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

Waleed Talib Abdullah Al Hashmi

PhD

University of York

Education

December 2018

Abstract

With most selection practices for initial teacher education programmes (ITEPs) focusing on applicants' cognitive attributes (academic records, subject knowledge, and so on), less attention is paid to evaluating applicants' non-cognitive attributes such as motivation and resilience. Evidence indicating the role of non-cognitive attributes in teacher effectiveness and the limitations of the current selection tools points to a need for better selection measures. Initial results on the development of situational judgment tests (SJTs) for the selection of prospective teachers in the UK are promising (Klassen et al., 2014b, 2017b), encouraging further research in a wider range of contexts. The current study extends this work to the context of Oman and aims to (a) report the development process of an SJT for selecting ITEP applicants in Oman and (b) describe the initial evaluation of the developed SJT by exploring the test's reliability and validity and applicants' reactions to it.

The research design consists of four phases. In Phase 1, the necessary non-cognitive attributes were explored with a convenience sample of key stakeholders (college tutors (n = 2), school principals (n = 74), supervisors (n = 63), teachers (n = 50)). The results suggested five non-cognitive domains. The domains were used in Phase 2 for developing an SJT with groups of working teachers (N = 116). The developed SJT was piloted in Phase 3 with a convenience sample of first year students (N = 171), and then implemented in Phase 4 with other criterion measures (N = 142).

The results show that the developed SJT has good internal consistency (α = .75). The correlation between the SJT scores and other measures indicates that the SJT correlates significantly with two facets of the Big-Five personality measure; that is 'conscientiousness' (r = 0.29) and 'agreeableness' (r = 0.20). In addition, the SJT has a positive and significant correlation with the participants' GPA (r = .31), and a negative (but non-significant) correlation with the scores of the interview currently used in the admission process (r = -0.17). The participants' responses to the SJT content and use as a selection tool were positive. Recommendations for policy makers and for further studies are discussed.

Table of contents

List of tables

List of figures

Acknowledgements

First and foremost, my sincere prayers and deep thanks to Allah, the Wise and the Knower, who bring light after darkness, and bestow knowledge after ignorance. I would not have been able to complete this thesis without His help to overcome the challenges and the moments of weaknesses.

I would like to extend my grateful thanks to my supervisor, Professor Rob Klassen, for his guidance and support during my research journey as a mentor and a friend. I would also like to express my thanks to Professor Chris Kyriacou for his invaluable comments and feedback on my work.

A great appreciation is also due to the government of my country (Oman) for sponsoring my study and all the facilities they provided me to accomplish my goal. I extend my thanks to all the academics, school principals, supervisors, teachers and students who generously offered their time, and support during collection of data.

Last but not least, all the thanks and love to my grandfather, grandmother, mother, brother and sisters for all the moral support and prayers they made for me and my small family. My deep love to my wife Faiza who stands beside me all the time, before during and after conducting this study. I extend my love to my beloved children: Majd, Waqaas, Maryam, Balqis and Saba. Oh Allah, save them all and help me to reward them for their patience and sacrifice.

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Dedication

To the souls of my late father (Talib), my grandmother (Thurayya) and my grandfather (Abdullah). You have always been a source of inspiration to me. Without you, I would not be who I am now.

I ask Allah to give you all the mercy and forgiveness. Ameen.

Chapter 1 Introduction

Teacher effectiveness is crucial for improving student attainment and meeting the objectives of the education system. Although there is currently no consensus on a clear definition of teacher effectiveness, research reveals the importance of teachers' non-cognitive attributes such as personality traits, beliefs, motives, and dispositions. Therefore, a better understanding of these attributes for prospective teachers prior to entering teacher education programmes could help to predict their future performance. Initial teacher education programmes (ITEPs) around the world focus mainly on the cognitive attributes of the applicants during the selection process, with less attention paid to non-cognitive attributes (Casey & Childs, 2007; Ingvarson et al., 2013). This study responds to concerns about the quality of new teachers in Oman by focusing on the current admission process for undergraduate students to ITEPs in the country. Specifically, it explores the development process and initial findings of situational judgment tests (SJTs) to better understand the non-cognitive attributes of new ITEP applicants in Oman.

This chapter gives an overview of the key aspects of the study in seven sections. Sections 1 and 2 highlight the background and the rationale for the problem found in the literature, and the context in Oman, respectively. Section 3 focuses on the significance and contribution of the study; and the key terms used in the study are defined in Section 4. The research questions and research design are illustrated in Sections 5 and 6, respectively. Finally, Section 7 describes the overall structure and the main components of the study.

1. 1 Background in the literature

This study concerns the importance of identifying non-cognitive attributes during the ITEP selection process for predicting future teachers' effectiveness. Several studies highlight this concern, looking more closely at both the cognitive and non-cognitive attributes of the applicants, in order to improve the quality of teachers (Barber & Mourshed, 2007; Jacobowitz, 1994; OECD, 2005). The selection process must consider the different reasons for choosing teaching as a job, looking not only at the best applicants in terms of academic performance, but also at other factors and characteristics, such as motivation for teaching (Gore, Barron, Holmes, & Smith,

2016). In addition, an early selection process could better predict how the new entrants would serve students in the future (Atteberry, Loeb, & Wyckoff, 2013). However, there are challenges to (a) understanding the key non-cognitive attributes that are necessary for screening the applicants, and (b) building a better selection procedure that is able to measure these attributes. Here, the issues around these challenges, as discussed in the literature, are highlighted for a better understanding of the rationale for this study.

Firstly, several attempts have been made to better understand the influence of non-cognitive attributes on teacher effectiveness. However, research in this area has reported at least three fundamental challenges to this: defining the attributes, their degree of influence, and the identification of the attributes (in effect, what they are and how they can be identified). The expression 'non-cognitive attributes' appears in different terminologies and has different components. However, the literature generally uses 'non-cognitive' to refer to attributes such as motivation, communication, and self-control, in contrast to 'cognitive', which refers to personal abilities in domains such as literacy and numeracy (Gutman & Schoon, 2013). In other words, non-cognitive attributes are whatever cognitive attributes are not (Kill, 2018). Several studies in education note the importance of non-cognitive attributes for teacher effectiveness (Klassen & Tze, 2014a; McGeown, St Clair-Thompson, & Clough, 2015). However, Brunello, and Schlotter (2011) argue that less consensus exists around the 'malleability' of non-cognitive attributes – or, the point up to which they can be changed. In addition, there is less consensus around identifying the key non-cognitive attributes, or the attribute(s) seen as the most important for teacher effectiveness. Part of the problem is the gap in defining the necessary key attributes of new teachers between schools, higher education and training and evaluation bodies (Relf & Hobbs, 1999). Besides, non-cognitive attributes are greatly affected by cultural factors, thus what is seen as necessary in one context might not be seen as such in another (Klassen et al., 2018; Zhou, 2016).

Secondly, the literature highlights challenges with the current selection practices used to measure the non-cognitive attributes of ITEP applicants. The widely used tests to screen applicants often miss - or, more accurately, do not adequately capture - the soft skills, personality traits, goals, motivations, and preferences that are valued in schools (Heckman & Kautz, 2012). Furthermore, the

tests are themselves limited, in terms of validity and reliability (Atteberry et al., 2013; Klassen & Kim, 2017c). At the heart of this dilemma in the selection process is the controversial question of whether teachers are 'born' or 'made'. The different views on this issue lead to different perspectives on the selection process. The 'born' stance considers selection a fundamental stage in the development process of teacher effectiveness, whilst the 'made' proponents rely on the role of the training programmes for producing good teachers (Klassen & Kim, 2017c; Kunter, Kleickmann, Klusmann, & Richter, 2013a). A combination of the two views, in the model developed by Kunter et al. (2013a), clarifies the role of non-cognitive attributes and the training programmes for the development of teacher effectiveness. The model in Figure 1.1 shows that the process of developing prospective teachers' competence and behaviours in learning opportunities (pre-service or in-service training) is influenced by two main factors: contextual (education policy, the characteristics of the ITEP, and so on), and personal (cognitive, non-cognitive, and background factors). The degree of influence plays a role in the outcomes of both students and teachers. Thus, the model illustrates implicitly the importance of considering the applicants' personal characteristics during the ITEP selection process.

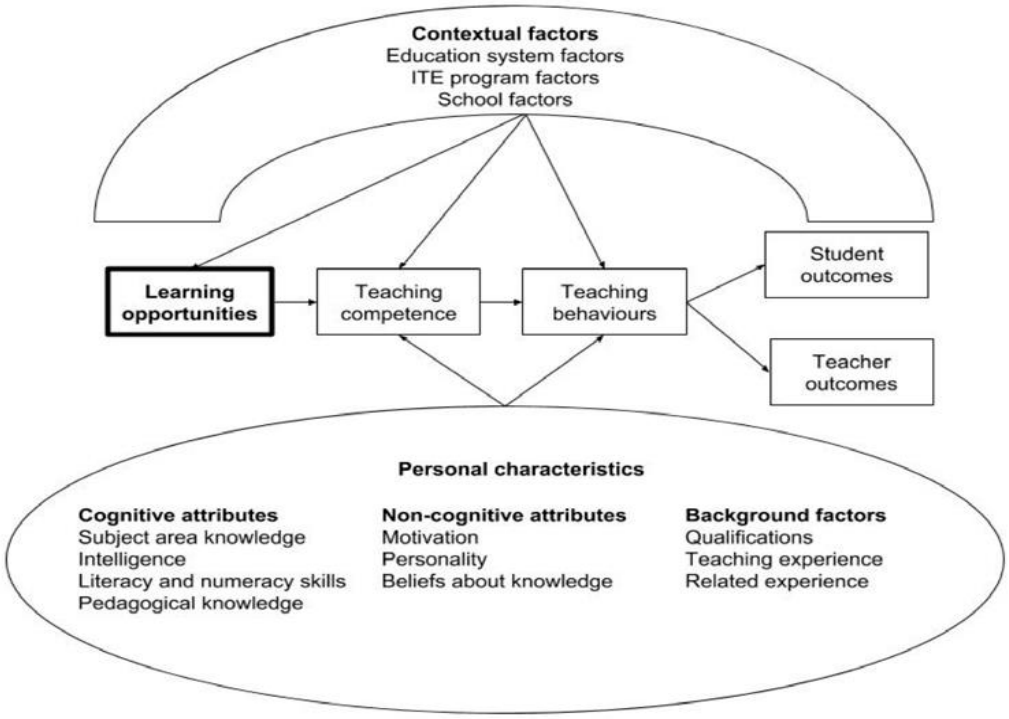


Figure 1.1 Model of the development of teacher effectiveness (in Klassen & Kim, 2017c; adapted from Kunter et al., 2013a).

Against the backdrop of the above, the challenges to understanding the non-cognitive attributes of effective teachers and the limitations of the current practices in ITEP selection reveal the need for further research. One way of achieving this and expanding the knowledge is consideration of selection in other professions. Medical schools in the UK and other countries, for instance, use a combination of screening measures, including SJTs (Patterson, Zibarras, & Ashworth, 2015b). An SJT is a kind of simulation test in which the applicant is presented with a variety of situations he or she would be likely to meet on the job. Recent research shows that SJTs can be used to assess and predict knowledge and attributes related to job performance (McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001; Weekley & Ployhart, 2006). Compared with personality tests, SJTs have good levels of reliability, predictive validity, and incremental validity for testing professional attributes (Patterson et al., 2012b). Nevertheless, SJTs have not been implemented widely in selection processes in Europe or other parts of the world, with the exception of the US (Lievens, 2006; Lievens, Peeters, & Schollaert, 2008). To the researcher's knowledge, SJTs have not

been widely applied or tested for teacher selection. Thus, this study attempts to explore the use of SJTs to overcome the challenge of understanding the non-cognitive attributes of ITEP applicants during the admission process. The study benefits from the initial findings of a current project for developing SJTs for teacher selection, led by Professor Robert Klassen at the University of York (Klassen et al., 2014b, 2017b). This study extends that research into a non-Western context (Oman).

1. 2 Context in Oman

The two challenges discussed earlier – defining the key non-cognitive attributes and the selection process –also exist in Oman. Although teacher effectiveness remains undefined, several studies warn that the quality of new Omani teachers is unsatisfactory in terms of non-cognitive attributes, such as motivation and commitment to the profession (Chapman, Al-Barwani, Al Mawali, & Green, 2012; OMoE, 2012). Al Tobi (2005) finds that almost half of all teachers surveyed wish to leave the profession. He recommends testing the perceptions of applicants throughout the selection and recruitment processes. Similarly, several studies in Oman recommend the modification of the selection criteria for trainee teachers to ensure a better selection of suitable candidates (Alkhausi et al., 2015; Al Barwani, 2002; Al Harthy, Jamaluddin, & Abedalaziz, 2013; Chapman et al., 2012; Issan, 2011).

Currently, the College of Education (CoE) at Sultan Qaboos University (SQU) in Oman is the main publicly funded source of Omani teachers in the country. The selection for CoE courses is based mainly on the applicants' academic results in secondary school (grade 12), with short interviews and aptitude tests for two subjects (physical education and art). Although the entrance requirements for the CoE at SQU attracts students with high academic performance in secondary school, a study investigating the career paths of the graduates found that about half of the participants were only somewhat or not at all committed to teaching as a career when they graduated (Chapman et al, 2012). In addition, the Al Barwani (2002) study at SQU indicated a nonsignificant correlation between secondary school certificate results and performance at college in terms of the students' grade point average (GPA). Further investigations at SQU suggest that both cognitive and non-cognitive characteristics contribute to students' success at the university, and that better selection measures may be needed for ITEPs in Oman.

1. 3 Significance and contributions of the study

The previous sections, detailing the background to this study in the literature and the context in Oman, highlight the importance of this study. In this section, the researcher addresses the contribution of this work to the education system in Oman, and to the literature in general.

The results of this study are intended to contribute to the selection process in Oman and to the educational sector in general, in at least four aspects. Firstly, it will help to develop a better understanding of the non-cognitive attributes of effective teachers in Oman. Moreover, it attempts to change the policy focus of the selection process to give more attention to applicants' non-cognitive attributes. Thirdly, it is the first study in Oman, to the researcher's knowledge, which attempts to improve the selection process with a tool developed by teachers. Finally, this study is a starting point for further research to test the validity of SJTs for predicting future teachers' performance.

In addition, the findings of this study will contribute to the growing body of research on the use of SJTs for the selection of prospective teachers, specifically in a non-Western culture (Oman). The findings will contribute to further studies seeking to understand the similarities and the differences between countries in terms of key non-cognitive attributes of prospective teachers.

In summary, at least four aspects shape the rationale and significance of this study:

- The lack of consensus in the literature around the key non-cognitive attributes of the effective (prospective) teacher is one aspect. This issue seems to be a priority in Oman, as studies in this area are scarce.
- The identification of non-cognitive attributes of effective teachers in a non-Western country (Oman) adds to the existing knowledge and supports future comparative research across cultures.
- In the researcher's experience, having spent 20 years in the educational sector in Oman, there is a real concern about the quality of new teachers due to the lack of non-cognitive attributes, including motivation and commitment. Part of the problem is the current admission procedures used by the teachers' preparation colleges.

- Despite the long history of SJTs, their use for the admission in the ITEPs is still new. This study benefits from the work of Klassen et al. (2014b, 2017b) in this area and expands the knowledge to a new context (Oman).

1. 4 Key terms

Here, brief definitions of the key terms used in this study are given, as follows:

- Non-cognitive attributes: includes all personality traits, beliefs, and dispositions such as motivation and enthusiasm. They are within-person variables, and they differ from cognitive attributes, which are related to the numeracy and literacy skills.
- Initial teacher education programmes (ITEPs) in Oman: colleges that offer a bachelor's degree in education to undergraduate students aiming to become teachers.
- Selection/admission process: process used to screen applicants for the ITEPs.
- Prospective teachers: Omani students enrolled in the ITEPs in the country to become teachers of grades 5-12 in the government Omani schools. Schools with lower classes (grades 1-4) are excluded because, currently, most of the ITEP graduates in the country enrol in schools with higher grades (5-12).

1. 5 The aim and research questions

Recent studies concerning the challenges facing the education sector in Oman confirm the need to test for the non-cognitive attributes of prospective teachers in order to improve the quality of future teachers. This is supported by the researcher's own 20 years of professional experience in the education sector. Identification of the key non-cognitive attributes is the first step to develop a better measuring procedure. In this study, the key attributes will become the inputs in the development process of the SJT.

Thus, the main aim of this study is to explore the SJT development process and initial findings on the use of the test to better understand the non-cognitive attributes of the undergraduate applicants to ITEPs in Oman. This will be done by developing the test and then testing its reliability and validity and applicants' reactions.

Specifically, this study aims to answer the following research questions:

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

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

1. 6 Research design

In order to address the above research questions, the methodological approach taken in this study is a combination of qualitative and quantitative methods, divided into four phases, as shown in Table 1.1. The research design is built from a review of the literature, specifically adapting the work of Klassen et al. (2014b, 2017b) to the context of Oman.

Table 1.1 A summary of the research design

Phase	Aim	Tools	Participants
One	Exploring the key non-cognitive attributes	Literature review – analysis of official documents – interview – questionnaire	College tutors, school principals, supervisors, teachers
Two	Developing the SJTs	Focus group(s)	Expert teachers
Three	Piloting the developed SJTs	The initial SJTs	New ITEP students
Four	Implementing SJTs to explore reliability, validity, and applicants' reaction	SJTs + BFI + Applicants' feedback	New ITEP students

The participants are stakeholders from the Ministry of Education (MoE) and the ITEPs in Oman (college tutors, school principals, supervisors, teachers working in government school (grades 5-12), and new ITEP students). At all stages of the study, participants were asked to express their agreement to participate by reading and signing a detailed consent form, approved by the University of York Education Ethics Committee. Chapter 4 explains in detail the research methodology of the study.

1. 7 Organisation/ layout of the study

This study comprises seven chapters, including the introduction. The next chapter concerns the context in Oman. It gives an overview of the country and the educational system. It also discusses the current practices and challenges on the effectiveness of teachers and the ITEP admission process in the country. Chapter 3 reviews the literature in four sections: personnel selection, non-cognitive attributes, ITEP selection, and the SJTs. In Chapter 4, the methodology of each phase of the study is discussed. The findings and results are presented in Chapter 5 and discussed in Chapter 6. Finally, Chapter 7 concludes by summarising the main findings, the limitations, and makes some recommendations for policymakers and further studies.

Chapter 2 The Context in Oman

To better understand the study in the context of Oman, this chapter aims to (a) provide an overview of Oman and its education system, (b) explore teachers' effectiveness and the non-cognitive attributes of Omani teachers, and (c) highlight the initial teacher education programmes (ITEPs) admission process in Oman, with its current practices and challenges.

This chapter is structured in six sections. The first two sections give an overview of Oman and its education system. Related issues about teachers in Oman are highlighted in section three. Sections four and five focus on the preparation and selection processes for prospective teachers on the ITEPs, respectively. Finally, section six summarises the main findings of the review of the Omani context and explains how this contributes to the aim of the study.

2. 1 Oman: a general overview

This section gives a summary of the demographic, economic, and social aspects of the context in Oman. An awareness of these aspects may shed light on the factors which influence the education system in the country, specifically the nature of the teaching and learning processes in Omani schools.

Oman is located in the south-eastern corner of the Arabian Peninsula, overlooking an important sea trade route between the Gulf and the Indian Ocean. It is bordered to the south-west by Yemen and the west by Saudi Arabia, and by the United Arab Emirates to the north and west. Oman has an area of 309,500 square kilometres divested between coasts, islands, mountains, sands, and green lands. It comprises 11 governorates, namely Muscat (the capital), Dhofar, Musandam, the Dakhiliyah, the Buraymi, the Dhahirah, the North Batinah, the South Batinah, the North Sharqiyah, the South Sharqiyah, and the Wusta. Each of these governorates includes a number of wilayats (districts), with a total of 61wilayats (NCSI, 2017; OMoI, 2016).

Islam is the official religion of Oman. Many Omanis are Ibadhis, which is a practice of Islam, distinct from other forms in its explanations of some Islamic norms. It is only found elsewhere in parts of North Africa (parts of Algeria and Libya).

Historically, Oman had governmental control of some countries in East Africa and the Indian Ocean. This expansion gave much cultural diversity to Oman from its Arab neighbours, and much diversity in the ethnicity of its people, with some citizens originally from East Africa or Pakistan (Common, 2011).

Looking at demographic factors, according to the National Centre for Statistics and Information (NCSI), the total population in Oman in July 2018 was 4.6 million inhabitants, of whom 56.3% were Omanis. In mid-2017, a distribution of citizens by age included approximately 42% under 17 years old, 23% aged 18-29 years old, 29% between 30 and 59 years old, and 6% of more than 60 years old. Indicators for 2016 showed that the total fertility rate (live births per Omani woman) was four, whilst the crude death rate per 1,000 people was three. Life expectancy at birth was around 74.7 for males and 79.3 for females. Regarding the workforce, there were approximately 2.3 million workers in 2016, 90% of whom were in the private sector. However, the majority of Omani workers were in the government sector, with a total of 84%, with just 12% in the private sector. The official number of job seekers in Oman, in September 2017, was approximately 44,000, of whom 63% were female. Approximately 48% of registered job seekers were aged 25-29 years, and some 40% had a university degree or higher qualification (NCSI, 2018).

In terms of internal policy, Oman is led by the 'Sultan', and is thus described as a 'sultanistic regime' (Common, 2011). The basic law of the state in Oman, issued in 1996, provides the legal framework to define and oversee the work of the various state authorities, including the Sultan's functions. These include guiding ministries and other agencies on their general responsibilities and objectives and clarifying the public rights and duties of citizens. According to the basic law, the Sultan is the head of state and the supreme commander of the armed forces. Respect for him is compulsory and his directives must be obeyed. He is seen as an icon for the unity of the Omani nation, and he is the guardian of the country. Whilst the government (the ministries) can issue laws and regulations related to its work, these laws should not conflict with the policies and guidance of the basic law (OMoI, 2016).

In 2011, a year after the start of the 'Arab Spring', the name given to the demonstrations and revolutions that occurred in a number of Arab countries, including Oman, a new amendment was issued by the Sultan to the basic law. This gave legislative and audit powers to the Council of Oman (Majlis Oman). This Council

comprises three main bodies: the Council of Ministers (representing the government), the State Council (Majlis al Dawla), and the Consultation Council (Majlis al Shura). The members of the Majlis al Dawla are appointed directly by the Sultan from among Omani citizens who have assumed high levels of expertise in the public and private sectors. In contrast, Majlis al Shura's members are elected by citizens as representatives of their wilayats (districts). According to the new amendment, the two Councils must both confirm the decisions issued by the Council of Ministers. The amendments also grant the State Council and the Consultation Council the right to review and give feedback on all regulations proposed by the Council of Ministers before submitting them to the Sultan, who then issues them as law (OMoI, 2016). Despite these amendments, members of both councils, notably the Consultation Council, are seeking further powers and authority for use in the revision and formation of national decisions.

The social and economic development process in Oman occurs in line with strategic plans. The government issued two main strategic plans: 1976-1995 and 1996-2020. Each strategy is divided into five-year plans. In 2010, Oman was named the most improved country of the last 40 years, according to indicators of the development of income, health and education (Wyatt, 2013). However, and like its neighbours in the Arab Gulf Countries Committee (AGCC), the development process in Oman faces major challenges. Firstly, government income is strongly dependent on oil revenues – from 77% in 1995 to 81% in 2010. In 2016, gas and oil revenue comprised some 68% of the total government income. Moreover, the private sector is largely dependent on governmental financial support. There is also an increasing number of job seekers among the national citizens, alongside increased reliance on expatriate labour in many economic areas. Another challenge is the incompatibility of the large budget allocated for education and the quality of its outcomes (Al Barwani, 2002; NCSI, 2018; OEC, 2017a).

The main challenges facing the development process in Oman (dependence on oil, the increase in unemployment, and so on) have been widely discussed, with solutions being put into practice. The government is working on a third strategic plan (2020-2040), with broad participation from stakeholders, including citizens. Oman's vision for 2040 focuses on three strategic themes: people and society, economy and development, and governance and institutional performance. In addition, and prior to

the end of the current strategic plan, the government is focusing on five key economic sectors: tourism, transport and logistics, manufacturing industries, fisheries, and mining. Investment in these sectors is believed to promote diversity in the economy and produce new jobs for citizens (OEC, 2017a). Finally, the investment in education and its quality improvements are an ongoing concern of the government in Oman in order to ensure the development of human resources.

In line with the aim of this study, the next section gives a summary of the education system in Oman, including its practices, challenges, and proposed future initiatives.

2. 2 The education system in Oman

This section explores the education system in Oman in terms of its management, organisation, general indicators, and key obstacles. Achievements and challenges for both the school system and the higher education system are presented separately.

Education is one of the fundamental components of the Omani government's policies geared towards providing citizens with better modern lives. It is essential to develop the knowledge and skills of a people in order to build a more competitive economy. The philosophy of education in Oman emphasises the need to

> 'develop citizens who demonstrate their faith in Allah the Almighty and follow the principles and values of Islam. It also aims to encourage Omani citizens to practise their loyalty to their country and the Sultan, to understand current events and to respond to these events in an appropriate manner. The Philosophy of Education endeavours to encourage Omani nationals to acquire scientific thinking skills and contribute to achieving sustainable development across all sectors of Omani society' (OEC, 2017b, p.11).

The philosophy includes 16 principles (including identity and citizenship, national prestige and respect, good values and behaviours, education human rights and duties). The philosophy and its related principles are considered to be the main reference and national framework for all educational institutions in the Sultanate.

Education is provided primarily by two ministries. Firstly, the Ministry of Education (MoE) is responsible for supervising pre-school education and providing

school education for grades 1-12 in the public and private schools. The Ministry of Higher Education (MoHE) is responsible for higher education, supervising public and private universities and colleges. In addition to these two ministries, technical education and vocational training is the responsibility of the Ministry of Manpower, whilst medical science and nursing institutes are the responsibility of the Ministry of Health (OMoE, 2014a). Public expenditure on education in Oman was reported to be 4.6% of the GDP and 26.1% of total government spending. In 2014/2015, approximately 17% of the national budget went to the MoE, whilst some 6% went to higher education (OEC, 2016b).

The management of the education system in Oman is the responsibility of the Education Council, which was established in 2012. The Education Council plays a role in coordinating the different ministries and authorities responsible for implementing education in Oman. It is concerned with drawing up general policies of education and directing, following up and evaluating education according to state public policy and the requirements for the future development of the country. Members of the Education Council are representatives of the authorities responsible for the education system in the Sultanate. It is led by the Minister of the Diwan of the Royal Court (Al Jarida Al Rasmiyya, 2012; OMoE, 2014a).

Following this general overview of the structure and management of the education system in Oman, the next two points highlight the components, main statistical data, achievements, and challenges of the two main educational sectors: the school system and the higher education system.

2. 2. 1 The school system in Oman

The schooling system in Oman includes all grades, from the preschool level up to grade 12. It has seen significant change in terms of quantity and quality since the role of the Sultan in 1970. Prior to 1970, there were just three schools in the country, educating some 900 male students. In the academic year of 2015/2016, the figures show that there were 1,068 government schools, with 540,000 students in grades 1-12. Approximately half of the students were female (49.8%). There were 56,586 teachers, 82% of them Omanis. In addition to the teachers, there were 11,648 administrators and technicians. The mean class size was 27 students for grades 1-10 and 25 for grades 11-12. The mean number of students for each teacher was 9.5 (OEC, 2016a).

To improve the quality of the education system, in 1998, the government established a new reform of the schooling system called ‘basic education’. Basic education is one of the outcomes of the vision for Oman’s economy in 2020, presented to the government in 1995, which placed emphasis on human resources development as the main driver of sustainable development in the country (Al Barwani, 2002). The basic education system replaced the old three-stages system – elementary (grades 1-6), preparatory (grades 7-9), and secondary (grades 10-12). The current system also has three stages but has different categories. Cycle one of basic education includes students in grades 1-4. The cycle one schools have boys and girls in the same classes, while the teaching staff are all female. Cycle two covers students in grades 5-10 in single-sex schools. Schools for students in grades 11 and 12 are ‘post-basic education schools’. In some conditions, especially in rural areas, one school might include two or more cycles (OMoE, 2012). The data for the school-year of 2017/2018, in Table 2.1, identify the number of schools, students, and teachers in the government schools, according to the type of school and gender (OMoE, 2018b).

Table 2.1 Schools, Students, and Teachers in Public Schools in Oman (2017/ 2018)

Type of school	Schools				Students			Teachers		
	Male	Female	Coed.	Total	Male	Female	Total	Male	Female	Total
1st-cycle schools			276	276	82189	81188	163377	0	13765	13765
2nd – cycle schools	168	105		273	98185	73870	172055	8153	6104	14257
Schools (10-12)	44	45		89	27273	29487	56760	2379	2621	5000
Schools (11& 12)	15	10		25	9089	5899	14988	781	546	1327
Continuous Schools (multi-levels)	127	49	286	462	74332	97512	171844	6947	15089	22036
Total	354	209	562	1125	291068	287956	579024	18260	38125	56385

Source: OMoE (2018b)

As well as changing the structure of the schools, the basic education reforms included other quality enhancing policies. Students under the new reform, for instance, learn English and information technology (IT) from grade one. Moreover, the post-basic education curriculum includes both core and elective subjects. The new reforms emphasise a student-centred approach to teaching and learning. Class size is reduced to 30 students per class in cycle one, and 35 students per class in cycle two and post-basic education. School buildings have been improved by the introduction of learning resource centres, computer labs, rooms for curricular activities and for environment life skills, stores, canteens, and healthy activities. Furthermore, school management has been given greater authority for planning and self-management. Finally, and as a response to the inadequate amount of school time in Oman, compared with international indicators, the school year was increased by the basic education reform to 180 days of 'instruction' per year. The length of the school day has increased from six periods to eight periods (each period being 40 minutes) (OMoE, 2012).

Although the quantitative and qualitative improvements to the schooling system are remarkable, recent studies illustrate the range of challenges to learning quality and student achievement. Studies conclude that the main challenge that the basic education system faces is the low achievement levels of students in basic education

skills, as compared to the achievements of students at the international level (Oman newspaper, 2014). Trends in International Mathematics and Science Study (TIMSS), for instance, indicates that Omani students are far behind students in most participating countries. In TIMSS 2011, Oman was ranked 46th in mathematics, of a total of 50 participating countries (OMoE, 2011b). In addition, in Progress for International Reading Literacy Study (PIRLS) 2011, Oman's position was 44 of 45 participating countries (OMoE, 2011a). Although the latest data from TIMSS 2015 show improvements in the performance of Omani students, results remain low. For instance, in mathematics, Oman ranked 39 of 49 participating countries in fourth-grade results. Despite the low performance, the TIMSS attitude survey showed that Omani students felt confident of their ability; and teachers also felt they were 'very well prepared' to teach. The MoE suggest that both students and teachers have unrealistic expectations of their performance (OMoE, 2012). However, a further investigation is needed by linking students results with other measures of teachers' effectiveness.

2. 2. 2 The higher education system in Oman

The provision of higher education to the youth in Oman is essentially a new phenomenon. It began in 1986 with the first (and thus far the *only*) government university in Oman – the Sultan Qaboos University (SQU). In the academic year of 2015/2016, there were 96 higher educational institutions in Oman: 41 government institutions and 28 privates. In addition to SQU, the government institutions include the Colleges of Applied Sciences, the Colleges of Technology, the Institutes of Health, the Institute of the Shari'a Sciences, and the College of Banking and Financial Studies. The number of students enrolled in higher education in Oman was 135,493, of whom 59% were female. In addition to the private and government suppliers of higher education, some students study at overseas universities. In 2015/2016, there were 6,297 students enrolled in higher education outside the country, 39% of these were female (OMoHE, 2017).

Despite improvement in the number of higher education providers, the quality of the graduates from higher education programmes is considered to be a concern. Belwal, Priyadarshi, and Al Fazari (2017) find that basic generic skills are the key to developing students for higher education in Oman. The five most significant employability skills in Oman are computing skills, the ability to work in teams,

English language proficiency, prior training, and personality. The findings of a 2017 survey of graduates of higher education, conducted by the MoHE, reveal a gap between the skills learned in higher education and the required skills for the labour market. Employers in the labour market in Oman state that Omani graduates are typically weak in communication skills, self-confidence, time management, and dealing with work pressure (OMoHE, 2018). In addition, data from the Ministry of Higher Education indicate that, in 2014/2015, the drop-out rate from higher education institutes in the country was approximately 8%, with 64% of the leavers being male (HEAC, 2017). Al-Ani (2017) claims that the education system in Oman uses traditional learning methods which are unable to meet diverse learning and working needs. She suggests the need for alternative education tools that provide more opportunities for students to learn, thus producing graduates with skills that meet the demands of the working environment.

To conclude, both the schooling system and the higher education system in Oman have benefited from remarkable increases in input (number of schools, institutes, students), whilst the quality of their outcomes remains a challenge. The Education Council is developing a national strategy for education in 2040, including projects such as the School Education Law, the Higher Education Law, and the National System for Quality in Higher Education (OEC, 2014). A full analysis of these projects' success in enhancing the quality of the graduates is therefore not expected for another 20 years or so. However, it is fair to say that the effective implementation of the proposed projects requires effective teachers who are able to convert the initiatives into good practice. An overview of the current status of teachers in Oman is presented in the next section.

2. 3 Teachers in Oman

Before looking at preparation and selection for ITEPs, it is important to understand the current practices and challenges surrounding teacher effectiveness in the profession. Thus, this section gives an overview of the teaching force in the school system in Oman in six subsections. First, a general overview is given, including some statistics relating to teaching staff. Then, the tasks and responsibilities included in the teachers' job description are reviewed. The recruitment, training, and evaluation procedures are then presented in separate subsections. Finally, the current challenges

for teacher effectiveness in Oman, notably in terms of non-cognitive attributes, are discussed.

2. 3. 1 A general overview

Due to a lack of qualified teachers in Oman, schools used to rely on expatriate teachers. These accounted for approximately 92% of the total teachers in 1980. However, due to the policy of 'Omanisation' (in effect, replacing expatriate workers with Omani citizens), the percentage of Omani teachers in the profession has increased rapidly (OMoE, 2012).

Recent data from the MoE indicate that, in the academic year of 2017/2018, there are 65,385 teachers in the government schools. Of these, 68% are females because, as noted earlier, the first cycle students (grades 1-4) are taught only by female teachers. Omani teachers account for 84% of the total: 89% of the female teachers and 73% of the male. Some 96% of teachers hold a university degree or higher qualification (a bachelor's degree, higher diploma, master's, or PhD) (OMoE, 2018b).

2. 3. 2 Teacher's job description

The job description for a school teacher is included in the Guidance for the Schools' Jobs in Oman. This is a national document, prepared and issued by the MoE and implemented in all government schools. It contains a description of the tasks and responsibilities of each job in the school (school principal, social worker, career guidance officer, and so on) and the working standards for each job. The job description does not distinguish between male and female teachers. The guidance issued in 2015 includes five categories of teacher: senior teacher, subject/field teacher, teacher for learning difficulties, special education teacher, and pre-school teacher. Here, the focus is on the subject/field teacher.

The description for the subject/field teacher includes 39 tasks and responsibilities, both within and outside the classroom setting. A review of the tasks indicates that the teacher should possess good academic knowledge and various non-cognitive attributes. Focusing on the non-cognitive attributes: the teacher must be, for example, 'committed to the profession's ethics and the job roles', show discipline by 'attending the daily school's queue/assembly and associating with its organisation' and should 'strengthen the national and job loyalty'. In addition, teachers are asked to

‘be cooperative with school, peers, parents and the community’, ‘take care of pupils and advise pupils for good attitudes’, and show ‘good attitudes towards pupils with Special Education Needs (SEN)’ (OMoE, 2015).

As well as tasks and responsibilities, the Guidance for the Schools' Jobs includes some theoretical standards about the workload for each job. For teachers, the standard is two teachers for each class. In other words, if the school has 10 classes, regardless of their levels, the school requires 20 teachers. The school day in Oman has eight periods, thus 40 periods per week. The guidance states that the minimum number of the teaching lessons is eight periods per week for the senior teacher, and 20 for other teachers (with each period being 40 minutes) (OMoE, 2015).

2. 3. 3 The recruitment process for new teachers in schools

The recruitment process for new teachers in schools is conducted centrally by the MoE through a number of procedures and standards. Firstly, the applicant for a teaching job in the government schools must have a university degree in education (a bachelor’s degree in education) or a non-educational university degree in a specific subject plus a higher diploma in education. Applicants for an English teaching role must also provide an additional certificate in an international standardised test of English proficiency (in effect, applicants must score a minimum of 6.0 on the International English Language Testing System (IELTS) or a minimum of 547 points on the Test of English as a Foreign Language (TOFEL)).

Applicants who meet the qualification requirements must then undertake an admission test to the profession. The recruitment test was built by experts from SQU in Oman. It includes (a) questions related to the applicant’s discipline, representing 70% of the test, and (b) questions related to teaching methods and educational psychology, representing 30% of the content of the test. Applicants for individual skills’ subjects (physical education, art, and music) are set a practical test related to their discipline. The pass score for the admission test is 50% for physical education, art and music, and 60% for other subjects (Alroya, 2016).

Although the current procedures and requirements were designed to ensure the quality of the entrants to the profession, the validity of the admission tests (in effect, their ability to predict the effectiveness of the enrolled teachers), to our knowledge, has not yet been tested. In the academic year of 2016/2017, there were 2,928

applicants, 88% of whom were female. They graduated from 47 teacher education programmes: 19 in Oman and 28 from elsewhere (the United Arab Emirates, Jordan, Saudi Arabia, Qatar, Bahrain, Egypt, Kuwait, and Yemen). Approximately 38% of the applicants did not pass the admission test (Alroya, 2016). Similarly, the latest data from the MoE, indicate that in 2017/2018, 3,045 graduates (of whom 90% were female) applied for a teaching job, and 38% failed the admission test (OMoE, 2018a). The large proportion of failed applicants illustrates the challenge of improving the quality of new graduates from the teacher education programmes. Therefore, the MoE provides in-service training programmes to enhance the quality of teachers. The next section details the current practices in professional teacher development in Oman.

2. 3. 4 Teachers' professional development

In 2014, the MoE opened the Specialised Centre for Professional Training of Teachers, with a vision 'to include teachers as active partners in the development of education, using best international practice, leading to achievement of the highest standards' (OMoE, 2014b, p.7). The centre runs four strategic training programmes: the Centre Associates Programme, which targets one senior teacher from each school in Oman; the Centre Arabic Experts Programme for teachers in first cycle schools (grades 1-4); the Centre Mathematics and Science Experts Programme for teachers in second cycle schools (grades 5-10); and the Expert Supervisors Programme, which targets educational supervisors to enhance their skills in supervising and developing teachers (OMoE, 2014b). In addition, the MoE provides short in-service teacher training courses at three levels: central, educational governorates, and school levels (OMoE, 2014a).

After being recruited and before joining the school, new teachers are provided with a training package. It consists of 75 training hours, divided into three themes. The first theme seeks to acquaint new teachers with Omani education philosophy, education plans, curricula philosophy, and guidance on teachers' rights and duties within the framework of the applicable laws. The second theme deals with the educational aspects of teaching methods, evaluation methods, and some aspects related to class management and educational media. Finally, the third theme details the specialised aspects of each subject, as well as providing a printed manual containing the documents required by teachers (OMoE, 2014b).

Despite the provisions of such training programmes, there is insufficient evidence for the influence of the courses on classroom practice and teacher effectiveness. However, some reports identify low commitment by teachers when undertaking the training courses, especially given that the courses do not have implications for salary or promotion (OMoE, 2012). Albelushi (2004) found that whilst the interviewed participants had not thought about their suitability for teaching when they chose to enrol on ITEPs, the training courses provided did not influence their attitudes towards teaching.

2. 3. 5 Teachers' assessment and evaluation

The evaluation system for teachers working in Omani schools has two main components. Firstly, the formative evaluation during the school year focuses on teachers' performance in the classroom setting. It is conducted individually by a senior teacher, the school principal, and the teacher's supervisor. The second form is the summative evaluation, which is an annual appraisal completed by the supervisor and the school principal together. The criteria for both assessments is very similar. The teacher is evaluated in both teaching and non-teaching practices at school. The evaluation items include commitment and discipline; developing positive attitudes and values; caring about appearance; accepting advice and feedback; good relationships with school, peers, pupils, and parents; strong personality and good class management; innovation in work and social activities inside and outside school; self-development; effective classroom management and time-management skills; raising pupils' motivation; and directing pupils' self-learning (Alyahmadi & Al-Kiyumi, 2014; OMoE, 2012).

Data from the teachers' evaluation report in 2008 identify challenges such as weaknesses in subject knowledge, overload of administrative work, weaknesses in teaching strategies, and problems using electronic facilities and equipment (OMoE, 2012). Although teaching hours in Oman are seen as low, compared to international standards (OMoE, 2012), the workload for teachers during the school day is seen as high by teachers, which causes dissatisfaction (Amzat & Al-Neimi, 2014). In addition, the teacher evaluation aims to identify teacher training needs and promotion opportunities, based upon current performance. However, Alyahmadi and Al-Kiyumi (2014) note a common belief among participants that teacher evaluation has little influence on the development of teacher performance or careers.

2. 3. 6 Teachers' effectiveness in Oman

The previous sections highlight the main practices and challenges in recruiting, training, and evaluating teachers in Oman. Although current procedures aim to improve the effectiveness and quality of teachers in schools, it is worth saying that there is not as yet an official document that defines an effective teacher. In addition, research measuring the effectiveness of teachers in Oman is scarce.

Al-Ani, Al-Barwani, and Al-Buloshi (2012) define effective teaching in Oman from the perspectives of a sample of teachers, school administrators, supervisors, and students (N = 3,487). The researchers developed a questionnaire of 85 items, comprising six domains: personality traits, professional characteristics, teaching strategies, implementation skills, academic characteristics, and community relationships. They found that personality traits had the highest mean of the domains. In addition, Al Barwani, Al-Ani, and Amzat (2012) explored students' perspectives of the most important characteristics for effective teaching in Oman (N = 2,628 students). The results indicate that 'community relationships' are the most important factor, and 'teaching strategies' the second. Moreover, Al-Rawahi (2010) found that, for a physical education teacher, it was necessary to have a strong personality, be a good role model for students, show patience and tolerance, and communicate effectively with students. Although previous studies note the importance of non-teaching factors for defining effective teaching in Oman, teachers' often lack these. Studies indicate that attributes such as commitment, motivation, and attitudes towards teaching are significant for working teachers (Al Harthy et al., 2013; Chapman et al., 2012; Klassen, Al-Dhafri, Hannok, & Betts, 2011; Issan, Al-Nabhani, Kazem, & Al-Ani, 2011; Zayed, Abu Hilal, & Diabat, 2011).

Currently, the MoE – in accordance with the Education Council Decision No. 4/4/2014 – is building a national framework relating to the policies of teacher preparation, rehabilitation, training, and selection mechanisms. There are four proposed main documents: the Professional Standards for the Omani Teacher, the Professional Training Document, the Professional Ethics for the Omani Teacher, and the Document of Professional Routes and Licenses. Such a framework is essential for outlining the standards for the different processes involved in selecting, preparing, and evaluating teachers – providing a standardised evaluation system for teaching policies and practices. This proposed framework arrives at a critical time, with Omani

schools expecting to need approximately 22,000 more teachers by the academic year of 2020/2021 (Alroya, 2016). The committee developing the national framework was established by ministerial decisions 248/2015 and 141/2016. The committee members represent concerned bodies and stakeholders, including representatives from the ITEPs in Oman. The next two sections highlight the current practices and challenges for the preparation and selection of prospective teachers for the ITEPs in Oman.

2. 4 The preparation of prospective teachers for initial teacher education programmes (ITEPs) in Oman

Here, the preparation of prospective teachers is discussed in three steps. The first section gives a general overview of the history and development of the ITEPs in Oman. In order to understand the preparation process, the case of the College of Education (CoE) at SQU in Oman is presented. The last step highlights the main challenges related to the quality of the preparation for the ITEPs.

2. 4. 1 A general overview

In 1970, the school system was extended to all citizens, creating a strong demand for new teachers. Due to the lack of trained Omani teachers, the vast majority of the teaching staff in the country's few schools were foreigners. In some parts of the country at that time, any educated Omani citizen could apply to become a teacher, regardless of whether they had any formal qualification, as long as they could read and write. In effect, there was an emphasis on quantity rather than the quality of the recruited teachers.

The first training of Omani primary teachers began in the academic year of 1975/76. The students who had completed the first preparatory grade (grade 7) were enrolled on a two-year preparation programme. Another programme was developed in 1977/78: a three-year programme which accepted successful holders of the third preparatory certificate (grade 9). In 1984, six teacher institutes were established, providing a two-year programme which accepted students finishing secondary school (grade 12). The applicants for those institutes went through a series of admission exams, interviews, and medical checks. The successful graduates from this two-year programme were awarded a diploma certificate in education, and recruited as primary teachers.

Later, to enhance the quality of teachers and due to the introduction of basic education, the government extended the preparation period to a four-year programme. Hence, all six teacher institutes were converted in 1995 to educational colleges. Before this, in 1986, the SQU was established, including a university college for education with a four-year programme. However, as a result of the oversupply, the six colleges of education were converted to more comprehensive applied colleges of sciences, and five no longer offer teacher education.

Until the academic year of 2016/2017, the ITEPs for undergraduate students in Oman were provided at two main institutions: the CoE at SQU, and the Applied College of Science in Al Rustaq for preparing English teachers. In addition, some private universities provide a higher diploma in education for graduate students (for example, University of Nizwa, University of Sohar, and University of Dhofar).

As a result of the limited places on teacher education programmes in Oman, and the strong competition between the applicants, many students enrol in teacher education programmes in neighbouring countries, such as the United Arab Emirates and Jordan. However, an unpublished report from the MoE indicates that a large percentage of the graduates from ITEPs outside of the country perform poorly on the employment admission test and in the professional evaluation process. As a result, the Education Council established a decision (no. 5/3/2015) on 21 July 2015, stating that the recruitment of new teachers in the academic year of 2016/2017 was limited to graduates from ITEPs in Oman and to those who had studied abroad with permission and scholarships from the government. The Council agreed to provide more placements for training teachers by reconverting the College of Applied Science in Al Rustaq to a college of education and opening more educational diploma programmes in private universities in the country (Alroya, 2016).

2. 4. 2 The case of the College of Education (CoE) at Sultan Qaboos University (SQU)

Students who have finished secondary school are prepared in a five-year programme in the CoE at SQU. The first year is a foundation course for students in English, maths, and IT. According to the Staff Guidebook (2013), the programme comprises 125 hours in total: 12 hours to study university requirements courses, 40 hours for college requirements, and 73 hours for specialisation requirements. In

addition, the college cooperates with the MoE to place the final year students in schools for teaching practice in a real teaching and learning environment.

In the Staff Guidebook 2013, the CoE describes the distinguished education graduate as, 'A leader who is empowered with specialized knowledge, expert skills, values of the field and society, and has the ability to utilize contemporary research findings to maximize self-learning through reflective practice and life-long learning in order to provide diversified optimal learning experiences for all students' (SQU, 2017, p.11). The CoE identifies five themes of candidate proficiencies: academic rigor and specialised experiences, diversified teaching, dispositions and values, research culture and lifelong learning, and technological skills. The five themes cover the conceptual framework for preparing and evaluating students in the college. The CoE recently received accreditation for seven years (2016-2022) under the National Council for Accreditation of Teacher Education (NCATE) standards (SQU, 2017).

2. 4. 3 Challenges in preparation programs

Here, the researcher focuses on challenges related to the preparation process and the quality of the graduates from the ITEPs in Oman. Firstly, according to data from MoE teacher recruitment tests, the graduates from the CoE at SQU are of a higher quality than the graduates from other colleges, notably those from abroad. However, in the academic year of 2013/2014, whilst approximately 81% of the CoE graduates at SQU in six subjects passed the recruitment tests, graduates in biology and Arabic did not meet the standards (Oman newspaper, 2014).

Secondly, looking at the quality of the educational courses in the preparation programmes, Al-Rawahi and Al Balushi (2011) measured student achievement levels in terms of professional competencies. A sample of 167 students from the CoE at SQU reveals that the most influential elements of the teacher preparation programme were the teaching practicum, the curriculum, and instruction courses, and the psychology courses, whilst the foundation and administration courses were not significant. Al Ganbousy, Al Harthi, and Kazem (2012) evaluated ITEPs in the CoE at SQU from the perspectives of the graduates of the 1990-2007 cohorts (N = 639). The results indicate that 40% believe there was an overlap between the courses, and that lecturing was the main teaching method used, and field trip method the least. In addition, the most common difficulty faced by graduates was the discrepancy between

knowledge acquired in college and the required knowledge in the field. Similarly, Alhashmi and Eissa (2010) identify an inadequacy in the practical educational courses, based on research with a sample of graduates from the Arabic education programme (N = 105). Finally, a joint study between the MoE and the World Bank found that although tutors in the ITEPs in Oman have good academic qualifications, their school teaching experience is limited. The study also reveals that students on ITEPs had just 6% of their total credits from courses in teaching practice (OMoE, 2012).

This review of previous studies illustrates the challenges for the content and teaching practices of ITEPs. The current decisions in Oman for limiting the acceptance of new teachers from the graduates inside the country, and the significant demand for new teachers, highlight the need for better outcomes from the ITEPs. It is not only important to enhance the quality of the process in the preparation courses; in addition, there is real concern about the quality of the new entrants. Many studies recommend modifying the selection criteria for trainee teachers to ensure a better selection of suitable candidates (Alkhausi et al., 2015; Al Barwani, 2002; Chapman et al., 2012; Issan, 2011). The next section discusses the current selection practices for ITEPs in Oman.

2. 5 The selection of prospective teachers for initial teacher education programmes (ITEPs) in Oman

This section highlights the current procedures used in the selection for ITEPs in Oman, in three subsections. First, an overview is given of the practice of selection for higher education programmes in Oman in general. The researcher then introduces the selection process for the CoE at SQU, as this was the only college which offered places in all subjects for undergraduate students until the academic year of 2016/2017. The last section presents the main challenges for ITEP selection in Oman.

2. 5. 1 Selection for higher education in Oman: a general overview

To facilitate the enrolment of students to higher education institutions, the Ministry of Higher Education in Oman established the Higher Education Admission Centre (HEAC) in 2006. Annually, the HEAC coordinates with higher education providers to keep abreast of the programmes they offer, the requirements for these, and the expected number of students for each programme. To simplify the application

procedures, the system enables students who are in their final year at school (grade 12) to submit their applications electronically to any higher education programme they choose, provided they meet its requirements. The applicants identify their preferences and are allocated seats according to their 'weighted average score' in the General Education Diploma (GED), or other equivalent certificate. The weighted average score is calculated using the following formula: (average grades of all subjects studied by an applicant) × 40% + (average of the subjects for a programme) × 60%. However, some programmes require students to sit an admission test and/or undertake an interview as an additional requirement.

The requirements differ from one programme to another. However, the applicants to any programme must be Omani, have a general education diploma or equivalent for the current academic year (newly graduated), and be in the age range of 16-25 years. The HEAC's data for 2016/2017 indicate that 517 undergraduate programmes, in different disciplines, were offered to students – approximately 49% in Oman and the rest scholarships outside the country. A total of 29,747 applicants were offered seats, from a total of 32,172 applicants who passed the GED exam or its equivalent (HEAC, 2017).

2. 5. 2 Selection for initial teacher education programmes (ITEPs): the case of the College of Education (CoE) at Sultan Qaboos University (SQU)

The requirements for teacher education programmes at the CoE at SQU vary according to the subject. A student who applies for the English programme, for example, should have GED results of 90% (or above) in English and 65% (or above) in Arabic and Islamic Studies. However, for subjects such as arts and physical education, applicants should also succeed in a competency exam in addition to their academic results (OMoHE, 2014). Once an applicant has met the standards, an announcement is sent by HEAC (via SMS) to offer the placement. Applicants who receive and accept an offer then become officially enrolled in the programme and are asked to complete the registration process.

During the induction week for new students, the CoE at SQU conduct interviews with the accepted students. According to the interview form for new undergraduate applicants, each interview is allocated 10-15 minutes and conducted by two or three college tutors. The applicant is evaluated according to seven aspects, based on the conceptual framework of the college, namely: care for academic

specialisation, enjoyment working with students with special education needs (SEN), attitudes towards and appreciation of Islamic and Omani values, awareness of the research culture, technological skills, language and communication, and general professional appearance. According to the interview form, the candidate should be informed that the interview is a prerequisite of admission to the CoE and the required score is 60% or above. However, and up to date, there is not much known about how 'effective' the interviews are for screening the applicants in the admission process.

In 2015/2016, 606 placements in education were offered to students who had finished secondary school. Some 60% (365 seats) of the educational programmes were offered by the CoE at SQU, and 100 seats were offered for English teaching programmes at the College of Applied Science in Al Rustaq. A further 232 were offered as scholarships outside the country. The number of placements increased to 933 in 2016/2017 due to the conversion of the College of Applied Science in Al Rustaq to a college of education. The data indicate that the process is competitive, particularly for female students. The number of students who applied for placements in education programmes – identifying them as their first, second, or third choice – was 4,925 in 2015/2016 and 6,577 in 2016/2017. In both academic years, approximately 70% of the students were female.

2. 5. 3 Challenges facing initial teacher education programmes (ITEPs) selection in Oman

In general, ITEP selection in Oman is highly dependent on applicants' results in grade 12 (the last grade in the school system). The applicants' non-academic skills are not given equal attention during the acceptance decision-making process. Al Barwani (2002) conducted research at SQU and found a non-significant correlation between secondary school certificate results and students' academic performance at college. She claims that exams in secondary school were typically information-based tests and students prepared for the exams by focusing on memorisation of knowledge, rather than using advanced skills such as problem-solving and analysis. A further investigation conducted at SQU to determine the skills that have a negative effect on students' achievement indicates that faculty members emphasise the weakness of students' preparation at secondary school, followed by students' weakness in self-learning skills, motivation to learn, self-reliance, and time-management skills. The

findings suggest that both cognitive and non-cognitive characteristics contribute to students' success at the university.

In the period 2013 to 2016, statistical data from SQU indicates students at the CoE perform well compared with students at other colleges. An average of 80 per cent of students completed their training programme in this period. About 1% of the registered students transferred from the CoE to other colleges while about 1-3% joined the CoE. Only 1% of students left their training because of academic or non-academic reasons (SQU, 2017). However, and although the entrance requirement for the CoE at SQU attracts students with strong academic performance at secondary school, a study of graduate career paths revealed that approximately half of the participants were only somewhat or not at all committed to teaching as a career after graduation (Chapman et al., 2012).

Although previous studies in Oman highlight the importance of enhancing the selection process for ITEP applicants, there is, to our knowledge, no published study that has developed a better selection method or proposed a model that better predicts the most effective applicants. Al Numani (2006) proposes some theoretical standards, though these are for hiring school teachers once they have finished their training programmes. Some research confirms the importance of personal interviews and admission exams for the selection process (Al Mahree, 2006; Gneema, 1996). However, Al Mahree (2006) and Madkoor (2005) argue that most interviews and admission tests are conducted quickly and with poor efficiency.

2. 6 Summary and contribution to the Study

As noted at the beginning of this chapter, the review of the context in Oman aims to highlight two main issues: (a) the admission process for the ITEPs in Oman, with its practices and challenges, and (b) teachers' effectiveness in Oman, notably in the non-cognitive attributes. The following points summarise the main findings from this chapter:

- Firstly, as a developing country, Oman emphasises improvements to its education system in order to produce high quality graduates with good levels of knowledge and skills. To achieve this, teachers are seen as the cornerstone

of the new initiatives contained in the national strategy plan 2040 for a better education system in Oman.

- Competitiveness in the labour market and changes to the global economic and social context are putting pressure on teaching as a profession. Teachers are not only responsible for expanding students' knowledge of academic subjects but must also develop students' non-academic skills. Hence, the preparation of prospective teachers for the ITEPs in Oman should be developed to better reflect changes in teaching tasks.
- Although the teacher education programmes in Oman have been developed to ensure higher quality teacher preparation, the selection process remains essentially the same. It is based largely on academic achievement, with little concern for the characteristics and non-cognitive attributes of the applicants. Thus, in order to enhance the quality of the school system in Oman, the education sector should focus on developing its selection mechanisms for prospective teachers.
- Currently, there are promising initiatives in Oman for enhancing the selection and preparation of prospective teachers, such as the National Teachers' Framework. This study will contribute to these efforts by outlining practical solutions to assessing the non-cognitive attributes during the ITEP admission process.

Finally, teaching in Oman is a civil service job. It is difficult to remove poor quality staff, who can remain in a job for many years. International reports of education show that the top-performing school systems implement more effective selection procedures for testing the skills and attributes of applicants before they enter teaching programmes (Barber & Mourshed, 2007). At this time, only limited attention is paid in Oman to evaluating candidates' non-cognitive attributes at entrance to ITE. Discussion in the literature of the non-cognitive attributes of effective teachers and how these can be measured during selection practices will be reviewed in the next chapter.

Chapter 3 Literature Review

In line with the aim of this study, the literature review focuses on four main areas. The first two sections review the literature from the teaching and non-teaching fields on personnel selection and non-cognitive attributes. Section three highlights the area of selection for admission into initial teacher education programmes (ITEPs). Finally, the related literature on situational judgment tests (SJTs) is discussed. A summary is then given of the main findings and their contribution to the study.

3. 1 Personnel selection

Vinchur and Bryan (2012) state that as long as there are organisations, there will be a need to determine who should and should not be selected for those organisations. Personnel selection (or employee selection) has played a central role in individual-organisational (I-O) psychology and has been seen by psychologists as a widespread activity. This section highlights in three subsections some findings about the selection of individuals. The first gives a general overview of the definition and history of personnel selection. Related theories and approaches are then presented. The final point highlights the fundamental questions for personnel selection.

3. 1. 1 Definition and historical perspective

Firstly, selection, as a human activity, can be seen as a natural behaviour used in people's daily routines. However, selection in the working or learning environment is simply a systematic process of making a decision about people. The process includes developing a series of activities to identify suitable applicants who are expected to make a positive contribution to the organisation's goals (Robert, 1989). The series of activities can generally include collecting and evaluating data and information on the applicants' backgrounds, academic knowledge, competencies, and other attributes. The decision based on the selection process might concern recruitment, transfer between jobs or roles, and/or promotion (Iles, 1999).

The first use of selection as a systematic process for organisational activities is thought to have been by US military forces during World War One to predict the performance of applicants (De Wolff, 1989). The practices of selection at that time were built on assumptions about the work environment (work done by individuals,

does not change very much, requires specific attributes, selection made by the organisation, and so on). However, concerns such as human rights, applicant privacy, the involvement of the court system, and many other changes in the nature of the work role, have resulted in a 'wake-up call' for personnel and selection psychology (Anderson & Herriot, 1997).

Since the 1990s, there has been renewed confidence in selection and recruitment systems, reflecting changes in the social and work environment. Four key issues have been identified as making changes to the dominant paradigm of selection. Firstly, selection is considered to be a strategy for change; thus, the focus is on knowledge, skills, abilities, and other factors (KSAOs). Secondly, multiple levels of interactions are introduced in the working environment between the person-work, person-task and person-organisation levels. Thirdly, the scope of selection now looks at the cross-cultural applicability, rather than just the local settings. Finally, research into selection has generated wider theoretical frameworks, including applicants' rights in the selection process and the psychological impact on candidates (Anderson & Herriot, 1997). However, the rapid changes in selection practices makes it difficult to evaluate the impact and value of the selection outcomes. In addition, changes in selection practices have affected working conditions and have sometimes been met by strong resistance (Anderson & Herriot, 1997; Robert, 1989).

3. 1. 2 Theories and approaches

There are different approaches and theories of personnel selection, affecting the development process and the outcomes for selection methods. Herriot (1989) identifies two main approaches to selection. Firstly, the traditional/prevailing/classical model starts by analysing the job, then selecting the criteria to measure, and choosing a measurement method. A validation study is then conducted and, depending on the results, a predictive test battery is set up. However, according to this model, selection is for prediction only — leaving out other aspects, such as relationship with the applicant, costs and benefits, and the social context of the selection. There is also the technological approach, which uses the design cycle from engineering science. Selection, according to this approach, has six steps: define the purpose and functions of the selection procedures, analyse the requirements that the procedure should meet, synthesise/make a design (creates or adapt), simulate to test the procedure, and

decision-making (accept or reject). The limitation of this approach is that it might be seen as a complex process requiring much time and effort.

A decade later, Iles (1999) expanded the approaches/models of selection and assessment and assigned them to five categories. The classical approach (the conventional model) involves determining the numbers and types of people to be recruited at specific times; defining the skills, competencies, and abilities required through job analysis; recruitment to attract people; making selection and placement; establishing performance management after selection; human resource development; and organisational change. Secondly, the strategic management approach is where recruitment should match needs. This focuses on 'who', 'why', and 'when' — leaving the 'how' questions for the psychometric model. The psychometric approach concerns the technology and efficiency of selection. It has a wide range of assessment methods, depending on what it is trying to assess. The fourth approach is the social process approach. It is interested more in understanding, through the selection process, the relationships between the candidate, the selection tool, the assessor, the organisation, and the social context. Thus, the major tool here is the observation and interview with the applicant in work situations. Finally, the critical discourse perspective on assessment focuses on the way in which power, knowledge, and practice support and reproduce each other.

Regardless of the number of selection theories and approaches mentioned in the literature, the selection process can, in practice, utilise different approaches at the same time. There are many factors which determine the strategy of selection. In the next section, the researcher highlights this issue by looking at the two fundamental questions concerning any personnel selection process.

3. 1. 3 Fundamental questions in personnel selection

Despite the revolution in selection and approaches to it, the two fundamental questions on any selection process remain the same: what should be assessed, and how is it done? Regarding the first question, Ryan and Ployhart (2014) summarise three views of what to assess. In the first view, the focus is on individuals, such as predicting applicant turnover. The selection, in the second view, aims to predict performance at unit or organisational levels (assessment related to KSAOs, for example). In the last view, selection aligns with organisational strategy, assessing

according to a competency-modelling process. In practice, the ‘what’ question focuses on cognitive attributes, non-cognitive attributes, personality, interests, emotional intelligence, and other constructs. In contrast, the answer to the ‘how’ question includes different methods and tools for selection, such as interviews, assessment centers (ACs), situational judgment tests (SJTs), and self-report measures (Ryan & Ployhart, 2014).

There are various influences on the answers to these two questions; such as the size of the organisation, the occupational areas, the cost of making a mistake (the level of risk), and the number of candidates for a vacancy (Herriot, 1989). Reed and Tsaur (2008) show that the type of the productivity in a sector affects its admission policies. They claim that in an economy where only cognitive skills are relevant for production, an entrance examination regime is efficient: in effect, admissions are exclusively determined by a person’s cognitive skills. However, if productivity also depends on non-cognitive skills, a generalised admission regime (meaning that the criteria for admission depend on an individual’s overall abilities) can weakly dominate the examination regime. However, a review of the research into employee selection by Ryan and Ployhart (2014) shows that selection is in a 'curious position'. The researchers claim that the basic question relating the predictivity of selection hypothesis remains unchanged. In addition, they argue that traditional selection research remains active and engaging, but other areas and direction should receive greater emphasis, such as globalisation (shift to a multicultural view) and technology. Researchers in selection should consider aspects such as psychometrics, scoring, validity, predicting a broader range of criteria, and investigating the role of context in selection. Other scholars suggest the need to expand the predictor measures to include outcomes such as health, relationship with the local community, and satisfaction. Such expansion should also consider the cultural context (Shemitt & Ott-Holand, 2012).

In general, personnel selection is a vital strategy for an organisation. It is a dynamic process and is affected by several factors (the mission of the organisation, the nature of the work, the measurements tool, the applicants’ needs, and so on). The policy makers in charge of creating a selection process should be aware of those factors. Fundamentally, the two main questions for any selection process (the ‘what’

and the ‘how’ questions) must be clearly answered and revised according to the quality of the outcomes.

3. 2 Non-cognitive attributes

As seen in the previous section, personnel selection relies on differences between individuals in terms of personal characteristics, which, in many studies, are divided into cognitive and non-cognitive attributes (Duckworth, 2009; Kell, 2018). This section reviews the literature relating to the two concepts, particularly non-cognitive attributes, in five points. The first two points illustrate the definition(s) and the importance of the attributes, respectively. Point three reviews the related literature on the stability of non-cognitive attributes and the influence of gender and culture. The last two subsections discuss the ‘what’ and ‘how’ questions related to the measuring of non-cognitive attributes in the selection process.

3. 2. 1 The definition(s)

There is a strong consensus in the literature around a definition of the term ‘cognitive’. In many studies, ‘cognitive’ is synonymous with intelligence and the ability to understand complex ideas and solve intellectual problems (Borghans, Duckworth, Heckman, & Ter Weel, 2008; Brunello & Schlotter, 2011; Kell, 2018). Defining the term ‘non-cognitive’, however, has proven more controversial. Different terminology is used in the literature, such as soft skills, personality traits, non-cognitive skills, attributes, non-cognitive abilities, character, and socioemotional skills. Even where scholars seek to provide a clear definition of the terminology in their own work, they rarely succeed without some ambiguity.

Heckman and Kautz (2012) use the term ‘personality traits’. They claim that that their choice has a sense of stability and the possibility of heritability, whereas ‘skills’ and ‘character’ can mostly be learned. In contrast, Kautz, Heckman, Diris, Ter Weel, and Borghans (2014) use the term ‘non-cognitive skills’. Gutman and Schoon (2013) use 'non-cognitive attributes' to refer to attitudes and behaviours such as motivation and self-control, whilst ‘cognitive domains’ refer to personal abilities in domains such as literacy and numeracy and are commonly measured by academic tests. The word ‘domain’ is also used by El-Baz and El-Sayegh (2015) to determine the attributes that are appropriate for the aims and nature of the job. In education, Garcia (2014) focuses on non-cognitive skills that can be learned in schools. Thus,

she defines non-cognitive skills as 'the patterns of thought, feelings and behaviour of individuals that may continue to develop throughout their lives, and that play some role in the education process' (p.26).

Integral to this argument about the different names and definitions is the challenge of distinguishing between the two concepts. Duckworth and Yeager (2015) believe that non-cognitive attributes are conceptually independent from cognitive attributes and that all the different terms relating to non-cognitive attributes refer to the same 'conceptual space'. However, Borghans et al. (2008) argue that even though cognitive and non-cognitive attributes can, conceptually, be distinct, this is an empirically challenging task because each has an influence on the other. Kell (2018) attempts to solve the dilemma by introducing the term 'cognition'. He argues that many working in the field of non-cognitive attributes do not have an understanding of 'cognitive skills' and how they are related to 'cognition'. He states that, 'Whilst cognitive skills constitute a variety of cognition not all cognition entails the higher-order, complex mental activity that defines cognitive skills' (p.25). Thus, he claims that replacing 'cognitive' and 'non-cognitive' with 'intellectual' and 'non-intellectual' would be beneficial.

Here, the researcher defines non-cognitive attributes as within-person variables, including all personality traits, beliefs, and dispositions such as motivation and enthusiasm.

3. 2. 2 The importance

Despite a lack of consensus on the definition, both attributes are important. Labour economics research indicates that cognitive and non-cognitive competencies play a crucial role in determining employee productivity. Even in academic and intellectual tests, examiner marks are likely to be affected not only by cognitive skills but also by motivation and personality (Gutman & Schoon, 2013; Harris & Sass, 2014). Heckman and Kautz (2012) provide evidence of the importance of personality for predicting and causing outcomes in economic and social life. In addition, Brunello and Schlotter (2011) state that many studies illustrate a strong consensus on the importance of non-cognitive attributes in both school attainment and labour market outcomes —considering these to be as important as the influence of cognitive skills. Chamorro-Premuzic and Furnham (2004) present a possible conceptual framework

for better understanding of the interface between cognitive and non-cognitive (intelligence-personality). They distinguish between three different levels of intelligence and assume that aspects of personality that determine performance in an intelligence test are essentially different from those that determine responses on personality measures. Therefore, they emphasise the need to include personality traits in the selection process for better prediction of applicants' future performance.

Despite the consensus on the relationship between the two concepts, there is no agreement on the best non-cognitive predictor of academic achievement. A review by Stankov (2013) reveals that many non-cognitive measures are poor predictors of intelligence. He identifies that measures of rationality, self-assessment of intelligence, openness to experience, and self-concept correlated by up to .35 with cognitive performance. The correlation reaches .45 between measures for self-belief (self-efficacy and anxiety) and achievement tests. However, the findings also assume that the best predictors of cognitive performance are measures of confidence. Similarly, Stankov, and Lee (2014) note that measures of maladjustment and motivation/goal orientation have the lowest correlations with achievement, whereas measures of confidence have the highest predictive validity. The other measures of self-belief are in the middle. In Duckworth, Peterson, Matthews, and Kelly (2007), 'grit' accounts for an average of 4% of the variance in success outcomes, including educational attainment and grade point average.

3. 2. 3 (In)stability: the influence of age, gender and culture

Some factors influence the development of non-cognitive attributes, such as genetics, nurturing, early childhood education, and health (Garcia, 2014). Here, the role of age, gender, and culture in the stability of non-cognitive attributes is discussed, as these factors are related to the context of the study.

Research states that cognitive and non-cognitive attributes can change as an individual age, in varying ways and to different degrees (Heckman & Kautz, 2012). Cognitive abilities, for example, tend to increase sharply during childhood, reaching the highest levels in late youth, and then decrease slowly. In contrast, some personality traits, such as conscientiousness, grow gradually from childhood to late adulthood (Borghans et al., 2008). Brunello and Schlotter (2011) assume that some non-cognitive attributes can be altered up until the end of the teenage years, whilst

others can continue to change throughout one's life. In their meta-analysis, Roberts, Walton and Viechtbauer (2006) study the degree of change in personality across different ages. The results show that individuals' performance increases in measures of social dominance (a facet of extraversion), conscientiousness, and emotional stability in young adulthood (ages 20 to 40). In contrast, performance on measures of social vitality (a second facet of extraversion) and openness increased in adolescence, but then fell in old age for both constructs. In addition, agreeableness changed only in old age. Heckman and Kautz (2012) present heritability studies that suggest personality traits tend to be about 40–60% heritable. This suggests that individual behaviour is tied more to the person than the situation.

Non-cognitive attributes are greatly affected by cultural factors, despite showing stability at particular life stages (Zhou, 2016). Borghans et al. (2008) suggest that contexts and incentives affect personality traits. The influence of culture over age was tested by Bleidorn et al. (2013), who conducted a cross-cultural test for a sample of young adults from 62 nations (N = 884,328). They found strong evidence that personality developed from early to middle adulthood, with cultural differences having a significant effect. The results reveal that cultures which commence adult-role responsibilities earlier than others were also marked by earlier personality maturation. Moreover, the majority of young adults in most cultures show similar age trends in personality. This is explained by the role transitions for young adults being similar across cultures at the same ages.

Finally, evidence-based studies indicate a significant difference between males and females in terms of personality traits. In four meta-analyses, Feingold (1994) identifies that males are more confident and have slightly higher self-esteem than females. Women have higher levels of anxiety, extraversion, and trust. There were no significant gender differences in terms of social anxiety, impulsiveness, activity, ideas (reflectiveness), locus of control, and orderliness. In addition, Schmitt et al. (2016) note that the trait differences are generally affected by gender roles and gender equity in each culture. They suggest that gender differences in most aspects of personality (the Big Five traits, self-esteem, depression, values, and so on) are larger in cultures with more equal gender roles and socio-political gender equity.

3. 2. 4 Non-cognitive attributes for selection: what?

In the previous sections, the literature on the naming, importance, and stability of non-cognitive attributes was reviewed. In spite of the arguments around definition and (in)stability, it is clearly noted in the literature that there is consensus on the importance of non-cognitive attributes for economic and social life. Therefore, the question which might come next is, which attribute(s) is more important to assess in the selection process? On this point, the researcher highlights the process of building a framework to identify the key non-cognitive attributes for the purpose of selection.

Surveying over 400 employers in four organisations in the US, Casner-Lotto and Barrington (2006) explored the skills required for new entrants to do better in the workplace. From the employers' perspective, they note that the most important skills are professionalism/work ethic (including personal accountability, effective work habits, working productively with others, and time and workload management), communications, teamwork/collaboration, and critical thinking/problem solving. The survey method was also used by Kim and Park (2013) to investigate the competencies required for training programmes of airline cabin crew members. The survey questions concerned general information and asked respondents to rate the importance of the competencies required of airline cabin crews, using a seven-point Likert-type scale. The results indicate a total of eight competency domains: appearance and attitude, physical fitness, customer-oriented skills and company loyalty, knowledge of foreign cultures and languages, emotional intelligence, skills for in-flight services, past work experience, and interpersonal skills.

In medicine, the works of Patterson et al. (2000) and (2013b) identify 11 key competencies in the selection criteria for doctors entering training as general practitioners (GPs). The competencies were elicited using interviews with stakeholders, critical incidents focus groups, behavioural observation, and a validated questionnaire. The results of the study (2013b) show that the most highly rated domains were empathy and perspective taking, communication skills, clinical knowledge and expertise, and professional integrity. In another study, Patterson, Ferguson, and Thomas (2008) used a series of job analyses to produce a competency model to select for postgraduate medical training in three secondary care specialties. Four job analysis methods were conducted, including observation, critical incidents focus groups, critical incidents interviews, and reviews of research literature. The

results were then applied to develop a validation questionnaire. The model comprised 14 general competency domains common to all three specialties. The findings indicate a wide range of attributes beyond clinical knowledge and academic achievement, such as empathy and sensitivity; communication skills; conceptual thinking and problem solving; organisation and planning; professional integrity; legal, ethical and political awareness; and coping with pressure.

In law, Shultz and Zedeck (2011) conducted research to enhance law school admission decisions using broader tests to assess applicants on both professional effectiveness and academic achievements. The key factors in lawyer effectiveness were found by interviewing groups of stakeholders and focus groups. The results were validated using a survey asking respondents to rate examples on a five-point Likert scale. The results give a list of 26 cognitive and non-cognitive factors, such as analysis and reasoning, fact-finding, organising and managing one's own work and that of others, passion and engagement, integrity/honesty, stress management, and community involvement and service. A similar triangulation methodology was used to build a competency domain model for engineering managers, including a review of related literature and interviews with academicians and practitioners. The importance of the competency domains and sub-domains was then tested using a survey (El-Baz & El-Sayegh, 2015). The findings illustrate that leadership and interpersonal competencies are the most important of the competencies in the developed model.

To sum, building a framework for the key non-cognitive attributes begins by collecting data from stakeholders. Agreement on the identity and importance of the attributes is a good way of overcoming the challenge of finding a consensus on the terminology. The data collection can be achieved using different methods, such as interview, observation, focus groups, and survey. It is also important to acknowledge the purpose and the context of building such frameworks, as these may influence the method(s) and the outcomes. Appendix 1 presents a summary of the main domains and attributes described in previous studies. The summary is used as one of the inputs when building phase one of this study (identifying the key non-cognitive attributes for ITEP selection).

3. 2. 5 Non-cognitive attributes for selection: how?

Once the key non-cognitive attributes have been identified, the next step is to ask how they can be measured. Currently, there is no systematic global measure for non-cognitive attributes (Zhou, 2016). However, the literature identifies some common measures, which are discussed here.

Literature shows that the 'Big Five' model, or five-factor model, is the most accepted framework for measuring personality traits (non-cognitive attributes). The origin of the model is built on the lexical hypothesis of Allport and Odber (1936), which assumes that the most important individual differences are determined in language. The researchers analysed personality-describing words in dictionaries. Later, and building on the work of several different psychologists, the personality traits were organised into five dimensions. These five factors have been known as the Big Five since Goldberg (1971), and they are as follows: openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism. The most significant criticism of the model is that it was derived from a factor analysis of test scores, rather than from predictive criteria in performance on real-world tasks (Borghans et al. 2008; Heckman & Kautz, 2012). However, several instruments have been developed to measure the Big Five dimensions, including the 44-item Big Five Inventory (BFI), the 60-item NEO Five-Factor Inventory, and the 10-item inventory (Gosling, Rentfrow, & Swann, 2003; McCrae & Costa, 2004; Rammstedt & John, 2007).

The Big Five model has been tested in studies of gender, age, and culture. Costa, Terracciano, and McCrae (2001), for example, analysed gender difference using data from 26 cultures (N = 23.031). The findings show that females report themselves as being higher in neuroticism, agreeableness, warmth, and openness to feelings, whereas males are higher in assertiveness and openness to ideas. The effect of age was studied by Donnellan and Lucas (2008), using two large datasets from the UK and Germany. Participants ranged in age from 16 to mid-80s. The results reveal that extraversion and openness are negatively associated with age, whereas agreeableness is positively associated with age. Jolijn et al. (2003) investigated the constant in the structure of the Five-Factor Personality Inventory (FFPI) across 10 European and three non-European countries. The five-factor structure is clear in all samples except in the smallest (USA, N = 97). Within each country, more than 80% of the items were equally stable.

In Oman, Kazem (2002) used an Arabic translation of the Big Five factors list of Costa and McCrae (60-item NEO), with a sample from Sultan Qaboos University students (N = 63). The factor analysis shows that the structure of the Big Five is not clear, demonstrating three bi-poled factors (conscientiousness/extraversion; agreeableness/openness; neuroticism/extraversion), and two uni-poled factors (neuroticism/agreeableness). He recommends retesting the model with a larger sample. In contrast, Salleh, Al-Kalbani and Mastor (2010) used the Adolescent Personal Style Inventory, a Big Five measure for adolescents, with a sample of high school students in Oman. The results suggest that the model is suitable for defining the personality structure of this population.

Research indicates a relationship between the Big Five dimensions and job performance. With a sample of employees of a pharmaceutical company (N = 159), Rothmann and Coetzer (2003) found that emotional stability, extraversion, openness to experience, and conscientiousness are related to task performance and creativity. In addition, emotional stability, openness to experience, and agreeableness explain 28% of the variance in participants' management performance. In a meta-analysis, Barrick and Mount (1991) note that conscientiousness is consistently related to the job performance criteria of all occupational groups. Zhou (2016) suggests not using the Big Five scales to measure the non-cognitive skills that can be developed by training or education, due to the stability of the five dimensions. However, a study by Lakhal, Sévigny, and Frenette (2015), with a sample of students enrolled in two compulsory undergraduate business courses (N=165), indicates that the Big Five factors explain 6-13% of the variability in performance on group work, oral exams, written exams, multiple choice tests, and practical work.

Despite the wide acceptance of the Big Five model for measuring personality, interviews are the most commonly used selection tool (Ryan & Ployhart, 2014). The construct validity of personnel selection interviews was explored by Salgado and Moscoso (2002) through a series of meta-analyses. They divided interviews into two groups: conventional interviews (checking qualifications, experience, and self-evaluation information) and behaviour interviews (focus on job knowledge). The results show that the measured constructs differ according to the type of interview being used.

Furthermore, multiple mini interviews (MMIs) are internationally used for selecting students in healthcare training programmes, with applicants asked to respond to scenarios at a series of 'stations' at certain times. Each scenario is designed to assess specific values or attributes. A systematic review of the empirical research in the domains assessed by the MMIs shows that 32 personal domains were assessed. The most frequent domains were communication skills, teamwork/collaboration, and ethical/moral judgment (Callwood et al., 2018). In an empirical study, Lemay, Lockyer, Collin, and Brownell (2007) found that the MMIs used for applicants to medical school were able to assess different non-cognitive attributes and offered a fairer and more defensible assessment than the traditional interview.

Previous studies indicate that there are different types of interview used to assess different constructs, which can be costly, in terms of time and resources, to develop and implement. Furthermore, a recent meta-analysis shows that the length of an interview is unrelated to reliability validity (Thorsteinson, 2018). However, Makransky, Havmose, Vang, Andersen, and Nielsen (2017) evaluated the predictive validity of the admissions procedure, including a cognitive ability test followed by MMIs, with a sample of students at the University of Southern Denmark. The results show that despite the high cost of using an MMI, the rewards in terms of lower drop-out rates and higher levels of academic achievement were likely to outweigh the costs.

An assessment centre (AC) is another measurement tool used in the selection process. In a typical design, the applicant is presented with multiple exercises, allowing judgments to be made about his behaviour in different dimensions, including communication, problem-solving, planning and organising, and so on (Gibbons & Rupp, 2009). The criterion-related validity of ACs was investigated by Arthur et al. (2003), using meta-analytic procedures. The results show a range of estimated true criterion-related validities from .25 to .39. In his review, Lievens (2017) states that assessment centre exercises and SJTs enable better understanding of the key research questions related to variability among people and the link between traits and behaviour.

On the basis of meta-analytic findings, Schmidt and Hunter (1998) identify the validity of 19 selection procedures for predicting job and training performance. They note different levels of validity for predicting future job performance among the methods and combinations of methods. Individual methods varied in their validity,

from no validity (graphology) to high validity (for example, general mental ability tests and work sample measures). The combinations method had the highest overall multivariate validity and utility for job performance.

Salvatori (2001) reviewed the use of various selection tools in the health professions literature. Overall, grade point average (GPA) is the best predictor of academic performance, though it has a weaker relationship with clinical performance. Admission test results are an accurate predictor of performance in some findings, but not in others. The value of personal interviews and written submissions as selection tools is, still, unclear. The selection methods used by medical schools were reviewed by Patterson et al. (2016a), using studies published between 1997 and 2015. They found eight selection methods, namely: aptitude tests, academic records, personal statements, references, SJTs, personality and emotional intelligence assessments, interviews and MMIs, and selection centres (SCs). Academic records, MMIs, aptitude tests, SJTs, and SCs are found to be the more effective methods. In another study related to selection for UK general practice, Patterson, Lievens, Kerrin, Munro, and Irish (2013a) found that a combination of a clinical problem-solving testing, an SJT, and a selection centre were the best predictors of work performance and training outcomes. Finally, a systemic review by Nielsen and Friderichsen (2017), which explored the alternative admission criteria used in selection for higher education, reveals that measures of non-cognitive skills — namely self-efficacy and admission interviews — were able to predict academic performance at university.

Despite the variety in selection measurements, the literature shows that there are challenges associated with every measure. Duckworth and Yeager (2015) argue that all measures have pros and cons and there is no sense in trying to rank them from best to worst. They suggest that rather than trying to seek out the ‘most valid measure’, it is better to seek out the ‘most valid measure for the intended purpose’. They recommended using, where possible, a multi-method approach which can increase reliability and validity. In addition, Ryan and Ployhart (2014) state that part of the problem is that selection researchers make mistakes when comparing research on a construct (for example, conscientiousness) with research on a method (such as interview). They suggest paying greater attention to multi-construct methods, particularly interviews, ACs, and SJTs.

Here, the researcher reviews some of the common measures used to assess the non-cognitive attributes of the applicants during the personnel selection process. The discussion about SJTs is covered separately in Section 4.3, and the next section highlights the selection process at ITEPs.

3. 3 Selection for initial teacher education programmes (ITEPs)

In the previous sections, the literature on personnel selection and non-cognitive attributes were reviewed. Here, the researcher discusses the selection for ITEPs in five subsections. First, a general overview of ITEPs is given. Then, some perspectives on the rationale for the selection process of the ITEP are discussed. The third point highlights the relationship between selection for the ITEP and teacher effectiveness. The last two points concern the 'what' and the 'how' questions used to assess the non-cognitive attributes during selection for the ITEPs.

3. 3. 1 Initial teacher education programmes (ITEPs): a general overview

Teacher education is one of the key policy tools for providing future teachers with knowledge, skills, and dispositions. There is widespread professional agreement that it is positively related to teaching quality and pupil outcomes (Menter et al., 2010; Tatto, 2008). However, there are different types of teacher education programmes and these vary according to different aspects, such as structure, requirements, duration, and mission. This section gives an overview of the meaning and the different types.

ITEPs are variously described as 'general secondary education, general higher education, specialist higher education in a particular subject, professional courses relevant to teaching, and supervised teaching experience in schools' (Eraut, 2000, p.453). The design of an ITEP can be determined by state government, local governments, or by the programme providers (universities, colleges) (Eraut, 2000). Hobson et al. (2008) use the term 'initial teacher preparation' (ITP) rather than initial teacher training (ITT), or initial teacher education (ITE). They claim that the word 'training' underestimates the view of teaching as intelligent consciousness, and that the word 'education' is more closely connected to declarative knowledge than procedural knowledge. In addition, they object to the use of the term 'pre-service training/education' because some students might already serve in the teaching profession.

Despite previous objections to the terms ‘training’ and ‘education’, both are operationally used. Through comparative analysis of ITEPs in England and Norway, Stephens, Egil tønnessen, and Kyriacou (2004) found that initial teacher training in England is a ‘training model’ that aims to give trainee teachers necessary teaching skills — such as classroom instruction, managing students’ activities, designing homework, and providing pupils with a secure learning environment. This contrasts with initial teacher education in Norway, which is an ‘educative model’ which helps trainees to reflect on the practical implications of educational theories, instructing students in subjects, performing leadership skills, acting as a member of a caring profession, promoting Norwegian values, and providing pupils with a safe learning environment.

Across the world, ITEPs have different structures, but they can generally be classified into two categories: (a) the ‘consecutive model’, which involves students completing degree-level study in a particular subject before enrolling in an ITEP and requires undergraduate study at the university level; and (b) the ‘concurrent model’, which combines the study of a particular subject with teacher education and training and requires successful attainment of secondary school qualifications (Hobson, Ashby, McIntyre, & Malderez, 2010). The duration of ITEPs varies as a consequence of these two models. Most countries have 3-4-year undergraduate concurrent programmes, and 1-2-year postgraduate or consecutive programmes. Finland is assumed to be the sole exception, with prospective teachers undertaking a five-year master's degree programme (Conway, Murphy, Rath, & Hall, 2009).

The different forms of ITEP can be seen as an outcome of the diversity in several factors, such as the teacher education missions at the national/state level (Kennedy, 2015), the providers (Beauchamp, Clarke, Hulme, & Murray, 2015), and the influence of the stakeholders (for example, politicians, schools, academics) (Franchi, 2016). Recent research argues that teachers’ knowledge and practice, and presumably also pupil learning, might be influenced by the education programmes on which teachers were enrolled (Tatto, 2008). For example, Henry et al. (2014) found that members of ‘Teach For America’, a teacher preparation programme in the US, are more effective than traditionally prepared teachers, and out of-state traditionally prepared teachers are less effective than in-state traditionally prepared teachers.

Finally, despite the importance of ITEPs for teachers' preparation, studies raise concerns about their quality. In one longitudinal study, participants from six schools in Portugal negatively evaluated their initial preparation (Maria, 2001). The participants in Fantilli and McDougall (2009) argued that their pre-service learning was limited and insufficient to meet their needs in their initial years in the profession. The challenge for most ITEPs is not only to prepare teachers for life in the classroom, but also to develop teaching as a professional learning community (Conway et al., 2009). In Biermann, Karbach, Spinath, and Brünken (2015), the quality of the field experiences during the teacher education programmes in Germany was tested. The results indicate that both personality and features of the field experience are correlated with teaching skills. Conversely, Mohamed, Valcke, and De Wever (2017) argue that the field experience used in ITEPs in the United Arab Emirates is not effective because it focuses mainly on activities inside the classroom, with teaching activities outside the classroom receiving little attention. Dolan (2012) states that it is widely accepted today that ITEPs are insufficient for the lifelong professional needs of teachers and, as a result, recent educational policies in countries around the world have focused on lifelong learning.

Despite the criticisms of the quality of ITEP outcomes, reviews of the research into teacher education conclude that prepared teachers are generally better rated and more successful with students than teachers without preparation. Teachers who have more preparation are seen as more confident and productive with pupils than those with little preparation (Darling-Hammond, 2000).

3. 3. 2 The rationale for selection into initial teacher education programmes (ITEPs)

Darmody and Smyth (2016) offer two reasons for using selection procedures for ITEP admission. Firstly, it helps to identify the applicants who are most likely to succeed during the training programme and who will become good teachers. Secondly, the process helps to regulate the numbers of the applicants to the available number of places. In some cases, teacher education programmes receive many more applications than they can accept (Casey & Childs, 2007). In Ireland, for example, in 2008, there were 2,455 applicants for 800 places on the various postgraduate teacher education courses (Harford, 2010). Therefore, with increasing numbers of applicants, a screening process is necessary to choose the best participants for the preparation

programmes and to screen those applicants who may be unsuitable. In other circumstances, the teaching profession is not very attractive, or there is an over-supply of teachers at some levels and in some subjects. Countries with an over-supply of candidates normally use more rigorous selection processes (Darmody & Smyth, 2016). However, poor quality selection procedures can lead to high personal and institutional financial costs. Bowles, Hattie, Dinham, Scull, and Clinton (2014) note that more than 30% of students entering university in Australia do not complete their course and a further 30% do not remain in the profession for more than 3-5 years.

Hobson et al. (2010) analyse different points of view on the application of a selection procedure. Some criticise the selection process, saying that the selection process seems to ignore the complexities and nature of teaching by testing a set of attributes, which can drive away potentially good teachers. In addition, even in countries which use selection approaches, some candidates are more likely to withdraw during or after the training programmes. Other studies argue that the theoretical effects of such requirements on teacher quality are ambiguous (Angrist & Guryan, 2008). Conversely, there is research indicating that the selection process is important because it sets a minimum standard for the knowledge that individuals must have and discourages candidates who have unrealistic expectations or lack commitment to teacher education programmes and/or teaching as a career (Hobson et al., 2010).

Another argument is that the evaluation process used during the selection of new entrants to the ITEPs, especially that concerning teaching skills, raises questions about experience. Specifically, there is a question about the value in conducting such a process with applicants who have no previous teaching experience. Firstly, although the research suggests that teaching experience is associated with teacher effectiveness (Stronge, 2007), the relationship between the two is not linear. The findings of different studies indicate that effectiveness increases for teachers during the first 3-5 years, but then declines (Henry et al., 2014). Day, Sammons and Stobart (2007) suggested that the influence of experience on teacher's professional life dependes on other factors such as self-efficacy and resilence. Simillarly, findings from Kington, Reed, and Sammons (2014) suggested 'that lack of experience is not seen as critical to teacher effectiveness' (p. 550). In addition, Kyriacou (2007) states that teacher effectiveness depends on their ability and motivation after many years of

experience. Besides that, the limited timeframe for preparing future teachers in educational institutes may allow the development of pedagogical knowledge, but they cannot transform the beliefs and attitudes of prospective teachers unless they already have those attributes (Jacobowitz, 1994). Therefore, the selection process can help to identify, at early stages, applicants with high levels of effectiveness. In addition, Casey and Childs (2007) state that pupils in the classroom need an effective teacher, regardless of his/her experience. Finally, applicants to ITEPs were themselves once school pupils, thus they have some experience of the profession's requirements.

In practice, the selection for ITEPs is, undoubtedly, exist. The previous debate might be seen as being more about the outcomes of the selection process. The question about the relationship between selection and teacher effectiveness is crucial to understanding the rationality of the selection process. In other words, what is the evidence that a good selection process will lead to a better teaching performance? The next section reviews the research concerning this issue.

3. 3. 3 Selection and teacher effectiveness

Firstly, theoretical and empirical studies show that good quality teachers have a significant and positive effect on student performance. The characteristics of the teacher have been shown to account for approximately 30% of student achievement (Hattie, 2009). A study by Kim, Dar-Nimrod, and MacCann (2017) suggests that teacher personality (using Big Five personality domains) is more important for student socio-emotional outcomes than academic outcomes. Furthermore, Kunter et al. (2013b) found that teachers' pedagogical content knowledge, enthusiasm for teaching, and self-regulatory skills had a positive effect on instructional quality and, therefore, affected student outcomes.

In terms of the importance of teacher effectiveness, the question is asked: can good selection produce a good teacher? A simple question with no simple answer. There remains little evidence of a relationship between the selection of applicants for the ITEP and future outcomes in the profession. Darmody and Smyth (2016) refer this to the complicity on separating selection effects from a range of institutional and social effects into teacher performance. However, it is important to understand the ability of the selection process to predict the quality of future teachers.

One aspect of the challenge is constructing a clear definition of teacher effectiveness. The term 'teacher effectiveness' is discussed extensively in the literature, but there is little agreement about its meaning (Stronge, 2007). The House of Commons Education Committee (2012) states that defining 'teacher quality' is complex because of the contribution of many different factors. Furthermore, 'effective teaching' is identified in the literature under different names, such as effective teaching, creative teaching, veteran teachers, quality teachers, and good enough teachers (Casey & Childs, 2007). The definitions also differ according to stakeholder perspectives. In Schumacher, Grigsby, and Vesey (2015), the definition is seen from the perspective of working teachers, whilst Strikwerda-Brown, Oliver, Hodgson, Palmer, and Watts (2008) define an 'effective teacher' from the perspective of the students. Day, Sammons and Stobart (2007) included two definitions of effectiveness: teachers' perceptions of their effectiveness and students achievement.

To overcome the challenge of finding a single definition, the research instead considers the characteristics that shape an effective teacher. One review of evidence-based research identifies six components of great teaching: pedagogical content knowledge, quality of instructions, classroom climate, classroom management, teacher beliefs, and professional behaviours (Coe, Aloisi, Higgins, & Major, 2014). Kane, McCaffrey, Miller, and Staiger (2013) define effective teachers as having three characteristics, namely: sensitivity to students' needs, knowledge of subject-matter content and pedagogy, and the ability to put that knowledge into practice. The Classroom Assessment Scoring System (CLASS), in Pianta and Hamre (2009), assesses classroom quality in three domains: emotional support, classroom organisation, and instructional support. By using the term 'inspiring teacher', Sammons, Lindorff, Ortega, and Kington (2016) found seven components that distinguish inspiring practice, namely: positive student-teacher relationship, good management, positive and supportive climate, formative feedback, high quality learning experiences, enjoyment, and student engagement and motivation. Furthermore, Kyriacou (2007) notes three main teaching skills: knowledge (of the subject, pupils, curriculum, teaching methods); decision-making before, during and after a lesson for better educational outcomes; and action taken to aid pupil learning. Appendix 2 summarises the attributes that related to teacher effectiveness found in the literature. In general, most research on effective teachers/teaching candidates focus on

three dimensions: the teacher's knowledge (of the subject and more), skills (teaching and non-teaching), and behaviours (explicit and implicit).

To reiterate, evidence-based research on the influence of the selection process on prospective teachers and their effectiveness in the profession is scarce. Casey and Childs (2007) state that such studies should follow candidates from the point at which they enter the ITEP, through at least the first five years of teaching. In their 2011 study, they investigated the relationship between two admission criteria (GPA and a written profile) and the readiness of teacher candidates in mathematics. They found no significant relationship between either the assessment of practice teaching or readiness. GPA predicted only 5-12% of the variance in course instructors' judgments of teachers' preparedness. Heinz (2013) explores the rationale behind various selection criteria for ITEPs in Ireland, finding a lack of evidence for the predictive value of previous academic achievement in the academic and practical components of the programme.

To expand our search on the relationship between teacher effectiveness and the selection process, the researcher looked at studies evaluating the selection process for recruitment into the profession. Goldhaber, Grout, and Huntington-Klein (2014) evaluated the selection tools used by Spokane Public Schools (SPS). They found that the screening instruments predict teacher value-added in student achievement and teacher attrition, but not teacher absence. In addition, Angrist and Guryan (2008) found that standardised tests (Praxis) used to recruit teachers to most states' public schools in the US were related to increases in teacher wages, but there was no evidence of improvement to teacher quality.

In summary, previous studies provide little evidence of a relationship between selection process and teacher effectiveness. However, taking a theoretical perspective, Kunter et al. (2013b) distinguish between three approaches in the research on teacher quality which might be related to the selection process. The first approach assumes that good teachers show stable cognitive characteristics, hence the recruitment and selection process is crucial. The second argument focuses on the profession-specific knowledge developed during teacher education programmes. The concept of professional competence is the third approach, and this explains differences in teacher performance. It focuses on the importance of profession-specific teacher attributes — such as knowledge, beliefs, motivation, and self-regulation — which are key aspects

that determine teachers' success. In addition, Klassen and Kim (2017a) develop a teacher selection model to better understand the relationship between teacher selection and teacher's effectiveness, as seen in Figure 3.1. The model shows that most studies in teacher selection fail to understand the relationship because they focus on correlations between measures (arrow 4), as it is difficult to directly assess the relationship between the selection measure and latent teacher effectiveness (arrow 5). However, the relationship can be explained by a series of inferences (arrows 1, 2, and 3), based on theoretical and empirical relationships.

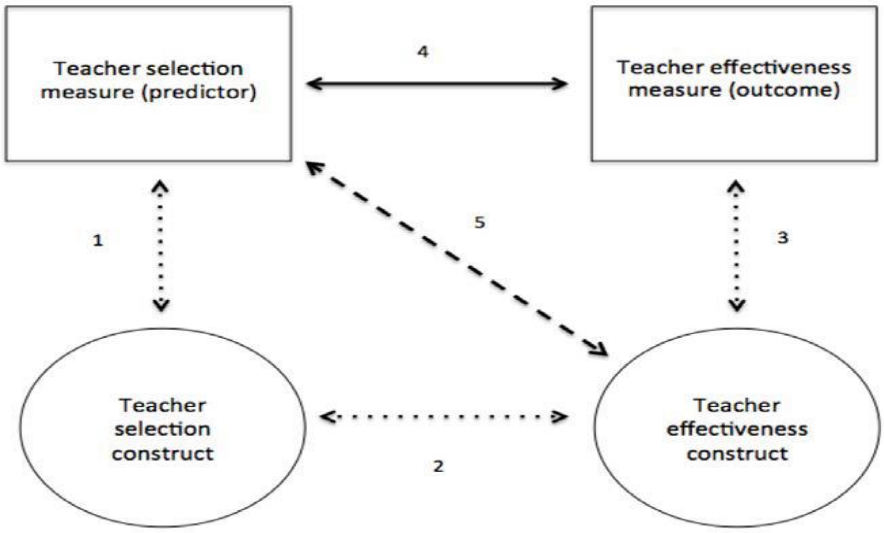


Figure 3.1 Teacher selection model by Klassen & Kim (2017a) (adapted from Binning & Barrett, 1989).

To conclude, although the relationship between the selection process and teacher effectiveness requires more evidence, it can be assumed that a bad selection procedure (or no selection) cannot produce good teachers. Teacher effectiveness is a result of a series of consequences, one of which is the ITEP selection process. To ensure an effective selection process, it is vital to first answer two fundamental questions: what to measure, and how to do it? The next two sections will review the literature in these two areas.

3. 3. 4 The Selection for initial teacher education programmes (ITEPs): what?

The main aim when selecting candidates for ITEPs is ensuring the best possible future teachers in the profession. Thus, understanding the key attributes that relate to teacher effectiveness, especially in terms of non-cognitive attributes, is important to know what to assess at the selection process.

In education, research into specific non-cognitive attributes indicates that they are potential predictors of success. Confidence, for example, is seen as the strongest non-cognitive predictor of academic achievement (McGeown et al., 2015). Moreover, grittier teachers outperformed their less gritty colleagues (Robertson-Kraft & Duckworth, 2014). A meta-analysis of self-efficacy beliefs found that pre-service and

in-service teachers' self-efficacy beliefs influence their commitment to the teaching profession (Chesnut & Burley, 2015). Teachers' enthusiasm, as a non-cognitive attribute, positively influences students' interest (Keller, Goetz, Becker, Morger, & Hensley, 2014) and has a positive effect on pupils' motivation (Kunter et al., 2013b). Furthermore, caring is found to be more important factor in teaching than in any other occupation, except nursing (Harris & Sass, 2014). Moreover, the results from Teaching and Learning International Survey (TALIS) indicate that teacher-student relations have a significant impact on teachers' job satisfaction (OECD, 2014). Although non-cognitive attributes play an important role in teacher effectiveness, there is no single non-cognitive attribute that can be named as the sole predictor of these outcomes (McGeown et al., 2015). Kyriacou (2007) states that, owing to the nature of teaching, it is difficult to devise a list of general skills on which to focus. However, he confirms the need for such a list in order to help teachers to develop their classroom practice. Such a set of skills is also important for different educational aspects, such as training, selection, and evaluation.

Several studies seek to determine the key non-cognitive attributes which influence teacher effectiveness. The interview and survey responses on effective teaching in numerous studies list characteristics and behaviours such as caring, listening, understanding, knowing students, fairness and respect, social interactions with students, promotion of enthusiasm and motivation for learning, and attitude towards the teaching profession (Stronge, 2007).

A study exploring perceptions of teaching soft skills in Taiwan identifies six factors: positive attitudes, open-mindedness, interpersonal relationships, teamwork, communication skills, and creativity (Lee & Lee, 2011). In Europe, the Education and Training 2020 Strategy identifies reflective practice, ongoing learning, engagement in research and innovation, collaboration, and commitment to school development as the minimum requirements of teachers, beyond pedagogical skills (Caena, 2014). Furthermore, Mohamed et al. (2017) suggest a framework of teacher competencies concerning readiness-for-the-job. The competencies are clustered into six main domains, each representing teachers' roles: knowledge and instructional skills in teaching and learning; organisation/management skills; knowledge of diverse learners; effective collaboration with colleagues, parents, social services and the community; attitude to professional development; and development of ethical stand.

Focusing on the key non-cognitive attributes in the ITEP selection process, Sautelle, Bowles, Hattie, and Arifin (2015) identify in the literature six psychological constructs of effective teaching, each of which can be assessed when selecting teachers for training programmes and which allow differentiation between candidates. These constructs are extraversion, agreeableness, conscientiousness, resilience, self-regulation, and cognitive ability. The graduate participants named enthusiasm for the subject, the ability to communicate, and the ability to work with others as criteria important for the selection process (Turner & Turner, 1997).

In practice, the selection process for Zurich University of Teacher Education in Switzerland, for example, measures five competencies: communication; cooperation, in terms of 'awareness of others'; assertiveness, including convincing others; motivation; and fact-finding (Bieri & Schuler, 2011). In a study of 48 ITEPs in Taiwan, eight criteria were found: academic ability, character and moral conduct, written expression, general educational knowledge, values and attitudes toward education, motivations and enthusiasm for teaching, psychological aptitude and personality traits, and social and interpersonal skills (Wang & Fwu, 2007). The Malaysian Educators Selection Inventory (MEdSI) test and interviews screen trainee teachers, assessing personality, integrity, career interest, and emotional intelligence (Hashim, Damio, & Hussin, 2013).

As a backdrop to the above, there are many attributes related to teacher effectiveness that it might be necessary to measure during the ITEP selection process. However, 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 successful applicants must already have those necessary attributes that cannot be learned on the programme. According to their proposed model, the minimum requirements for admission can be determined by subtracting the preparation provided by the programme from the minimum requirements for a good beginning teacher. Although the model provides a useful framework, it has several weaknesses, including a lack of consensus of the requirements for a good beginning teacher and other operational decisions. However, despite the difficulties with the implementation of such a framework, it is crucial to begin by clarifying which non-cognitive attributes can be used in the selection process for applicants with no teaching experience.

Identifying the key non-cognitive attributes to be assessed during the ITEP selection process is important, but it is not easy. Those frameworks built for other professions, as explained in Section 3.2.4, can be adapted to the teaching context. Furthermore, it is important to, first, determine the general features of the necessary attributes. In general, it is felt that the targeted non-cognitive attributes should have three features: (a) they should be seen as a priority for novice teachers (with no teaching experience) for solving teaching incidents, (b) they should be difficult to teach or develop effectively through ITEPs, and (c) they should reflect the implicit attributes of the applicants (rather than their personal appearance). Appendix 3 offers a summary of the key non-cognitive attributes assessed during the selection of candidates for ITEPs in a number of countries. The methods used to measure applicants are reviewed and discussed in the following section.

3. 3. 5 The Selection for initial teacher education programmes (ITEPs): how?

This section reviews the methods used to measure non-cognitive attributes during the ITEP selection process. Firstly, it is important to note that the variety of teacher education programmes has resulted in varied selection policies. One report on policies for recruitment and selection of students for teacher education programmes found that countries varied significantly. In some countries (for example, China), the criteria vary according to the age group which applicants intend to teach. In some countries, entrants must pass a national examination (for example, Malaysia, Singapore and China); whereas in others, each institute has its own exam (France, New Zealand, and Canada) (Hobson et al., 2010).

Ingvarson et al. (2013) gathered data from 750 programmes in 500 teacher education institutions in 17 countries (including Oman). They found that the most common basis for selecting students was the applicant's general academic achievement in the final year of secondary schooling. Similarly, Darmody and Smyth (2016) selected eight case studies (Australia (New South Wales), Austria, Canada (Ontario), Finland, the Netherlands, Scotland, Spain, and Sweden) and examined the differences in the level of demand for teacher education programmes. The findings show that all of the sampled countries use secondary school qualifications to select students, and many also used other measurements, such as interviews and admission exams. In addition, a review of the measures used throughout North America found that GPA is the most widely used measure to assess academic ability, owing to its

availability. A written profile is the second most widely method used. This asks the applicant to answer specific questions about the relevant experience and interest in teaching. Interviews are also used to gather information about applicants. Other measures include letters of reference and standardised tests (Casey & Childs, 2007).

In another review, a survey of 74 university-based ITEPs in the UK indicates that all programmes measure academic attributes using university academic transcripts, whereas all non-academic attributes are usually evaluated using a combination of individual and group interviews (97%) and evaluation of behaviour during group activities (62%) (Klassen & Dolan, 2015, in Klassen et al., 2017b).

In some well performing education systems, such as Finland and Singapore, the countries use multiphased selection measures. Selection in Singapore involves three steps before acceptance onto teacher training programmes. Firstly, the applicant presents a CV to indicate his/her academic qualifications, and then takes an admission test in literacy. Finally, successful applicants are interviewed to evaluate their attitude, aptitude, and personality. Finland also has a three-step selection procedure. In the beginning, the applicants take a national admission test to measure their literacy, numeracy, and problem-solving skills. A university assessment test is then taken in order to measure abilities in processing information, thinking critically, and synthesising data. Finally, successful applicants are interviewed to check their motivation to teach, motivation to learn, communication skills, and emotional intelligence (OMoE, 2012).

In Malaysia, different sets of criteria are used by different authorities, and the prioritisation given to each criterion differs between these authorities. Generally, the teacher-candidate selection process comprises the following stages. First, applicants are filtered according to their academic achievement. The candidates then sit the MEdSI as an entry examination, which evaluates the applicants on intrinsic qualities such as personality, interest in a teaching career, integrity, and emotional intelligence. Finally, applicants are interviewed (Mat Kasim et al., 2012; Othman et al., 2008; Ramli et al., 2013). Selection based on assessment centre principles is used at the Zurich University of Teacher Education, which allows participants to demonstrate competencies related to the role for which they are applying (Bieri & Schuler, 2011).

The decision about which tool to use depends on the construct being measured. Wang and Fwu (2007) note that each construct is assessed using a variety of measures and most ITEPs use more than one means of evaluating a single criterion. For 'academic ability', most ITEPs use applicants' academic records or subject-related tests. For 'character and moral conduct', applicants are examined through their official records and recommendations letters. 'Oral expressions' are evaluated using interviews, public speeches, and other classroom situations. For 'written expressions', applicants' statements and other language tests are reviewed. In terms of 'general educational knowledge', ITEPs administer written exams on educational issues, practices, and theories, whereas for 'attitudes and motivations', most programmes use a combination of personal interviews, autobiographical statements, and recommendations. For 'psychological aptitude', standardised tests, or personal interviews and recommendations are used. Finally, for 'social/interpersonal skills', some ITEPs ask for records of community service and leadership activities, whilst others consider written statements and recommendations.

Despite the variety in the selection procedures, a recent meta-analysis examining the relationship between teacher selection methods and outcomes in 27 studies shows a weak effect size (r = .12) (Klassen & Kim, 2017c). This effect size is weak for both cognitive and non-cognitive admission tests. Caskey et al. (2001) note the need to continue the search for the most effective admission selection procedures. In their review, they argue that admission processes for education programmes should be concerned with important issues, such as the need for extensive work on moral questions, applicants' interests and specific demands, and the resources available for decision-making. Furthermore, Heckman (2000) highlights faults in current policies around education and job training around the world which exclude social adaptability and motivation.

Studies suggest that, rather than focusing on certain selection measures, there should be a comprehensive model for selection. Bowles et al. (2014) propose a model for identifying teaching candidates, comprised of three phases: the application phase, the assessment phase, and the structured behavioural interview. In the application phase, applicants are asked to answer questions related to previous teaching experience, their justification for applying, their educational achievements, and their prior work experience. In the assessment phase, the personal attributes and

capabilities are assessed (general cognitive ability, measures of personality, self-regulation, resilience, social interaction, cultural sensitivity, and self-awareness). In the final phase, a modified selection centre approach is taken, with behavioural interviews. Likewise, a six-stage model for ITEP selection is proposed by Klassen and Kim (2017a). Stage one sees the identification of the critical attributes (cognitive and non-cognitive) that will be targeted and included in the selection process. The academic records required for the programme are checked in stage two. The authors suggest screening the applicants at stage three, using assessments of literacy and numeracy skills, non-cognitive attributes measures (SJTs), and any other cognitive ability assessments. The screened applicants then go to evidence-supported methods at stage four, including (a) simulated teaching practice, (b) structured individual interviews or multiple mini-interviews, and (c) SJTs (if not used in stage three). Applicants' scores at stage four are then used to select the successful candidates. Finally, the components at the previous stages are evaluated and linked to the students' performance during and after the ITE programme.

As has been noted across and within countries, a variety of selection methods are used for ITEP selection. Appendix 3 summarises the selection methods and theoretical models that are used in ITEP selection. In this chapter, the researcher identified SJTs as a measurement used for the selection processes in other professions. The ability of the SJT to assess ITEP applicants has not been extensively studied. The next section will review the existing research on SJTs and outline how this measure could be developed and implemented for the ITEP selection.

3. 4 Situational judgment tests (SJTs)

Kaplan and Saccuzzo (2009) define a psychological test as 'a set of items that are designed to measure characteristics of human beings that pertain to behaviour' (p.6). A simple definition but each of its components ('set of items', 'designed', 'to measure', 'human characteristics', and 'behaviour') has different meanings and interpretations in the literature. However, there is increasing interest in these tests from organisations and the business sector for use in selection, recruitment, and training processes. Although cognitively-oriented tests are usually preferred in these processes, there is a growing need for tests that widen the competencies tested and which can be administered to large groups of applicants. SJTs are believed to meet

these requirements and are increasingly popular in personnel selection (Lievens et al., 2008).

This section reviews literature in SJTs in ten sections. First, the definition of the SJT is explained. The second subsection gives an overview of the implications of the SJT throughout history and up to the present day. The theoretical basis is illustrated in the third point, whereas an outline about SJTs across cultures is given in point four. Point five presents the development process of the SJTs. Studies of subgroup differences, reliability, validity, and the applicants' reactions to the SJT are reviewed in points six, seven, eight, and nine, respectively. The last point highlights two main threats to the use of SJTs; namely, coaching and faking.

3. 4. 1 The definition of situational judgment tests (SJTs)

SJTs are simulation tests in which the applicant is presented with a variety of situations that he/she would be likely to meet on the job. They are seen as 'a predictor of performance', as they aim to measure judgment in work settings. Normally, SJTs consist of a set of situations and responses for each situation, with the test-taker asked to identify the appropriate response(s). The test-taker is asked to show his/her level of agreement with statements concerning work-related behaviours (Corstjens, Lievens, & Krumm, *in press*; Lievens et al., 2008; McDaniel et al., 2001; McDaniel & Nguyen, 2001; Weekley & Ployhart, 2006). Box (1) shows an example of an item used in an SJT, questioning how a prospective teacher could interact with the challenging behaviour of a pupil in the classroom.

As students in your classroom begin a writing task, one of them, Kata, starts throwing paper around and distracting the others. You know from previous incidents that Kata often becomes frustrated when she does not understand how to complete activities; she often displays this by being disruptive.

Would you....

a) Ask her to leave the class?
b) Show her how to get started on the task?
c) Encourage her by telling her that she is capable of completing the task?
d) Ask a passing teacher to talk to her?

Box (1). Example of an item used in an SJT (from Klassen, 2016)

Referring to the previous definition, research has raised questions about the nature of SJTs, such as how SJTs differ from other simulation tests, the extent to which SJTs are situational, whether they capture certain or multiple construct(s), and how judgment is made in the SJTs. In the following paragraphs, the related literature to these issues is reviewed.

The simulation tests used for selection comprise sets of tasks that present situations and ask participants to respond as though they were actually doing the job. The responses are interpreted as a potential indicator of applicants' future behaviour. Generally, simulation tests vary according to the way in which they are presented or what is known as the 'fidelity' of the test. Tests presented as an exact approximation of real job situations are 'high-fidelity simulation tests' (for example, ACs). Tests that consist of a simple written presentation of the tasks and responses are called 'low-fidelity tests' (Motowidlo, Dunnette, & Carter, 1990; Ployhart & MacKenzie, 2011). Historically, SJTs have been defined as low-fidelity simulations tests because they are typically presented in a written format and aim to assess context-dependent situations. However, recent research has developed SJTs in video format, measuring more general domains (Krumm et al., 2015). In addition, Patterson et al. (2015a) argue that SJTs have substantial advantages over other selection methods when measuring applicants' attributes. They claim that panel interviews, for example, have a lack of standardisation and are likely to be biased, whilst personality tests do not have high face validity. The practice in medicine shows that SJTs offer an objective standardised method for assessing a broad range of attributes for large numbers of applicants, show a good face validity to candidates, can be used for applicants who have no previous job experience, and have fewer group differences than other selection measures.

Secondly, research on SJTs has explored the extent to which SJTs depend on specific situations and how people judge these situations. Firstly, whilst most SJTs are quite similar, in that respondents are asked to make a judgment about a work-related situation (McDaniel et al., 2001), recent research shows it is possible to develop context-independent SJTs (generic SJTs) for use in different occupations (Motowidlo, Ghosh, Mendoza, Buchanan, & Lerma, 2016). This new SJT paradigm provides the opportunity to expand research across different jobs. Secondly, the 'judgment' approach focuses on assessing the effectiveness of the response options. However,

recent research expands on that perspective to better understand the judgment processes used by different participants. The results show that people differ in their understanding and interpretation of the situations, hence their different judgments. In other words, the results indicate that 'situational judgment' has more incremental validity than 'response judgment' (Rockstuhl, Ang, Ng, Lievens, & Van Dyne, 2015).

The third issue concerns what SJTs really measure. Although some studies claim that SJTs measure specific constructs or groups of constructs, they are more often seen as measuring multiple constructs, and they are difficult to isolate from general cognitive ability. Most studies correlate SJT scores with measures of cognitive ability or personality (for example, the Big Five). Others find that SJTs correlate with experience and job knowledge. Hence, the research claims that it is better to look at SJTs as a measurement method, such as interviews and ACs, which can be designed to measure a variety of cognitive and non-cognitive constructs, rather than a measure of a single construct (Lievens, 2006; Patterson et al., 2015b; Whetzel & McDaniel, 2009). However, despite the lack of evidence for the suitability of SJTs for measuring specific constructs, Guenole, Chernyshenko, & Weekly (2017) state that it is important to build such tests to allow for more accurate feedback in development settings, which could also improve our understanding of the construct validity.

The arguments around the construct(s) measured by SJTs, the validity, and the different forms of SJTs are discussed in the following sections. However, first, a clarification is given of the history and current implications of SJTs.

3. 4. 2 History and recent implications

Here, the history of SJTs and their current implications are highlighted. Research on the history of SJTs does not identify the first use of the term 'situational judgment tests', but the first use of the concept saw 'situations' and 'responses' presented in a written-format assessment.

SJTs go back to civil service and military examinations in the US in the 19th century. The first widely used version, containing response options, appeared during World War II and measured the judgment of soldiers. From the 1940s, a number of SJTs were developed, including the practical judgment test and supervisory practices test. In the late 1950s and early 1960s, organisations began using SJTs as part of the

selection process to improve performance and success rates. The test was criticised in the 1930s and 1940s for its low correlation with social characteristics, and it was declared to be more appropriate as test for general intelligence. However, SJTs were 'reintroduced' to applied psychologists through the works of Motowidlo and colleagues in 1990. There has since been a dramatic increase in research on SJTs, and they have been widely developed in different formats and applied in numerous sectors (Campion, Ployhart, & MacKenzie, 2014; Lievens et al., 2008; McDaniel et al., 2001; Weekley & Ployhart, 2006; Whetzel & McDaniel, 2009).

SJTs are currently used for selection purposes, especially screening applicants within the healthcare sector. In the UK, SJTs have been used alongside other measurements to select doctors for UK Foundation training since 2013, and for postgraduate training in public health, psychiatry, ophthalmology, and other fields. SJTs are also used in dental foundation training (DFT). Internationally, SJTs are used in medical school admissions in Belgium and Canada, and in postgraduate recruitment in Australia (Patterson et al., 2015b). In 2015, SJTs comprised approximately 50% of the assessment marks in the selection process for UK DFT (Affleck, Bowman, Wardman, Sinclair, & Adams, 2016). Box (2) shows an example of an SJT item used in postgraduate medical education, from Patterson et al. (2015b).

On the morning ward round, your registrar/specialty trainee said that Mrs Anderson is medically fit following her total knee replacement and could be discharged if Occupational Therapy feel it is appropriate. The occupational therapist has assessed Mrs Anderson and believes it is safe for her to go home with a care package that has been arranged. It is now 4 p.m. and the nurse informs you that Mrs Anderson is demanding to see a doctor, as she does not feel that she is ready to go home yet. An elective admission is waiting in the day room for Mrs Anderson's bed.
Rank in order the appropriateness of the following actions in response to this situation (1 = Most appropriate; 5 = Least appropriate).
A. Ask Mrs Anderson about her concerns.
B. Ask a senior colleague to speak with Mrs Anderson.
C. Ask the bed manager if he can find another bed for the elective patient.
D. Explain to Mrs Anderson that the bed has already been allocated and she has to go home.
E. Ask the occupational therapist to come and speak to Mrs Anderson with you.

Box (2). An example of SJT item for postgraduate medical education (Patterson et al., 2015b).

The use of SJTs in the education sector remains scarce. A project led by Robert Klassen at the University of York is currently exploring the use of SJTs for entry into ITEPs in different countries. The work began in the UK with a sample of practising teachers, teacher educators, and ITEP applicants. The work with the participants resulted in the development of an SJT targeting three non-cognitive domains; namely, organisation and planning, resilience and auditability, and empathy and communication. The SJT was piloted to candidates for a primary ITEP. The results show a near-normal distribution and good reliability for the participants' scores. The SJT also indicates significant positive correlations with the scores in an administered interview (Klassen et al., 2017b). The SJT developed in the UK has also been revised, developed, and piloted in Australia. The results indicate the benefit of a fourth targeted domain for applicants in the New South Wales (NSW) Department of Education in Australia; namely, culture and context. This new domain is considered necessary for selecting teachers who will be working in rural and remote settings (Durksen & Klassen, 2017). The project has been further conducted in Finland. Initial results show the importance of adding a fourth non-cognitive domain to meet Finnish needs for ITEP selection: namely, 'cooperation and fostering of community'. In addition, the findings from Finland indicate a relatively modest correlation and mostly positive applicant perceptions (Metsäpelto & Poikkeus, 2017).

3. 4. 3 Theoretical basis of the situational judgment test (SJT)

Motowidlo and colleagues (Motowidlo et al., 1990; Motowidlo, Hooper, & Jackson, 2006) propose a better understanding of the theoretical basis of the SJT, with two main perspectives. Firstly, the traditional perspective of the SJT, which is based on 'behavioural consistency theory'. According to this, past behaviour is the best predictor of future behaviour. Hence, the situations in the SJTs are important and should be strongly related to actual performance in the future job, as SJTs capture context-dependent knowledge.

The second perspective is that of general domain knowledge, which views SJTs as capturing relatively context-independent knowledge (rather than that developed from specific job experience). The theoretical basis for this perspective is that SJTs can be explained according to the 'implicit trait policy (ITP)'. ITP suggests there are differences between people in terms of implicit beliefs about the importance of personality traits for determining behavioural effectiveness. According to this

theory, an individual's effectiveness judgment can be related to the weight he/she gives to the trait in the specific situation, which reflects fundamental socialisation processes (parents, schooling, and so on) and personal dispositions. In other words, ITP assumes that applicants' responses to critical situations in SJTs can indirectly express their implicit traits (Lievens & Motowidlo, 2016; Corstjens et al., *in press*; Patterson et al., 2015b; Whetzel & McDaniel, 2009). This theory was tested by Oostrom, Born, Serlie, and van der Molen (2012), using multimedia SJTs to assess individual differences between participants in terms of leadership skills. The results confirm that SJTs are able to capture individual differences in implicit trait policies for extraversion and conscientiousness. Furthermore, leadership behaviour can be predicted more accurately by implicit trait policies for extraversion than for leadership experience and the associated personality trait. In another study, Kell, Rittmayer, Crook, and Motowidlo (2010) show that emotionally stable and conscientious actions are more effective in task situations, whereas open and agreeable actions are more effective in interpersonal situations. However, Patterson et al. (2012b) state that the extent to which ITPs change after some period of development (for example, after early adulthood) remains unknown.

Although these theoretical perspectives give explanations for why SJTs are used to predict work performance, they also give opportunities for more future theory-based research on SJTs (Lievens et al., 2008). Moreover, these perspectives should be better understood prior to the development of an SJT. Corstjens et al. (*in press*) recommend using the general domain perspective for entry-level selection, with context-specific SJTs being more useful when applicants already have work experience. More details of the development procedures used to build an SJT are presented in Section 5.3.4, but the next section highlights the use of SJTs across cultures.

3. 4. 4 Situational judgment tests (SJTs) across cultures

As noted before, this study is built on initial work conducted in the UK, extended to the context in Oman. Thus, it is important to understand how SJTs vary across cultures. This section reviews some of the related literature in this area.

Firstly, the term 'culture' has different meanings and comes with different terms, such as society, race, and ethnicity. In the Handbook of Cultural Psychiatry,

Tseng (2001) defines culture as 'the unique behaviour patterns and lifestyle shared by a group of people which distinguish it from other groups', noting that it 'is characterised by a set of views, beliefs, values, and attitudes toward things in life' (p.26). Schwartz (1999) states that the cultural values of a group of people influence the meaning of their work by representing, implicitly or explicitly, common thoughts about what is good, right, and desirable in a society. Cultural values (for example, success, justice, freedom, social order) work as norms that tell people what is appropriate in various situations.

To understand the differences in thinking and social action between different cultures, Hofsted (2001) collected data from more than 50 modern nations. He found that countries can be classified into five main dimensions: power distance, uncertainty avoidance, individualism/collectivism, masculinity/femininity, and long-term/short-term orientation. This framework was used in cross-cultural studies as a theoretical perspective to explain the differences in findings between different cultures. Even countries with similarities in geography and other life patterns also have differences in behaviour. In Kolman, Noorderhaven, Hofstede, and Dienes (2003), a survey of a sample of university students reveals important differences between value orientations in Western Europe (represented by the Netherlands) and Central Europe (represented by the Czech Republic, Hungary, Poland and Slovakia). Furthermore, there are significant differences between the four Central European countries.

The theoretical background of the SJTs illustrates that the fundamental socialisation processes (parents, schooling, and so on), which can be seen as products of culture, have an influence on individual judgments of the effectiveness of a particular response to a certain situation. Thus, the application of the SJT in a culture other than that originally intended is not warranted, as the correct or appropriate response to a specific situation might differ as a function of cultural values. The SJT with more cognitive ability content is believed to exhibit more cross-cultural validity than that with more non-cognitive items (Lievens, 2006).

In practice, Lind (2005) notes that the moral judgment test (MJT) has been successfully validated in 29 different language versions and is well suited for cross-cultural research into moral development and education. Similarly, Lievens et al. (2015) examine the transportability of an integrity SJT that was originally developed in the US to a Spanish context. The findings suggest that most SJT items (16 of 19)

are realistic for a Spanish context, and there is strong consensus in the scoring scheme. In addition, correlations between the SJT integrity scores and ratings on a self-report integrity measure do not differ significantly between the two contexts. Despite these optimistic results, there are few studies exploring the use of SJTs in countries other than those in which they originally developed (Lievens et al., 2015).

3. 4. 5 The development process of a situational judgment test (SJT)

Patterson et al. (2015b) state that, 'SJTs represent a reliable, valid, well-received and fair selection method when designed appropriately' (p.12). However, despite their long use, there is no consensus in the literature on how SJTs should be developed, scaled, or scored. In addition, the increase in studies which examine SJTs has produced an increase in ways of developing SJTs (Weekley, Ployhart, & Holtz, 2006). The development process differs from one study to another, combining and separating procedures or introducing new methods. Here, the resaercher highlights the main steps by reviewing the work of Lievens et al. (2008), McDaniel and Nguyen (2001), Patterson et al. (2015b), Ployhart and MacKenzie (2011), Weekley and Ployhart (2006), and Weekley, Ployhart, and Holtz (2006).

The first step is collecting statements about situations that arise on the job. Situations, or item stems, form the basis of any SJT. There are two main methods of collecting, or developing, situations. The most common is the critical incident approach, in which stories about critical situations on the job are provided. The critical incident technique is a set of procedures for collecting direct observations of human behaviour. The word 'incident' relates to any observable human activity that requires a person to perform an action. To be critical, the incident must arise in a situation where the performance seems fairly clear to the observer and its consequences leave little doubt about its effects (Flanagan, 1954). Typically, the situations are collected by subject matter experts (SMEs) (incumbents, senior teachers, supervisors), who are asked to recall good or poor examples of incidents in the work setting. Another source for the critical incidents, in some working contexts, is archival records. The second method uses a model to develop the incidents. The model can be built from a job analysis or theory of effective performance. The literature review identifies the attributes necessary for the work. The item stems in both methods vary according to length, complexity, and fidelity (video or written format).

Secondly, the responses to the scenarios/situations can be collected either at the same time or in separate steps. The response options are behavioural in nature and use both effective and ineffective options to identify better judgment. For each situation, there are multiple responses that can be generated, but these are then reduced to 4-6 options. The responses are normally presented in a short-written format, even when the situations are in video format. The responses can also be collected by SMEs writing various effective and ineffective responses. Another method is the construct-based response options, in which the options target certain construct(s). The applicants are asked to evaluate the responses in different ways, as seen in the following steps.

The third step is to build the response instruction (the question type). The response instructions for the situations can have different forms. The two main forms are the knowledge format ('should do') and the behavioural tendency format ('would do'). Despite the number of studies seeking to determine the impact of these two formats on SJT properties, the type of instruction used depends mainly on the test specification, the context, and/or the targeted applicants. At the end of this stage, the critical incidents and responses are revised and edited by considering similar incidents and the length, complexity, and format of items, and excluding items that show legal attentions.

The next step is to determine the effectiveness of the responses (the answer key). As there is often no absolute answer in an SJT, the answer key refers to the best judgment of a situation from the given responses (what is most likely to be the right answer?). The literature shows that this step can be taken using three main methods. The first method is the rational key, where a pool of SMEs or excellent employees are asked to make decisions about the effectiveness of the responses. Secondly is the empirical approach, which uses a correlation method and certain criterion measures. Finally, the least frequently used method is reliance on theory to determine the effectiveness of the responses. However, Whetzel and McDaniel (2009) state that there is insufficient evidence in the research to judge which scoring strategy is substantially better than the others.

Finally, there is the need to build the scoring key to determine how the test-takers performed in the test. There are two broad categories, according to the type of question. The first technique is the forced-choice method. Here, the test-taker is asked

to choose the best answer(s) or to identify both the best and the worst option. Thus, he/she receives a number of points (one or more) if the answer chosen is correct, according to the answer key, and no points for the wrong choice. The second method is to ask the test-taker to rate the effectiveness of the options in a Likert-type scale. In this method, the distance-measure approach is used to score the answers. The test-taker is given higher points if his/her rating is closest to the answer key of the SMEs, and lower points if not.

Once the SJT stems, responses, and scoring key have been constructed (paper and pencil, or electronic format), the next step is to pilot the test to ensure that it is fair, reliable, and measures what it is intended to measure. After an SJT has been piloted, an analysis of the data can be conducted to ensure that the SJT items perform well psychometrically.

Among the practical studies of developing an SJT for medical school selection, where SJTs have been in use for a long time, Patterson, Ashworth, Mehra, and Falcon (2012a) explain the process of building, piloting, and evaluating a SJT designed to select candidates for UK DFT. The development process begins by identifying the relevant professional attributes for dentistry (for example, empathy and integrity). Test items and options are then developed by SMEs working with experienced psychometricians. The developed test is piloted and evaluated in terms of the normal distribution of the scores, the internal reliability, and the correlation of the scores with those of the admission interview. In addition, candidates' reactions are evaluated using a questionnaire. An SJT was also developed and piloted for the two-year generic training programme which bridges medical school and specialist/general practice training. The development process was explained in a technical report by Patterson, Ashworth, Murray, Empey, and Aitkenhead (2014). A similar process is used to develop SJTs in other fields, such as employee integrity (Becker, 2005), emotional intelligence (Sharma, Gangopadhyay, Austin, & Mandal, 2013), and leadership (Peus, Braun, & Frey, 2013).

As noted earlier, at each step of the SJTs' development, there are different approaches or methods. Research has tested those alternative approaches to identify the best development procedure. Krumm et al. (2015) tested SJTs with situations (item stems) removed. Their results show no significant difference made by the presence (or absence) of the description of the items. Lievens, Sackett, and Buyse

(2009) examine the differential effects of knowledge and behavioural response instructions. The results identify no meaningful differences in low-stakes settings. Ployhart, Weekley, Holtz, and Kemp (2003b) compare two different formats in a selection setting: web-based tests and paper-and-pencil tests. The results indicate that web-based tests have positive benefits compared to paper-and-pencil measures. In addition, Stemler Aggarwal and Nithyanand (2016), St-Sauveur, Girouard, and Goyette (2014), and Guo, Zu, Kyllonen, and Schmitt (2016) tested the different scoring approaches, and the instructions approaches were studied by Ployhart and Ehrhart (2003a). Other studies have sought to develop SJTs using a single-response format (for example, Crook et al., 2011; Motowidlo et al., 2009).

To conclude, research indicates there are differences in the procedures and means used to develop SJTs. In their review of empirical studies since 1990, Campion et al. (2014) found that SJTs vary in 12 areas: situation and response development, key development, scoring methods, scenario presentation, stimulus medium, response medium, response format, instruction format, context, constructs assessed, research design, purpose of study, number of items, sample size, and number of dimensions. Thus, in general, there is insufficient evidence that results can be generalised to better develop the SJT. However, the success of this development process can be measured by testing the main psychometric proprieties of the developed test, such as reliability and validity. In addition, the developed test should have good applicant reactions and reduce sub-group differences. The next sections highlight these features.

3. 4. 6 Sub-group differences in situational judgment tests (SJTs)

Ployhart and Holtz (2008) believe that many organisations seek to implement selection mechanisms that have less impact on minority groups and lead to greater diversity in the workforce. However, some of the most valid selection procedures reveal differences in scores between certain demographics (for example, non-White, female). There is no entirely effective or ideal strategy for reducing subgroup differences and adverse impacts. Since subgroup difference is important when implementing a selection method, studies measure the statistical differences between groups using mean differences (*d*). A *d* of one indicates that one group is one standard deviation above the mean of another (Whetzel & McDaniel, 2009).

SJTs are found to have a less adverse impact than other selection methods (Lievens et al., 2008; McDaniel & Nguyen, 2001; Patterson et al., 2015b). A meta-analysis by Whetzel et al. (2008) indicates that, on average, White test-takers performed better on SJTs than Black (d = .38), Hispanic (d = .24), and Asian (d = .29) test-takers. Female examinees performed slightly better than male test-takers (d = −.11). The differences were tested in terms of (a) loading of cognitive (loading of g) or personality on the SJT, and (b) the response instructions (knowledge and behavioural). The results indicate that differences between Black, Hispanic, Asian, and White people are largely explained by the cognitive loading of the SJT, whereas the personality loadings show that Black–White and Asian–White differences are smaller in emotional stability, and Hispanic–White differences are smaller in conscientiousness and agreeableness. Regarding male–female differences, cognitive loading has a minimal effect; and the differences are larger, favouring women, in conscientiousness and agreeableness. Knowledge response instructions appear to have greater race differences than behavioural tendency instructions. It is concluded that SJTs have less adverse impact on minority groups than cognitive ability tests do.

Similar findings are presented in Lievens et al. (2008). The White-Black differences are considerably reduced for the non-cognitive domains of job performance. In addition, video-based SJTs show less adverse impact than written SJTs, and SJTs with behavioural tendency instructions have less adverse impact than those with knowledge instructions. Regarding gender difference, females score slightly better than males on SJTs. The researchers argue that this gender bias might be due to differences in the personality traits assessed by the SJT situations. The scenarios are often interpersonal in nature, and females tend to score higher on traits such as agreeableness or sociability. Finally, research into medical education and training suggests that SJTs have less adverse impact in terms of ethnicity and gender than other selection tools do, and could promote widening access compared to indicators of academic attainment (Patterson et al., 2015b).

However, it is important to note that meta-analysis studies struggle with the issue of publication bias (Whetzel et al., 2008). In addition, most research included in the studies discussed above were conducted in Western countries, mainly the US. Further studies in different (non-Western) cultures would allow for a better understanding of possible sub-group differences in SJTs.

3. 4. 7 Reliability of the situational judgment test (SJT)

Reliability is a fundamental aspect, which must be tested for a measurement, as it indicates that the measurement (or scale) is free from random error. In other words, it suggests consistency across time or items. It can be measured by (a) testing the similarity (or differences) of the scores of a group of items built to measure the same construct (internal consistency), (b) measuring the stability of the scores over a given time period (test–retest reliability), and/or (c) by correlating results from two versions of the same test (parallel forms reliability) (Punch, 2013).

There are challenges to the methods used to measure the reliability of the SJT. One of the challenges in measuring the internal consistency of an SJT that each of its items may target several dimensions or constructs, which implies difficulty in underestimating its reliability (McDaniel, Hartman, Whetzel, & Grubb, 2007). Schmitt and Chan (2006) suggest that there is no clear factor structure of an SJT, hence the attempt to analyse the internal structure of the SJT using an exploratory factor analysis (EFA) has often produced 'disappointing' results. This is confirmed by Sorrel et al. (2016), who reveal that the methods of assessing the reliability of SJTs — factor analysis techniques and Cronbach's alpha coefficient — have proven inadequate. Similarly, Kasten and Freund (2015) state that factor analyses in the literature often reveal a dominant factor, which might justify the account of SJT reliability for the general score rather than different sub-scores.

Despite the problematic nature of measuring the reliability of SJTs using the coefficient alpha, most studies report this. Since 1990, 88.4% of studies on SJTs report reliability using the coefficient alpha, whilst only 5.5% use the test-retest measure, 3.4% parallel-form, and 2.7% split half reliabilities (Campion, Ployhart, & MacKenzie, 2014). However, the test–retest or parallel forms reliability are seen as more accurate for examining reliability (Lievens et al., 2008; McDaniel & Nguyen, 2001).

The reliability value of the SJT is analysed in many reviews and meta-analyses. A meta-analysis by McDaniel et al. (2001) highlights Cronbach's alpha coefficients of $\alpha = 0.43–0.94$. Patterson et al. (2015b) state that the internal consistency of the SJTs used in medical and dental contexts is approximately $\alpha = 0.7$ or more. Catano, Brochu, and Lamerson (2012) computed the corrected weighted mean alpha from 56 alpha coefficients and found a value of $\alpha = .46$. The other two forms of reliability

(test-retest and parallel-form) are also reported in SJT research. Studies show a range of $r = 0.20$ to $r = 0.92$ for the test-retest reliability (Ployhart & Ehrhart, 2003a; Ployhart et al., 2003b), whereas Chan and Schmitt (2002) report a parallel-form reliability of 0.76. Generally, and regardless of which test is used, the research broadly shows that SJTs have moderate to good levels of reliability (Patterson et al., 2015b).

Finally, it is important to note that the reliability of SJTs is affected by their characteristics and the context in which they are tested. The results show that the reliability of the SJT scores is low, and falls below recommended levels in high-stakes settings (rather than low-stakes settings). Secondly, compared to the simple 'pick the best' scoring approach, both the 'pick best/pick worst' and the Likert scale approaches expect higher reliability. In addition, the results indicate better estimates of reliability when using the theoretical approach than the empirical key approach. Moreover, SJTs with more items show higher internal consistency than those with fewer items (Kasten & Freund, 2015). Campion, Ployhart, and MacKenzie (2014) indicate that SJTs in video-format tend to have lower reliability. They claim that the information given in video-format tests is much greater than in the written tests and this is likely to contribute to a higher variance in scores. They find also that the type of measured construct can affect reliability. For instance, social skills constructs are likely to have lower reliability than teamwork constructs.

3. 4. 8 Validity of the situational judgment test (SJT)

Validity simply means the degree to which an instrument measures what it is developed to measure (Punch, 2013). There are different ways to evaluate the validity of an SJT. These include measuring the following: (a) content validity, to ensure that test items are sufficient and cover the test objectives, which can be achieved by making professional judgments; (b) criterion-related validity, which correlates the score of the test with those of other tests measuring the same factor(s); (c) construct validity to ensure that performance on the test is fairly explained by appropriate constructs or concepts, where comparisons are made with measures of similar constructs; (d) concurrent validity, which correlates results with those on other tests assessing the same performance; (e) face validity, to ensure that the test tests what it is designed to; and (f) predictive validity, where results accurately predict subsequent performance (Cohen, Manion, & Morrison, 2013). In general, interpretations of what

an SJT measures are based on correlations of SJT scores with external measures (Jackson, LoPilato, Hughes, Guenole, & Shalfrooshan, 2016). Although SJTs have been used for a long time in personnel selection, concern with their validity is a recent phenomenon (Weekley & Ployhart, 2006).

In a study of the construct validity of the SJTs, Christian, Edwards and Bradley (2010) classify the construct domains assessed by SJTs in the literature. They identify that most studies measure leadership (37.5%), interpersonal skills (12.5%), basic personality tendencies (9.56%), teamwork skills (4.41%), and job knowledge and skills (2.94%), and 33% of the SJT studies have unclassified constructs. In addition, they conducted a meta-analysis of the criterion-related validity of each construct domain. They state that the mean validity of the studies measuring teamwork skills was .38, leadership skills was .28, interpersonal skills was .25, and conscientiousness was .24. In another meta-analysis, the validity of indicates a moderate correlation with general mental skills (McDaniel & Nguyen, 2001). Additionally, the results indicate that an SJT designed to measure the construct of integrity has a significant correlation with the dimensions of honesty-humility (integrity), conscientiousness, extraversion and agreeableness dimensions, ranging from .16 to .36 (Husbands, Rodgerson, Dowell, & Patterson, 2015).

The predictive validity of the SJT is mostly tested using longitudinal studies. For example, Patterson et al. (2016c) evaluate the predictive validity of an SJT for entry into postgraduate GP specialty training in Australia. The results show that the participants' performance on the SJT and the overall selection score significantly predicted all three end-of-training assessments (r = .12 to .54), indicating good predictive validity. Moreover, an exploratory longitudinal study was conducted to evaluate the validities of selection tests (including SJTs) that used in the recruitment process for candidates applying to training in UK general practice in 2009. The results indicate positive and significant relationships between the selection tests (Koczwara et al., 2012). In another study, Ahmed, Rhydderch & Matthews (2012) state that SJT is a better predictor of workplace-based simulation exercises at a selection centre performance. The SJT was found to be the most effective independent predictor, based on evaluations of three shortlisting methodologies in selection for postgraduate training in general practice (Patterson, Baron, Carr, Plint, & Lane, 2009). Finally, using a longitudinal and multiple-cohort design, Lievens (2013) found video-based

SJTs as measures of interpersonal behaviour had significantly more value than cognitive tests for predicting interpersonal GPA and doctor performance.

Furthermore, studies suggest that SJTs have an incremental validity over other tests. Lievens and Patterson (2011) evaluated the validity of three tests (knowledge tests, SJTs and ACs) in advanced-level high-stakes settings. The results show that both the SJT and the AC had incremental validity over the knowledge test, whilst the AC had incremental validity over the SJT. In the context of admissions to medical and dental studies in Belgium, Lievens, Buyse, and Sackett (2005) found that the SJT had incremental validity over cognitively oriented measures for curricula that included interpersonal courses, but not other curricula. A similar result is presented by Patterson, Lievens, Kerrin, Zibarras, and Carette (2012c) in their review of studies investigating the effectiveness of multiple selection instruments used for medical education and training recruitment in high-stakes processes. Their results suggest that the SJT is the best single predictor of performance, with incremental predictive power over cognitively oriented tests. Finally, Chan and Schmitt (2002), in a study of 160 civil service employees, note that SJTs provide incremental validity for the prediction provided jointly by cognitive ability, the Big Five personality traits, and job experience.

However, Whetzel and McDaniel (2009) note the limitations of the validity research. They argue that most SJT validity studies rely on concurrent designs where respondents are incumbents who have little motivation because the test results will not affect their careers. Tests given to job applicants indicate greater motivation and 'fake' responses in order to present oneself in a better light. In addition, results from meta-analyses may not be sufficiently accurate to judge the validity of a specific SJT because different models measure different constructs. Despite that, Patterson et al. (2015b) state that there is good evidence in healthcare that SJTs have added value over other selection measurements for predicting job performance.

3. 4. 9 Applicants' reactions to the situational judgment test (SJT)

It is important to understand how applicants assess the fairness and equity of the selection procedures and how organisations and institutes could improve perceptions in order to raise job acceptance or to decrease the possibility of claims. Applicants' reactions have significant consequences in the selection process and for the

organisation policy as a whole (Weekley & Ployhart, 2006). The publication of the Gilliland (1993) classic organisational justice model of applicant perceptions had a significant impact on the study of applicants' reactions. McCarthy et al. (2017) reviewed 145 primary studies and several meta-analyses published since 2000. They found evidence that applicant reactions have significant effects on attitudes, intentions, and behaviours. The Gilliland (1993) theoretical model includes 10 procedural justice rules in three broad categories. The formal characteristics category includes job-relatedness, chance to perform, reconsideration opportunity, and consistency. Under the explanation grouping is feedback, information known, and openness. Finally, the interpersonal treatment domain includes treatment at the test site, two-way communication, and propriety of questions (Bauer et al., 2001).

Patterson, Zibarras, Carr, Irish, and Gregory (2011) used organisational justice theory to study applicants' reactions to the selection methods (including SJTs) used for medical training in the UK. They developed an evaluation questionnaire for completion by the applicants immediately after the selection stages. The results indicate positive perceptions of fairness of all the selection methods, and an affirmation that all were job-related. However, initial candidate reactions were less positive to the SJT than to the other selection methods (for example, CPST). This was explained due to the candidates' preferences for methods with clear answers based on facts, whilst that not seen by the candidates in the SJT. In another study, Patterson et al. (2012c) note that the SJT receives lower face validity ratings than the knowledge test. Participants perceive the knowledge-oriented test to be more relevant to their role than the SJT, which focuses on non-cognitive attributes. As a solution, research suggests the importance of increasing information given to candidates about the SJT through different interventions, such as using the recruitment website (or other means) to give detailed information about the rationale for the SJT, what the test is measuring, and how it is scored.

In education, Klassen et al. (2014b) studied applicants' reactions to taking the SJT for entry into primary and secondary ITEPs in UK universities, using a theoretical framework of organisational justice. Participants were invited to provide feedback, after completing the SJT, on seven items: content relevance, difficulty, and fairness, and SJT differentiation, fairness, appropriateness, and measurement. The data indicate that the reaction to the content and format of the SJTs is good and most

applicants (76.7%) found the tool favourable. The results of open-ended questions recommend that separate selection tests should be created for primary and secondary applicants.

3. 4. 10 Threats to use of situational judgment tests (SJTs): coaching and faking

The above sections concern the challenges to the constructs that SJTs measure, and how these can affect the reliability and validity of the test. As for other measurement methods, there is a growing interest in understanding the effect of coaching and faking on SJT scores. This section highlights the results of studies on those issues.

Stemig, Sackett, and Lievens (2015) examined the effect of coaching type on SJT score and on construct-related and predictive validity in the context of medical school admissions. The results suggest that commercial coaching techniques have less effect on SJT scores than the organisationally provided methods. In addition, the criterion-related validity of the SJT scores is not degraded by the availability of coaching. The study suggests making effective, organisationally endorsed coaching available to all applicants in order to overcome the unfairness of coaching. In another study, the coachability of two situational judgment tests, the College Student Questionnaire (CSQ) and the Situational Judgment Inventory (SJI), used in the college admission process, are examined by Cullen, Sackett, and Lievens (2006). The participants were trained in the use of strategies for raising scores on each test using a video-based training programme. The scores on the CSQ appear to be more affected by coaching than by the SJI. The study concludes that the difference in the effectiveness of the coaching programmes was not due to one training method being better than the other, but rather the SJI strategies themselves were more difficult to understand and apply than the CSQ strategies. This weaker effect of coaching on SJT score is also acknowledged by Simon, Walsh, Paterson-Brown, and Cahill (2015), who state that there is no difference in SJT scores for students who use additional resources, including textbooks and study courses, to revise for the test.

Concerning the ability to fake responses to the SJT, some studies investigate whether participants can intentionally change or fake responses. Peeters and Lievens (2005) examined the fakeability of an SJT of college students' performance by assigning an honest and a fake condition. In the fake condition, participants were

instructed to respond as if they were taking part in a college admission exam to obtain the highest scores. The results suggest that the scores in the fake condition were significantly higher than those in the honest condition, and faking had a negative effect on the criterion-related validity and the incremental validity of the SJT. Thus, the results indicate that faking is a possible threat to the use of SJTs in a high-stakes selection process. However, Nguyen, Biderman, and McDaniel (2005) state that SJTs can be faked, but the degree of faking may vary according to the response format. They claim that the knowledge response format is more resistant to faking.

3. 5 Summary of the literature review and contribution to the study

The two main aims of this study are as follows: (a) the identification of the key non-cognitive attributes necessary for prospective teachers in Oman, and (b) an exploration of the ability to use SJTs in the selection of applicants for ITEPs in Oman. To achieve these aims, it is important to review and discuss related research and studies. Hence, in this chapter, four related areas have been reviewed, namely: (a) personnel selection, (b) non-cognitive attributes, (c) the ITEP selection process, and (d) the use of SJTs in the selection process. In each section, the associated themes, such as definitions, theoretical bases, and the selection methods used in both educational and non-educational fields are highlighted and discussed.

Here, a summary is given of the main points raised in the previous review and how they contribute to the aims of this study:

- Firstly, the selection of individuals, for any organisational purpose, must be seen by policy makers as an important strategy which can directly affect the future successes (or failures) of the organisation. The risk of a failed selection strategy is higher in professions and contexts where it is difficult to fire ineffective employees, such as teaching.
- The two fundamental questions in any selection process are ‘what’ and ‘how’. The answers to these must be identified in a systemic evidence-based process. The outcomes must be continuously piloted, tested, and evaluated according to the outcomes and the needs of both the applicants and the organisation.

- Despite the argument about their definitions and consistency, non-cognitive attributes are important for economic and social life. They must be part of any selection process.
- Building a framework of the key non-cognitive attributes necessary for selection is an evidence-based process. The identification of key non-cognitive attributes for ITEP selection could benefit from the processes used in other professions (for example, medicine and law). The context of the study, the stakeholders' perspectives, related theories, and the successful practices could be input into the framework building process. The review of the literature in this section contributes to the first aim of this study. In Phase one, in the research method, a framework of the key non-cognitive attributes of effective prospective teachers in Oman is built, using (a) the summary of the non-cognitive attributes used for selection in teaching and other professions (see Appendixes 1, 2 and 3), (b) a review of the related policies and regulations used for the teaching profession in Oman, and (c) a collection of data from stakeholders, using interviews and questionnaires, these being the methods used in similar studies.
- All selection measurements have advantages and disadvantages. However, most ITEPs use traditional untested selection methods, paying little attention to the non-cognitive attributes of the applicants. Recent studies in personnel selection and evidence-based research into selection practices used in other professions present SJTs as promising for development and testing to assess applicants to ITEPs.
- The current work of Klassen et al. (2014, 2017b) on building and testing the use of SJTs for ITEP selection is the starting point for this research. This study has extended the initial non-cognitive framework and the developed SJTs in the UK to a non-Western context (Oman).

Chapter 4 Research Methodology

This chapter presents the methodology used to accomplish the aim of the study and to find answers to the research questions. Specifically, it illustrates how the data were collected and analysed, and where they came from. The structure of the chapter comprises eight main sections. Firstly, the aim of the study and the research questions are presented, then Section 2 illustrates the research design. The method and participants are summarised in Section 3. The instruments, participants, and procedures of the four phases of the study are highlighted individually in Sections 4 to 7. The last section focuses on the ethical considerations of the study.

4. 1 Aim and research questions

The literature emphasises the importance of the research questions in order to organise and direct the study, keeping the researcher focused on the goals, and it provides a framework for the writing of the research. Research questions differ from the research aim in that they are explicit statements about what is to be investigated. The process of building good research questions is not simple. It involves identifying the problem, generating possibilities, splitting general questions into more specific ones, discussing the researcher's views with others, and determining what is important according to the available time and resources. The knowledge of the researcher is also important and can be used as a starting point (Punch, 2013; Bryman, 2015).

As noted in the Introduction, the main aim of this study is to explore the development process and initial findings around using SJTs to best understand the non-cognitive attributes of undergraduate applicants during the ITEP admission process in Oman. A review of similar work around admission to medical schools (for example, Patterson et al., 2000, 2008, 2012a, 2013b) and, more recently, admission to teacher education programs (Klassen et al., 2014b, 2017b) has generally begun by searching for the key non-cognitive attributes that form the specifications of the SJT, and then exploring the properties of the test in terms of reliability, validity, and applicants' reactions. These properties are some of the principles necessary for psychological testing (Kaplan & Saccuzzo, 2009). In addition, Eignor (2001) states that, due to the wide range of tests, major professional organisations in the US (the

American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME)) have jointly created certain fundamental elements known as the 'Standards for Educational and Psychological Testing'. Although the standards do not specify a limit for the quality of a test, the required quality is mostly influenced by the purpose of the test within the targeted context (Cronbach, 1990).

Therefore, building on the researcher's knowledge of the education sector in Oman over more than 20 years and a review of similar studies in the literature, this study aims to address the following research questions:

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

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

4. 2 Research design

The research design acts as a framework for a researcher collecting and analysing data, while the techniques used for collecting data are 'research methods' (Bryman, 2015). At this point, the general framework of the study is illustrated, while the research methods are presented in the next section.

For better development of the research design, the literature on methodology provides different philosophical views of social research. Some of the philosophical perspectives are related to the nature of the research phenomena (ontology), while others discuss the manner in which research should be conducted (epistemology). In addition, there are views of the ethical behaviours and values necessary to support the research (axiology). Methodology raises a philosophical question about how a researcher can obtain the desired knowledge. The answer to these philosophical questions of ontology, epistemology, axiology, and methodology forms is what has been called in the literature a 'paradigm'. Where 'research' is a way of knowing and understanding, the 'paradigm' is the guidance which directs our thinking and actions in a particular way. It influences what is to be researched, what questions can be asked, which methods will be used, and how the findings will be interpreted (Bryman, 2015; Coe, 2012; Mertens, 2014).

The literature includes different views of these philosophical questions. In ontology, for example, which looks to the nature of reality, some see social phenomena as an external concept which cannot be controlled (in effect, objectivism/realisms), while others identify it as a result of different interactions (constructivism). The assumptions of epistemology also vary. Research can be conducted using scientific approaches (positivism) or other approaches that suit the qualities of people and social institutions (interpretivism). However, despite attempts to classify research according to certain philosophical view(s), the determination of one particular paradigm for research is complex and perhaps not possible (Bryman, 2015).

Johnson and Onwuegbuzie (2004) describe that difference in assumptions and views as the 'paradigm wars' (p.14). They analysed the views of two well-known and deep-rooted paradigms: the 'quantitative' paradigm, which classically follows the positivist philosophy, and the 'qualitative' paradigm, which follows the constructivist and interpretivist philosophy. The debate between the two traditional paradigms has led to the provision of a framework for designing and conducting mixed methods research, which follows the pragmatist philosophy. According to this view, research methods should be guided by research questions in the way that offers the best chance of obtaining useful answers. Recently, there has been increased interest in a combination of the two approaches. Quantitative and qualitative methods are more

powerful when used in combination than in isolation (Gorard & Taylor, 2004; Punch, 2013).

As noted in the previous point, the research design of the study benefited from a review of the designs used in similar studies (for example, Durksen & Klassen, 2017; Klassen et al., 2014b, 2017b; Patterson et al., 2000, 2008, 2012a, 2013b). To our knowledge, this is the first study conducted in Oman of the development of an SJT for selecting prospective teachers. Therefore, an explorative research design with a mixed method of qualitative and quantitative approach was used. This explorative mixed-method approach is suitable when a researcher has little or no scientific knowledge about a given phenomenon, as it allows for more flexibility when seeking relevant data (Stebbins, 2001). The research design consists of four complementary phases. The next section gives a general overview of the methods used in the phases of the study, while the details on each phase are presented in the following sections (4.4 to 4.7).

4. 3 Research method

As noted earlier, this study consists of four phases. The goal of Phase one was to find the specification of the SJT (in effect, the key non-cognitive attributes seen as necessary for prospective teachers in Oman – thus answering Research Question 1). This phase consisted of four main steps. Step one focused on analysing related official documents from the Ministry of Education (MoE) and an ITEP in Oman. A semi-structured interview was then conducted with a sample of stakeholders (in effect, tutors in ITEPs, school principals, and teachers' supervisors). The results of the two steps were then compared to the three domains found in the UK (Klassen et al., 2014b) to produce an initial framework of the domains in Oman. Finally, the initial domains/attributes were rated by a sample of stakeholders, using a closed questionnaire. The outputs of Phase one were used to develop the SJT in Phase two.

In Phase two, an SJT targeting the domains found in Phase one was developed with a sample of working teachers. The development process consisted of five main steps: collecting incidents (situations), developing response options, response instructions, determining the response effectiveness (in effect, the answer key), and building the scoring method. The developed SJT was piloted in Phase three. In Phase

four, the SJTs and other criterion measures were implemented to find the reliability, validity, and applicants' reactions (in effect, answering Research Question 2).

Table 4.1 summarises the four phases of the study in terms of the objective, procedures, and outcome. It is important to note here that the work of Klassen et al. (2014b, 2017b) – specifically the three non-cognitive domains and the 35-item SJTs found in the UK – were used as inputs in phases one and two. This will be clarified in detail in the next sections.

Table 4.1 A Summary of the Four Research Phases

Phase One	Aug. 2015 – Jan. 2016	Objective: - Identifying the SJTs' specification (i.e. key non-cognitive attributes of prospective teacher in Oman) (Answer Research Question 1). Procedure: - Analyse related official documents in Oman. Exploratory semi-structured interview (*N* = 8). Findings from the steps above were compared with Klassen's work. Exploratory closed questionnaire (*N* = 181; 58% females). Outcome: - List of the key non-cognitive attributes/domains (Answer RQ1).
Phase Two	Mar. – Apr. 2016	Objective: - Developing the SJTs for Oman. Procedure: - Translating the SJTs built in the UK to Arabic. Collecting additional items from teachers in Oman. First review by the researcher. Second review by a group of expert teachers (*N* = 8; 50% females). Building the answer key through expert teachers (*N* = 108; 48% females). Building the scoring key. Outcome: An initial SJTs for selecting applicants into ITEPs in Oman.
Phase Three	Aug. – Sep. 2016	Objective: - Piloting the SJTs to check reliability, items' quality and face validity. Procedure: - A sample of applicants at an ITEP in Oman (*N* = 171; 53.4% female). Analysing the results. Outcome: The final SJTs to be used for the implementation phase (Phase 4).
Phase Four	Feb. – Mar. 2017	Objective: - Implementing the SJTs to answer Research Question 2. Procedure: - Another sample of applicants at an ITEP (*N* = 142; 73.9% females). Analysing the results. Outcome: Answering the second research question (reliability, validity, and applicants' reaction).

Although the developed SJT targeted applicants for ITEPs, the participants in the phases of the study were different groups of working teachers and ITEP students. The working teachers participated in phases one and two, finding the key non-cognitive attributes and developing the SJTs; whereas, the ITEP students participated in phases three and four in order to evaluate the properties of the developed SJTs. More details of the procedure and the participants in each phase are given in the next sections.

4. 4 Phase one: building the SJTs' specifications

The goal of Phase one was to build the specifications for the SJTs by identifying the key non-cognitive attributes important for prospective teachers in Oman. These attributes formed the inputs for developing the SJT items in Phase two. To fulfil the goal of this phase, and due to a lack of previous research in this area in Oman, an explorative research approach was taken.

Here, the explorative approach was conducted through four complementary steps. At the first step, related official documents from the MoE and ITEPs in Oman were reviewed. An explorative semi-structured interview was then conducted with key stakeholders (in effect, tutors at an ITEP, school principals, and teachers' supervisors). Findings from the two steps were then compared to the findings in the UK, resulting in an initial framework of key non-cognitive attributes. Finally, an explorative closed questionnaire, completed by a larger sample of teachers, supervisors, and schools' principals, was used to rate the importance of the initial attributes. The next points identify the procedure and participants for each step.

4. 4. 1 Step one: review of official documents

Documents are defined as materials that can be read, which were not produced for the aim of the study, and which are relevant and available for analysis (Bryman, 2015). They can take different forms, such as personal documents (diaries, letters), official documents from the state or private sectors, or mass media reports. Although the reliability of official documents is criticized for the manner in which they privilege the 'top-down view' of education of policy-makers, documents produced by organisations are a rich source of data for social science researchers. However, documents do not speak for themselves but require careful analysis and interpretation (Cohen, Manion, & Morrison, 2013; Punch, 2013).

Here, a review of the official documents focuses on finding the key non-cognitive attributes necessary for teachers, or prospective teachers, in Oman. This section highlights how documents were selected and analysed.

4. 4. 1. 1 ***Document selection procedure***

The review of the context in Oman in Chapter 2 highlights the lack of a single document describing the non-cognitive attributes necessary for teacher effectiveness in Oman. Hence, two main sources were targeted to obtain this: the Ministry of

Education (MoE), which is responsible for working teachers, and one of the main ITEPs in Oman, which is responsible for selecting and preparing future teachers.

In the targeted ITEP, the focus was on the documents that illustrate the competencies used for selecting and preparing the new students (prospective teachers). The related documents from the MoE in Oman focused on the job descriptions of teachers and the teachers' evaluation criteria. An analysis of these documents allowed identification of the non-cognitive attributes used to select prospective teachers, guide teachers in doing their jobs, and evaluate teachers' performance. Five main documents were selected.

Two documents were selected from the targeted ITEP. Firstly, the candidates' proficiencies are used in one of the ITEPs in Oman to select and prepare students. The proficiencies reflect the ITEP conceptual framework, which has five themes: academic rigor and specialised experiences, diversified teaching, dispositions and values, research culture and lifelong learning, and technological skills. Each theme includes certain proficiencies that describe the distinguished graduate. 'Disposition and values', for instance, includes proficiencies related to commitment, cooperation and 'Islamic principles'. The second document is the 'interview form' used to assess the applicants in the admission process. Three documents were selected from the MoE: the teachers' job description, which describes the tasks and responsibilities of teachers in schools, the classroom supervision visit, which is used to evaluate teachers' performance inside the classroom, and finally the teachers' annual appraisal, which evaluates the teachers' performance during the schooling year.

Some of the targeted documents were available online through the official websites of the organisations (the MoE and ITEPs). However, other documents were obtained officially from the concerned bodies. The last section of this chapter highlights the ethical considerations associated with the collection of the documents.

4. 4. 1. 2 ***Document analysis procedure***

The interpretation of the documents was achieved mostly using qualitative content analysis, searching for underlying themes (Bryman, 2015; Marshall & Rossman, 2014). For Franzosi (2004), content analysis is a tool to analyse written materials such as documents and interviews. It requires the researcher to become familiar with the targeted texts by reading them many times. There are different

techniques for analysing qualitative data. Thematic analysis is suitable to develop a clear picture of the contents, while referential analysis is more concerned with how certain objects are presented in a text (and not what they are). Frame analysis goes further by seeking to understand how the meanings are built into the written material. Finally, structural narrative analysis is concerned with social actions and interactions. The most common approach is thematic analysis. Using this technique in documents, the researcher captures the dominant themes that suit the aim of the analysis.

Our aim in analysing the contents of the documents was to identify the explicit and implicit non-cognitive attributes related to teacher effectiveness in Oman. Thus, the researcher first read the documents numerous times to become familiar with the contents. Then, the written task, or the criteria contained in the documents related to non-cognitive attributes were highlighted. The highlighted texts were then copied into a Word document and printed. Similar texts from the documents were grouped and put into a table. Finally, the related attributes for each group were identified. Table 4.2 gives an example of the analysis process, and the main findings are presented in the next chapter.

Table 4.2 An example of the analysis process for the official documents

Task / criteria	Source	Related attribute
Establishes a time schedule to carry out its duties and responsibilities.	Job description	Planning
Preparing an integrated and effective annual / daily plan.	Classroom's supervision criteria	

4. 4. 2 Step two: the interview

While the previous step looks at the non-cognitive attributes found in official documents, the aim of this step was to collect rich explorative data about the key non-cognitive attributes from stakeholders' perspectives. To achieve this, semi-structured interviews with three stakeholder groups were conducted. Punch (2013) states that interview is a good tool to gain an understanding of individuals' perceptions, meanings, definitions of situations, and considerations of reality. The following points describe the instrument's development, the participants, the procedure, and how the data have been analysed.

4. 4. 2. 1 *Instrument*

The initial interview questions were constructed by the researcher and comprised four closed questions and one open question. The questions were designed to trace the necessary non-cognitive attributes. However, other questions sought the interviewees' perspectives of the availability of these attributes in current teachers and student teachers, with one question exploring the importance of enhancing the current admission system for ITEPs in Oman.

The interview targeted three groups: college tutors, schools principals, and teachers' supervisors. Very similar questions were used for all three groups, with slight changes to reflect their duties. For example, where supervisors and school principals were asked about current teachers, the college tutors were asked about student teachers. The interview questions were initially discussed with and revised by the researcher's supervisor. The first draft, including the questions and a consent form, was then translated to Arabic by the researcher.

The initial Arabic version of the questions and the consent form were given to two reviewers in Oman to check their clarity, relevance, and suitability. The questions

for the college tutors were checked by one of the academics at an ITEP, who did not participate in the interview process. The questions for supervisors and school principals were checked by a member of the supervision department at the MoE. The reviewers suggested some small amendments to improve the clarity of the questions. Appendix 4 presents the final English version of the interview questions and the consent form.

4. 4. 2. 2 ***Participants***

As noted earlier, three stakeholder groups were targeted (college tutors, teacher's supervisors, and school principals). These groups were selected for their responsibilities in preparing, monitoring, and evaluating teachers before and after joining the profession. The sampling procedure began by following the official protocol to obtain permission to conduct research in the government sector in Oman. A brief description of the research objectives, procedure, participants, and interview questions (in Arabic) was sent to the MoE and to a main ITEP in Oman. Approval was given via official letters sent to the concerned departments to facilitate the researcher's mission.

For the targeted ITEP, the researcher met the assistant dean of the college to explain the objectives of the study and seek nomination of two tutors for interview who were currently involved in the selection process for new students and the teaching programmes. Two participants were nominated. The consent form and the interview questions (in Arabic) were sent to these individuals by email prior to the agreed interview appointment.

Regarding the sample of school principals and supervisors, the researcher asked for these to be chosen by the education authority in which the researcher works. An official letter was sent to the targeted governorate from the Technical Office for Studies at the MoE. The researcher met the director of the HR development department in the governorate, who is responsible for supervisors and school principals. A brief description of the research was given at the meeting. Three supervisors (two males, one female) and three school principals (two males, one female) were recommended for interview. The consent form and interview questions (in Arabic) were sent to the participants prior to the agreed interview appointments.

4. 4. 2. 3 *Procedure*

The eight interviews were conducted at the beginning of the academic year 2015/2016, between 27 August and 3 September 2015. They were all one-to-one interviews and conducted in Arabic. They were not audio recorded but documented in a written note format. The interviews were undertaken in places and at times convenient for the participants. The time taken for each interview was approximately 30 minutes, as proposed in the covering letter. Since the interviews were conducted in friendly atmospheres, most interviewees took extra time to explain their perspectives.

Each interview began by thanking the participant for taking part in the study. The main goal was then explained, along with its expected contribution to the educational system in Oman. The participants were asked to show their agreement to participate by signing the consent form which details their rights as interviewees. During some of the interviews, the participants mentioned incidents related to the subject of the research, which the researcher asked for permission to include in the study, if necessary. At the end of each interview, a summary of the answers was presented to the participant and they were asked for any comments on this.

4. 4. 2. 4 *Data analysis procedure*

Creswell and Poth (2018) illustrate three main methods on analysing qualitative data: preparing and organising the data (transcripts); reducing the data into themes; and representing the data as figures, tables, and discussion. The eight interviews were first transcribed in Arabic. The transcription process was, simply, transferring the spoken words to a written format (Marshall & Rossman, 2014). A thematic analysis approach was taken, looking at the repeated themes and their similarities and differences (Bryman, 2015). Each transcript was read and re-read to ensure familiarity with the text. The non-cognitive attributes mentioned in each transcript were then highlighted. The attributes were compared within and across the interviews to avoid repetition. As the aim of this step was to explore as many non-cognitive attributes as possible, there was no need to compare the groups. Finally, all mentioned attributes, without repetition, were listed. The findings are presented in the next chapter.

4. 4. 3 Step three: initial attributes vs. Klassen's framework

The initial attributes found in steps one and two were put into a matrix (see Appendix 5), including the attributes of the three domains found by Klassen et al.

(2014b): empathy and communication, organisation and planning, and resilience and adaptability. In this step, the researcher sought to build an initial framework of the key non-cognitive attributes. The importance of the initial framework was then assessed with a sample of stakeholders in the following step.

The comparison process resulted in the establishing of two new domains for the context in Oman, namely: enthusiasm and motivation, and professional ethics. In addition, the first domain, 'empathy and communication', was changed to 'communication skills'. Here, 'domain' refers to a group of non-cognitive attributes that share the same features. As a result of this step, five domains (including 29 attributes) were found to be important for prospective teachers in Oman. These are distributed as follows: communication skills (seven attributes), organisation and planning (five attributes), resilience and adaptability (six attributes), enthusiasm and motivation (five attributes), and professional ethics (six attributes). These domains and attributes were used to build the questionnaire in the next step. More details on the results of this step are presented in the next chapter.

4. 4. 4 Step four: the questionnaire

This step aimed to explore the opinions of a larger sample of teachers, supervisors, and school principals regarding the importance of the initial list of non-cognitive attributes developed in the previous step. Questionnaires are a widely used instrument for collecting data from a large group of people and providing structured information (Cohen et al., 2013).

This step was fundamental to the study for two reasons. Firstly, it summarised the non-cognitive attributes found in the previous steps and presented these to a large sample of participants. Secondly, the findings of this step were a prerequisite of the framework used to build the SJT in Phase two. The next points explain the instrument, participants, procedures, and data analysis procedure used in this step.

4. 4. 4. 1 ***Instrument***

The first draft of the questionnaire was constructed by the researcher and included three parts (plus the consent form). Part one concerned the participants' personal information: gender, job, date of appointment, and educational governorate. The selection of these four independent variables reflected the main characteristics of the teaching force in Oman, and therefore helped to test the differences in

perspectives within and between these groups. The second part requested the participants' opinions of the importance of the five domains for teachers and for selecting candidates for ITEPs. The last part explored participants' agreement about the importance of the attributes for teachers and for ITEP candidates. Both parts two and three in the initial draft used five-point scales. The questionnaire was designed to be accessible online, using Google Docs. This form of online questionnaire is assumed to be more economical in terms of reaching larger samples in less time and allowing data to be collected quickly. However, researchers should be mindful of participants' level of internet access (Bryman, 2015).

The first draft of the questionnaire was piloted with a convenience sample of nine participants to check clarity, relevance, and suitability: school principals (N = 3), supervisors (N = 2), college tutors (N = 4). Six responses were received and analysed. Three major points of feedback should be mentioned. The first concerned the scale. The participants indicated the need to expand the five-point scale in order to give respondents more choices. This led to the introduction of a 10-point scale (1 = not important, 10 = very important). Secondly, some argued that the design of the question could make it difficult to distinguish between effective teachers and effective candidates. Thus, the final draft separates questions about effective teachers and effective candidates into parts two and three. The participants also provided comments on the clarity of some of the domains and the attributes. The 'professional ethics' domain, for example, was described as 'moral fitness' in the pilot draft. The feedback indicated a lack of clarity around the Arabic term for 'moral fitness' and a preference for 'professional ethics' as an expression, as this is widely used in the educational context in Oman. Appendix 6 presents the final draft of the questionnaire (in English).

4. 4. 4. 2 ***Participants***

The target population was teachers' supervisors, school principals, and teachers in government schools in Oman. The focus was on schools with grades 5-12 (cycle two and post-basic education). The researcher excluded grades 1-4 (cycle one), because the majority of the ITEP graduates in Oman work in school grades 5-12.

According to the data for 2015/2016, the targeted population comprised approximately 46,000 teachers, supervisors, and school principals, distributed

between 11 educational governorates. The literature indicates that, in a field of such large organisations, the use of convenience samples is more common than probability sampling, and sample sizes are mostly affected by considerations of time and cost (Bryman, 2015). Therefore, three convenience governorates were chosen from the 11 educational governorates in Oman. In each governorate, the sample targeted all supervisors and school principals in schools with grades 5-12. Due to the large number of teachers, a convenience sample of teachers in four schools (two male schools and two female schools) was targeted from the local educational authority in each governorate. Table 4.3 presents the target population and sample (approximate data for the academic year 2015/2016).

Table 4.3 Population and sample for the questionnaire

Governorate	Supervisors			School principals			Teachers		
	Population	Sample	%	population	Sample	%	Population	Sample	%
Muscat	274	274	100	105	105	100	6031	300	5
Batinah S	263	263	100	90	90	100	5296	250	5
Sharqya N	190	190	100	68	68	100	3406	200	6
Other governorates	1488	0	0	560	0	0	28578	0	0
Total	2215	727	33	823	263	32	43311	750	1.7

4. 4. 4. 3 ***Procedure***

Permission to distribute the questionnaire in the three targeted governorates was sent by email from the researcher to the Technical Office for Studies at the MoE in Oman. The governorates were then asked to distribute the questionnaire, via a link format, to the participants. The implementation process began in December 2015 and continued until January 2016. Although the questionnaire was in an electronic format and could be completed within a few minutes, the response rate was very low. Hence, the researcher travelled to Oman to visit the three governorates in order to boost the participation rate. Ultimately, only 181 of the participants responded. It is worth mentioning that this low response rate is in line with the observation made elsewhere that survey response rates are declining in many countries (Bryman, 2015).

4. 4. 4. 4 *Data analysis procedure*

The data were collected online from participants, using Google Docs, and transferred to the Statistical Program for Social Science (SPSS 25). There were no missing data because the respondents were obliged to complete all of the questions in order to sign out. Due to the aim of this step, the analysis procedure was conducted as follows:

- Descriptive statistics in terms of frequency and percentages of the independent variables (gender, job, governorate, and experience) were calculated to show the distribution of the responses.
- The importance of the domains and attributes was tested by calculating the mean and standard deviations.
- Further analyses were used to investigate the difference in the participants' responses for each independent variable. The literature on data analysis indicates that different types of tests depend on the nature and distribution of the data and the number of the compared groups (Cohen et al., 2013). For the gender variable, there were two groups (males and females) and the data showed a non-normal distribution – thus, a Mann-Whitney test was used. A Kruskal-Wallis test was used for the other independent variables (job position, governorate, and years of experience) (Pallant, 2010). The standard assumed for the significant differences was $p < .05$.

The results are presented in the next chapter.

4. 5 Phase two: developing the SJT for Oman

The aim of this phase was to develop the SJTs based on the five non-cognitive domains found in Phase one. This section describes the method, participants, and procedures used in the development process.

4. 5. 1 Method

First, the method used to develop the SJT benefits from the theoretical and the empirical studies reviewed in Chapter 3 (Section 3.4.5). The review revealed no consensus in the literature on how SJTs should be developed, scaled, or scored. However, according to the work of Weekley, Ployhart, and Holtz (2006), the development process comprises five steps: collecting incidents (situations),

developing response options, response instructions, determining the response effectiveness (the answer key), and building the scoring method. A very similar process was used to develop SJTs in the works of Becker (2005), Patterson et al. (2012a), Patterson et al. (2014), Peus et al. (2013) and Sharma et al. (2013). In education, the work of Klassen et al. (2017b) on developing SJTs for teacher selection began by collecting items (scenarios and responses) through interviews with practising teachers, according to determined attributes. The collected items were then reviewed in a one-day workshop with eight experienced teachers. Finally, a concordance panel review, with 11 experts, was used to establish the effectiveness of the responses and build the scoring key.

Building on the literature review and the researcher's knowledge and conditions, the process of developing SJTs in Oman comprised five steps: collecting the incidents, building the response format, revising the items, building the answer key, and establishing the scoring key. These steps are summarised in Figure 4.1 and will next be presented individually and in detail.

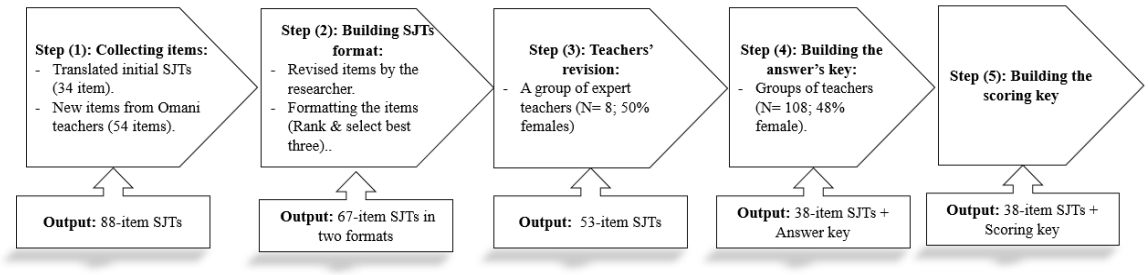


Figure 4.1 Summary of steps in Phase Two for developing the SJT.

4. 5. 2 Participants

To develop SJTs, most research uses a sample of subject matter experts (SMEs); that is people who are experts in the field (for example, incumbents, senior teachers, and supervisors). Here, the participants were working teachers from a number of government schools with grades 5-12. Due to the multi-step procedure and the limited time available for the researcher to collect data, teachers from only one educational governorate were targeted. They were either senior teachers or recommended by their school principals as good teachers. The participation of the nominated teachers varied at each step. Some teachers participated in the collection of incidents, others helped

with the revision of the test, and other groups contributed to building the answer key. The sampling procedure is highlighted for each step in the following points.

4. 5. 3 Step one: incidents collection

In Chapter 3 (Section 3.4.5), there were two main approaches to writing the SJT items: theory-based methods and critical incidents methods. The theory approach involves writing the items to reflect an underlying model or theoretical construct(s). However, the most common approach is the critical incident method, where SMEs are asked to write incidents related to performance on the job (Weekley, Ployhart, & Holtz, 2006). The critical incident approach was taken here to building the content of the SJTs. The collection of the incidents included simultaneously collecting stems and responses. The next points concern the instrument used, the participants, the procedure, and data analysis.

4. 5. 3. 1 ***Instrument***

Targeting the five domains found in Phase one, two main sources were used to collect the incidents.

The SJTs built by Klassen et al. (2014b) were first translated from English to Arabic by a bilingual translator. The researcher revised the translations and made some necessary amendments to suit the context in Oman (using Arabic names for teachers and pupils in the scenarios and responses). Thirty-four items were translated in order to measure three non-cognitive domains (communication, resilience and adaptability, and planning and organisation).

Based on the translated incidents, new incidents and responses from current teachers in Oman were collected. To collect the incidents, the researcher prepared a booklet (in Arabic) as guidance for the participants. This included a consent form, a brief description of the study (the aims and the steps), a list of the five non-cognitive domains and their definitions, examples of items and responses, and a form on which to write the incidents. Appendix 7 presents the form prepared for teachers to detail the incidents.

4. 5. 3. 2 ***Participants***

At this step, the participants were a group of senior teachers in schools with grades 5-12, from one educational governorate in Oman. They were recruited using a

snowball sampling procedure. This is a type of non-probability sample where the researcher contacts a small group of people, related to the research topic, and these people contact others (Bryman, 2015). To begin, a small number of senior teachers were contacted by telephone and asked to participate in the research. With their agreement, the instrument was sent to them by email and they were asked to contact other senior teachers and encourage them to participate. Ultimately, teachers (males and females) from 13 schools agreed to participate. The collected incidents were written anonymously. Each participant could write one or more items; thus the number of participants was not counted.

4. 5. 3. 3 ***Procedure***

The procedure for collecting the data (the incidents) was conducted remotely. As previously noted, the researcher prepared a form to collect the incidents and sent it to a number of senior teachers. One participant was appointed to voluntarily collect the written forms from the participants and to send them anonymously to the researcher by email or phone (as an image). Appendix 8 provides examples of the collected forms. This process took a number of weeks, and 54 new items were ultimately collected from the Omani teachers.

4. 5. 3. 4 ***Data analysis procedure***

Eighty-eight situations (stems and responses) were collected from the translated SJTs and the participants in Oman. The new collected situations from Oman were revised by the researcher. Some incidents were rewritten to match the original items in the translated SJTs in terms of length and complexity. Other items were deleted because of their similarity to other situations and/or their unsuitability to the context in Oman or for new teachers. By the end, 67 items (stems and responses) remained, and these formed the initial SJT. The analysis procedure for these items was based on the researcher's knowledge, with further analysis conducted by expert teachers in the following steps.

4. 5. 4 Step two: response format/instruction

The 67 items found in the previous step were written in two formats: 'ranking' and 'select best three'. This format followed the format used in the initial test built in the UK. For the 'ranking' format, the following instruction was given: 'In this part, you have a number of situations, followed by five options for each. Rank the five

options from the most appropriate (by giving number 1) to the least relevant (by giving number 5). For example, if option c is the best, c = 1; if b is the next best option, b = 2; and so on'. In contrast, the 'select best three' format was guided by the following instruction: 'In this part, you have a number of situations, followed by a number of options. Choose three options that represent the best ones for dealing with the situation, marking these with (√). The order of options is not important'.

There were 19 items in the 'ranking' format, and 48 items in the 'select best three' format. A list of the 67items was prepared for revision by expert teachers in the next step.

4. 5. 5 Step three: items' revision

The main aim of this step was to revise the 67-item SJT with the help of a group of expert teachers. The revision process had two goals. Firstly, the participants were asked to check the clarity and suitability of the items for the educational context in Oman and for use in the ITEP admission process. Secondly, teachers were asked to match each item with a suitable domain(s) (in effect, what is the item supposed to measure from the five given domains?).

4. 5. 5. 1 ***Instrument***

The 67-item SJT prepared in the previous step was used as the main instrument. It was presented, in Arabic, to the participants as a hard copy in written format. Alongside the SJTs, a consent form and answer sheet were prepared, as shown in Appendix 9.

4. 5. 5. 2 ***Participants***

At this step, the researcher targeted a small number of teachers to revise the items. A convenience sample of eight teachers (four males and four females) from four schools was recruited. The teachers were recommended as good teachers by their schools' principals.

4. 5. 5. 3 ***Procedure***

Firstly, the researcher visited each school principal, explained the aim of the study, and asked the principal to nominate two good teachers. The study and the nature of the participation were explained to the nominated teachers at each school. Once they had agreed to participate, each teacher was provided with the instrument.

The teachers revised each item in terms of clarity, suitability for the context in Oman, the options, the suitability for new teachers, the related domain(s), and the possible answers.

Teachers worked individually over six days. It was not possible to hold a group discussion with the teachers due to the researcher's limited time and the conditions of the teachers. Hence, the researcher worked alone to review the feedback of the teachers.

4. 5. 5. 4 ***Data analysis procedure***

The eight answer sheets were collected from the participants and analysed in line with the aim of this step. First, each item was revised according to the participants' feedback on each of the determined criteria (clarity, suitability, and so on). The results indicated agreement between the teachers on the appropriateness of the 53 items: 15 in the 'ranking' format and 38 in the 'select best three' format. Thirteen items were removed due to the participants' remarks. The participants' classifications of each item according to the five domains were analysed. The findings are presented in the next chapter.

4. 5. 6 Step four: building the answer key

The literature on SJTs states that there is no one correct response to the situations. The effectiveness of the responses is mostly determined using the rational key which recalls the judgment of the experts (Weekley, Ployhart, & Holtz, 2006). Therefore, a group of working teachers built the answer key for the developed SJTs. The instrument, participants, procedures, and data analysis are explained in the next points.

4. 5. 6. 1 ***Instrument***

The 53-item SJTs was then used. To save the participants' time, each teacher was tested with half of the 53 items (some teachers answered 26 items and others 27). The test was conducted in Arabic and in a paper format. On each test paper, the participants were first asked to give their gender and their date of appointment (years of experience). The test began with a general statement:

> 'As a good skilled and experienced teacher, your point of view on the correct responses of a new teacher to the following educational situations are of

interest. There are no right or wrong answers. However, your answer should be credible because it will represent the answer key in the actual application phase of the exam.'

The items in the 'ranking' format were then presented, followed by the items in the 'select best three' format.

4. 5. 6. 2 ***Participants***

A convenience sample of teachers (N = 108) from 10 schools, who did not participate in the previous steps, were asked, voluntarily, to participate. Of the participants, 48.1% were female and their mean work experience was 12.9 years (range 3-27). Participants were recruited via recommendation from their school principals as good teachers.

4. 5. 6. 3 ***Procedure***

The researcher visited each school twice. During the first visit, an explanation of the research was given to the school principal, who was then asked to nominate a number of good teachers (about 10-12) to undertake the test at a convenient date and time. At the time of the appointment, the researcher visited the school again and explained the research and the test to the nominated teachers. With their agreement, the instrument was distributed to the participants. The time afforded to answer the test questions was unspecified, but most teachers finished in 30-40 minutes. Although the goal of this step was to build the answer key, teachers were asked to give suggestions and feedback (if any) on the situations in terms of clarity or relevance, by writing comments beside the item(s).

4. 5. 6. 4 ***Data analysis procedure***

Following the rational model of seeking consensus (Weekley, Ployhart, & Holtz, 2006), all the participants' responses were entered by the researcher into a computer program (Excel, 2016). The answers for each option were summarised and calculated as percentages, by gender. The numerical figures did not show high agreement between the participants, thus for some items with low agreement, other comparison factors were considered when building the answer key. More details on the findings and the limitations are presented in the next chapter.

At the end of the data analysis stage, the answer key for 38 items was built (14 in 'ranking' and 24 in 'select best three'). The 38-item SJT was used in the next phase, after the researcher had made changes to the order of the questions, on the basis of the participants' suggestions. The 'select best three' format items were considered to be easier to answer and so were moved to the beginning of the test.

4. 5. 7 Step five: scoring key

After defining the most likely right answers in the previous step, the aim was then to determine the scores (weights) for the responses to the test. The research on SJTs has used different methods of scoring to enhance the psychometric properties of the tests. However, in their overview of the current research, Whetzel et al. (2009) indicate that, 'There is insufficient research to judge one scoring key to be substantially better than another. More research is clearly needed in this area' (p.196).

The scoring key used in this study was based on that implemented in selection for medical schools (Metcalfe & Dev, 2013) and, recently, in the work of Klassen et al. (2017b), as follows:

- For the 'ranking' questions, responses were scored according to their closeness to the answer on the answer key. The participant received four points for his/her answer on each option if it equalled the answer key ranking, three points if the difference between them was one, two points if the difference was two, one point if the difference was three, and zero points if the difference was four. The participant was awarded 20 points for each item if all options were in the right order. Thus, for the 14 ranking items, the total possible score was 280. Table 4.4 provides an illustration of the scoring process for the 'ranking' items.

Table 4.4 An example of the scoring process for the 'ranking' questions

The correct ranking	Score for possible answers				
	If ranked 1st	If ranked 2nd	If ranked 3rd	If ranked 4th	If ranked 5th
1 = B	4	3	2	1	0
2 = A	3	4	3	2	1
3 = C	2	3	4	3	2
4 = E	1	2	3	4	3
5 = D	0	1	2	3	4

- For the 'select best three' items, the participant scored four points for each correct option and zero points for the wrong option. No negative marking was used. Thus, for the 24 items in this form, the total possible score was 288.

4. 6 Phase three: piloting the SJTs

In Phase two, the SJTs were developed and revised by practising teachers to screen applicants for the ITEP. This phase explored the appropriateness of the developed test by piloting it with a group of new students in one ITEP in Oman. The literature indicates that piloting can be done in several ways: (a) by a small group of experts who check the items in terms of their suitability, relevance, validity, possible cultural bias, and remoteness from the test-takers' experiences; (b) by a small group of test-takers who give feedback on the items in terms of clarity, readability, difficulties in the wording, the format, and the time taken; or (c) by a large group of test-takers (Cohen, Manion, & Morrison, 2013).

The piloting in this phase was conducted with a group of new ITEP students in Oman and sought to determine the SJT's potential for implementation with another sample in Phase four. In specific, the piloting was conducted to (a) analyse the initial properties of the developed SJTs (distribution of the data and internal consistency); (b) examine the quality of the items; and (c) explore the initial applicants' reactions to the test and its suitability for the selection process in future. The instrument, participants, and procedures are explained in the next sections.

4. 6. 1 Instrument

The 38-item SJT developed in Phase two was then used. It was conducted in Arabic and presented in a paper-based format. It consisted of four parts: (i) personal information (gender, subject, and date of birth), (ii) the 'select best three' items (24 items with six options for each), (iii) the 'ranking' items (14 items with five options for each), and (iv) an open-ended question to obtain feedback from the participants about the test ('Kindly, give your point of view in terms of the test's suitability for use with ITEP applicants, specifically to ensure the selection of the best possible future teachers. Please also give any other comments that you have on the test'). The total possible score was 568 points.

4. 6. 2 Procedure

The pilot was conducted during the induction week at the beginning of the academic year 2016/2017. As mentioned in Chapter 2, there were two main ITEP providers in Oman in that year. The researcher contacted one of these, which offered to take part in the implementation phase (Phase four). However, they later apologised due to the demands of the numerous other activities scheduled during the induction week. As a consequence, the second ITEP was contacted. A letter, including the instrument, was presented to the administration unit to request official approval. Once approval had been granted, the participants were invited to voluntarily and anonymously take the test as a group at a predetermined time. Before taking the test, a brief description and a consent form were given by the researcher to the participants. An unspecified amount of time was given for the test, but most participants finished in approximately 40-50 minutes.

4. 6. 3 Participants

The targeted population at this step was the new undergraduate entrants to the ITEPs in Oman. As noted in the procedure, there were two government colleges offering an undergraduate ITEP in 2016/2017. According to the Higher Education Admission Centre (HEAC), there were 804 new students enrolled on the education courses for this academic year, of whom 56% were female. Students on the targeted ITEP, and at one of the scheduled activities in the induction week, were asked to take part in the research. A convenience sample of first year students agreed to participate. The total number of participants was 171, of whom 53.4% were female. The mean

age of the sample was 18 years (range 17-19). The sample's subjects were biology (33.3%), chemistry (22.2%), physics (19.4%), maths (23.6%), and English (0.7%).

4. 6. 4 Data analysis procedure

First, the test papers were collected and reviewed by the researcher. Twenty-seven papers were removed because they had missing data (in effect, no answers given for one or more of the SJT items). Hence, the data of 144 participants (56.3% female) were analysed. The participants' answers to the SJT items were entered, scored, and analysed, using the Statistical Program for Social Science (SPSS version 25). Bearing in mind the aim of the pilot, the analysis process included the following steps:

- The participants' answers were scored according to the answer and scoring keys built in Phase two.
- Descriptive statistics (means, standard deviations, and percentages) were obtained in order to explore the distribution of the data and to test the differences in the responses by gender and type of question. The standard assumed for the significant differences was $p < .05$.
- Cronbach's alpha was calculated to determine the internal reliability of the SJTs.
- Items' analyses, using partial correlation and items' difficulty, were used to examine the quality of the items, and consequently to produce a shorter version of the test to be implemented in the next phase. A factor analysis procedure was also conducted.
- A thematic analysis procedure was used to analyse the participants' feedback on the open-ended question.

The results are presented and analysed in the next chapter.

4. 7 Phase four: implementing the SJTs

After building the SJTs in phases one and two, and piloting the test in Phase three, the aim of Phase four is to explore the test's reliability and validity and applicants' reactions to it (in effect, answering Research Question 2). This section identifies the measures, data collection procedure, participants, and data analysis procedure used in this phase. The findings are presented in the next chapter.

4. 7. 1 Measures

Based on the goal of this phase and the findings of the literature review, three measures were distributed to the participants: the SJTs, the Big Five Inventory (BFI), and the applicants' feedback. In addition to these three measures, the interview scores obtained during the admission procedure and students' academic scores in terms of their cumulative grade point average (GPA) were obtained, along with the ITEP authorisation from the administration unit. The correlation between the participants' scores in the SJTs and the other external measures was used to explore the validity of the developed SJTs (Jackson et al., 2016). Details of those measures are as follows:

- The 29-item SJT developed in Phase three was used. It was presented in Arabic and in a pencil and paper format. The first section sought personal data about the participant: student number (academic number), gender (male/female), year of study, and subject. The student's number was necessary to obtain the interview score and GPA; and the gender of the participants was needed to calculate the differences in responses from males and females, and consequently to examine the fairness of the SJT. The 'year of study' was included to find the targeted sample (students in years one and two). The last variable was the 'subject', which allowed exploration of the variety of the participants. The second section was the SJTs items, which was divided in two parts: (i) part one – 'choose best three', with 15 items and six options for each; and (ii) part two – 'ranking', including 14 items with five options for each. The total score of the SJTs was 460 points (180 points for the 'select best three' and 280 for the 'ranking' items).
- The BFI, or 'five-factor model', is considered a comprehensive model of personality traits (Costa & McCrae, 1992) and has been used in many studies to measure the construct validity of the SJT (Chan & Schmitt, 2002; Lievens et al., 2008). Construct validity ensures that the performance of a test is fairly explained by appropriate constructs or concepts, with comparisons made of measures with similar constructs (Cohen, Manion, & Morrison, 2013). Hence, to explore the construct validity of the SJTs in Oman, participants were asked to complete an Arabic version of the 60-item BFI implemented in Oman by Kazem (2002), measuring neuroticism, extroversion, openness, agreeableness, and conscientiousness. It 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.

- The interview scores of the participants at the admission process were obtained from the ITEP, with the aim of measuring the concurrent validity of the developed SJTs. According to Cohen, Manion, and Morrison (2013), concurent validity is found by correlating results between tests assessing the same performance. 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.
- GPA is a measure of students' academic achievement. 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). Although the SJTs in this study were designed to measure non-cognitive attributes, the research indicates that SJTs have underline cognitive power and are correlated with cognitive ability (for example, GPA) (McDaniel et al., 2001; Patterson et al., 2013a). Hence, a correlation with participants' GPA can indicate the validity of the SJT. GPA score ranges from zero to four points.
- Participants' feedback on the SJT was sought using an Arabic translation of the measure developed by Klassen et al. (2014b) to explore applicants' reactions to the SJTs. The translation was completed by the researcher and revised by a bilingual academic in Oman. It contained seven items measuring participants' evaluation of the SJT as a measurement, as well as its content in terms of relevance, difficulty, fairness, differentiation, and appropriateness. The participants were asked to show their level agreement with each item, choosing one of the five options: (1) strongly disagree, (2) disagree, (3) neutral, (4) agree, and (5) strongly agree. In addition, participants were given

the opportunity to provide any comments about the SJT in an open-ended question:

> 'Kindly provide any other comments you may have about this test in terms of both its suitability for use within the ITEP admission procedures and a comparison with the current selection tools (secondary school results, the admission interview, and so on), as well as any other observations you wish to make.'

4. 7. 2 Procedure

Data collection took place February to March 2017, at an ITEP in Oman (not the one used in the piloting phase). The researcher met the teachers of the targeted students, explained the aim of the study, and arranged a mutually convenient time for the data collection. At the time of the agreed visit, the aim of the study and details of the consent form were explained and distributed to the students. With the students' agreement, the measures were distributed to the students in the following order: the SJT attached to the applicants' feedback, followed by the BFI paper. The maximum time allowed was one hour, but most students finished within 40 minutes. The participants provided their university numbers for the three measures, allowing the researcher to match the scores with their marks in the admission interview and the cumulative GPA.

4. 7. 3 Participants

The targeted participants were recent entrants to the ITEP. However, the academic year had already begun, and it was not possible to find classes limited to new students in education. In addition, when students enrol, they usually undertake a foundation year and are grouped according to their grades on the admission tests in maths, English, and information technology – regardless of their programmes (education or non-education). Therefore, the researcher targeted undergraduate educational students in their first or second year. These students were still completing their foundation year and had not been involved in any teaching practice. Hence, and with assistance from the ITEP administration unit, four classes were identified where most students matched the targeted sample of the study.

There was a total of 142 participants in the four classes, of whom 74% were female. The majority of the participants (87%) were in their first or second year.

However, the data showed that there were 18 participants in their third, fourth, or fifth year. Almost all subjects were represented in the sample, with one-third being English students (33%). Details of the participants are presented in Table 4.5.

Table 4.5 Distribution of the participants by the independent variables

		Frequency	%
Gender	Female	105	73.9
	Male	37	26.1
	Total	142	100.0
Year of study	1st year	66	46.5
	2nd year	58	40.8
	3rd year	12	8.5
	4th year	5	3.5
	5th year	1	.7
Subject	Missing	3	2.1
	Arabic Language	12	8.5
	Art	20	14.1
	Child Before School	3	2.1
	English	47	33.1
	Information	6	4.2
	Math & Science	22	15.5
	Physical Education	9	6.3
	Religious Education	20	14.1

4. 7. 4 Data analysis procedure

To accomplish the goals of this study, data were entered in the SPSS and analysed according to the following concerns:

- The distribution of the SJTs scores were analysed in order to examine the difficulty of the test and its ability to differentiate between the applicants using descriptive statistics (means and standard deviations), histograms, and normality tests.
- The group differences, by gender and year of study, were analysed using the independent sample t-test and tested by Cohen's *d*. The independent samples t-test was used to investigate if there was a statistically significant difference between the two groups in the total SJT scores. In other words, this compared

the mean scores in order to discover the extent to which we can claim that the difference between groups is real or due to sampling error by looking to the '*p*' value. A '*p*' value below .05 indicates a significant difference, which means that there is a low probability of it occurring by chance. However, the '*p*' value can be affected by the size of the sample, where a small difference in large samples can be significant. Hence, Cohen's *d* was used to estimate the difference by quantifying it in standard deviation units. A '*d*' value of .2 represents a small effect, whereas .5 and .8 represents medium and large effects, respectively (Pallant, 2010).

- The reliability of the test in terms of internal consistency was examined using Cronbach's coefficient alpha, which measures the inter-item correlations. The literature suggests a value of .7 and above has good reliability (Cohen, Manion, & Morrison, 2013). The same suggestion was used here to explain the results.
- The validity of the SJT was analysed by correlating the SJTs' scores with other criterion measures used in the study, using Pearson's correlation coefficient (*r*). Correlation coefficients give information about the strength and the direction of the relationships between two variables, and is arranged between +1 and -1. As a guideline, *r* between .1 and .29 represents a small positive correlation, *r* = .3 to .49 is a medium positive correlation, and *r* = .5 to 1 shows a large positive correlation (Pallant, 2010). Construct validity was tested by correlating SJT scores with the scores in the BFI factors, with scores in the interview and GPA used to explore the criterion-related validity. In each, the correlations for males and females were also calculated. The significance of the different correlations between males and females (in effect, *z* value) were tested using an online calculator (http://vassarstats.net/rdiff.htmal). The calculation of the *z* value allowed the researcher to assess the likelihood that the gender difference could have been due to chance. The difference is not statistically significant if the obtained *z* value is between -1.96 and +1.96 (Pallant, 2010).
- The applicants' reactions were explored by analysing the participants' feedback on use of descriptive statistics (frequency and percentage). In addition, the qualitative data from the open-ended question were analysed

using thematic analysis procedure, looking at themes repeated in the responses (Bryman, 2015).

The results are presented and analysed in the next chapter.

4. 8 Ethical considerations

Research in social science raises ethical questions from the beginning (for example, is it a good idea to conduct this research?), during the research (how will the participants be treated?), and after the research has been completed (how will the research be reported and presented to the public?) (Oliver, 2010). Bryman (2015) states that there is a growing concern around how to deal with these ethical issues because of their direct relation to the integrity of the work. In addition, the expansion of the field of research, not just in terms of the number of studies and researchers, but also the diversity of the methods and contexts, has highlighted the need for professional associations and committees to formulate codes of ethics (for example, the British Educational Research Association (BERA)).

The current study took into account the ethical issues agreed by the Education Ethics Committee at the University of York, where this study was carried out. The agreed ethical demands of educational research include the presentation of a written informed consent for participants to sign, an understanding of the impact of the study on the participants, and awareness of that impact throughout the process of handling the data and writing up the research. For this reason, the Education Ethics Committee established an ethical issues audit form to be completed by the researcher and approved by the supervisor (or other concerned bodies) before the collection of the data (Department of Education 'University of York', 2018). The ethical procedures implemented in this study follow that procedure, as explained below.

Before commencing the data collection, an ethical issues audit form was filled in by the researcher and approved by his supervisor. Additionally, two official letters confirming the researcher's status and giving a brief overview of the study were obtained from the supervisor and the Omani embassy in the UK (the sponsor). A document including a letter from the researcher, the two official letters, a summary of the study (objectives and research methods), the instruments, and the requirements, was then sent to the targeted organisations in Oman (the MoE and the ITEPs authority). Official permission was obtained from both the MoE and the two ITEPs,

allowing the researcher to access the official documents and collect data using the instruments (the interview, the questionnaire, and the SJT) from the targeted samples at each phase of the study. Appendix 10 provides a copy of the official permission (in Arabic).

During the data collection process, different informed consent forms for the participants were used for all of the steps. The forms gave the participants details of the aim of the study, the steps of the study, the nature and timeframe of their participation, their option to withdraw from the study during or after the data had been collected, the process of saving the data, and the anonymity of their responses. The researcher considered the ethical issues at each step of the study. The access to official documents, for example, was permitted by the concerned bodies. The ethical issues regarding use of documents concern the public use of the documents and ensuring that their use does not harm the organisation (Marshall & Rossman, 2014). Hence, some of the obtained documents were analysed but not explicitly included in the final research. Moreover, the ethical issues considered during the interview procedure included the importance of trust, good listening, and protecting the identity of the interviewees during the writing up of the research. Finally, participants in Phase four were informed that their GPA and interview scores would be obtained and used anonymously for the purpose of the study.

Chapter 5 Results

The aim of this chapter is to present the results of the research process described in Chapter 4. The chapter comprises four sections. First, the results from Phase one of the study are gathered by analysing the data from the official documents, the interview, and the questionnaire. The second section presents the findings from the development process for the Situational Judgment Test (SJTs) in Phase two. The results of the pilot study (Phase three) and the implementation (Phase four) are then presented and analysed. Finally, a summary is given of the main results of the four phases of the study.

5. 1 Results of Phase one: non-cognitive attributes/domains in Oman

The aim of Phase one was to find the essential non-cognitive attributes for applicants to the initial teacher education programmes (ITEPs) in Oman. These attributes were then used for developing SJT that can be used in the admission process. To achieve that aim, data were collected by following complementary steps, including analysis of the related official documents, and an explorative interview and questionnaire with some of the stakeholders (college tutors, teachers, supervisors and school principals). The results from the official documents were presented and followed by findings from the interviews. Results from the two steps were compared with the findings of Klassen et al. (2014b) in the UK. Finally, the results of the questionnaire were presented and analysed. The last point gives a summary of the main findings.

5. 1. 1 Results from the official documents

Tasks and criteria included in three documents from the Ministry of Education (MoE) and two documents from an ITEP, in Oman, were analysed. The analysis of the three documents from the MoE produced 22 attributes. The first document was the teachers' job description, including 39 tasks and responsibilities, which produced seven groups of attributes. For instance, some tasks required teachers to show commitment to the profession and discipline, whilst others needed planning skills. The classroom assessment and annual evaluation documents were also analysed, with

six and nine attributes found, respectively. On the other hand, the analysis of two documents from the targeted ITEP identified nine attributes.

In general, 31 attributes were found during the analysis of the five official documents. The necessary attributes of the working teachers (from the MoE documents) were generally focused on commitment, discipline, planning, self-development, and self-assessment. On the other hand, the documents from the ITEP for student-teachers stressed ethical values and attitude towards the profession. Good communication skills were seen as important for both. The final list of attributes is illustrated in Table 5.1.

The analysis did not produce precise results because, in some tasks, it was difficult to separate the cognitive and the non-cognitive attributes required for specific tasks. However, a review of the literature, as well as the researcher's own experience, was used to make better decisions (included or excluded). Moreover, this limitation could be overcome, as this was complemented by other steps (the interview and questionnaire).

Table 5.1 Summary of the attributes found from official documents in Oman

Source	Document	Attributes
Ministry of Education	Job description	Commitment (committed to the profession's ethics and the job roles)/ Discipline. Loyalty (strengthens the national and job loyalty). Planning & organization. Cooperative (with school, peers, parents and the community)/ involved in school activities. Care of pupils/ advise pupils for good attitudes/ attitudes about pupils with Special Education Needs (SEN). Self-development. Self-assessment.
	Classroom supervision criteria	Planning. Raise pupils' motivation. Management (effective classroom management - time management). Direct pupils for self-learning. Develop positive attitudes and values. Self-assessment.
	Teachers' annual appraisal	Care about his/her appearance. Accept advice and feedback. Good relationship with school, peers, pupils and parents. Strong personality and class management. Innovation in work. Social activities inside and outside school. Self-development. Planning & organization. Commitment & discipline.
The ITEP	Candidate's Proficiencies & the interview criteria	Observes Omani, Islamic and professional ethics/values in performing his/her professional tasks. Develops positive attitudes towards the profession and contributes effectively to it. Collaborates with schools, families and community to support student learning. Show strong and reliable concern to be a teacher. Good attitudes to work with SEN pupils. Show high consideration to the Islamic and Omani values. Show good problem-solving and decision-taking skills. Good communication skills (oral and non-oral communication, eye contact, active listening). Has a professional appearance and behaviour.

5. 1. 2 Results from the interview

Although the interview aimed to explore the non-cognitive attributes necessary for prospective teachers, it also explored the participants' perspectives (N = 8) on the effectiveness of the current selection procedure. Thus, the findings in this step are classified according to the interview's five questions.

Firstly, the participants were asked about the necessary non-cognitive attributes for prospective teachers. The participants from the ITEP (N = 2) referred to the proficiencies in the college's theoretical framework. One tutor said that *'the necessary non-cognitive attributes can be obtained from the College's Conceptual Framework, especially in the theme of dispositions and values'*. They also stressed the skills of *'time management*', *'taking responsibility'*, and *'honesty'* for student-teachers in the ITEP. On the other hand, school principals and supervisors (N = 6) showed a strong concern for teachers' motivation, enthusiasm, and ethics (honesty, fairness). A supervisor said: *'humanity is very important for teachers when dealing with pupils. Teachers must look at pupils as humans with different needs not just as learners'*. One school's principal said: *'teaching is not just a job but a profession- teachers must have values such as fairness, honesty and be objective'*. In general, the attributes most commonly named by the eight participants were *'motivation'*, *'enthusiasm'*, *'positive attitudes towards the profession'*, and *'desire for professional development'*. They also stressed ethics and morals, such as *'fairness'* and *'honesty'*.

The second question was about the availability of the necessary non-cognitive attributes among new students in the ITEP or the new teachers in schools. The participants, in general, gave negative responses. From the tutors' perspective, new students in the ITEP mostly had weak non-cognitive attributes. For example, one tutor said that *'new students lacked motivation, especially the male students'*. The teachers' supervisors were also concerned about the poor performance of new teachers in the non-cognitive attributes, especially the *'attitudes towards teaching as a profession'* and *'the belief about students' ability to learn'*. One supervisor said: *'most of new teachers can not control their anger with some pupils' behaviour- they have no resilience when dealing with pupils' needs'*. From the school principals' perspectives, most current new teachers also had weaknesses in the areas of *'enthusiasm'* and *'professional ethics'*, such as honesty and conscientiousness.

In the third question, the participants were asked about the ability of the current selection procedures in the ITEP, in Oman, to measure the non-cognitive attributes of the applicants. There was a consensus that the current selection process tools were insufficient to measure non-cognitive attributes as their focus, mainly for the applicants' secondary school results. The college tutors said similar things but these are just illustrative quotations from individuals: *'for most subjects, the decision to accept students on the ITEP was made at the national level by an external unit (Higher Education Admission Centre (HEAC))'* and *'the admission process interview at the college was insufficient to exclude applicants with low performance, except in two subjects: physical education and arts. Applicants to these two subjects are set an aptitude test, and those who fail are not accepted. However, the interview results for the applicants to the other subjects were not able to change the acceptance decision if the applicants had met the programme's academic requirements. There was one case of a student stating explicitly, during an interview, that he did not want to become a teacher, and the response was simply to make a note of this in the recommendation*'. The school principals and supervisors agreed on the weakness of the current selection practices. In addition, they thought that the ITEP preparation programme was not enough to develop the necessary non-cognitive attributes for the prospective teachers.

The fourth question asked the participants about any assessment tools used by the ITEP and the schools to evaluate the non-cognitive attributes. One tutor said: *'the college has started developing and assessing students' performance in both cognitive and non-cognitive programmes, according to the themes in the college's theoretical framework'*. On the other hand, the school principals and supervisors affirmed that teachers were assessed through certain criteria used in the class visit. However, they both expressed low satisfaction with the sequences of such evaluations, as they have little influence on teacher performance.

Finally, the respondents to the open-ended question were asked to add any comments they wanted to make on the subject. Three participants made no comments. Others expressed a strong need for a better selection procedure for prospective teachers in the ITEPs. They highlighted the need to measure applicants' attitudes towards being a teacher and the extent to which they have the necessary non-cognitive attributes to be a good teacher. One participant felt that *'female teachers had better non-cognitive attributes than males'*.

The focus of the analysis procedure was highlighting the non-cognitive attributes mentioned throughout the questions in the eight interviews. Twenty-five attributes were explored, and these are presented in Table 5.2.

Table 5.2 Summary of non-cognitive attributes found from the interviews

Attributes	College tutors	Supervisors	School principals
Positive attitudes towards teaching - Proud of his/her career - passionate about teaching	√	√	√
Responsible	√		
Time management	√		
Honest – reliable	√	√	
Social worker - care of the community	√		
Communication	√	√	
Motivation	√		
Self-assessment	√		
Fairness		√	√
Wise in solving problems		√	
Flexible		√	√
Self-development	√	√	√
Show humanity (with pupils, peers, parents...)		√	
Planning & organisation		√	
Enthusiasm		√	√
Patient		√	
Believe about students learning	√	√	
Has religious faith – conscientious			√
Has the spirit of the citizenship	√		√
Initiative			√
Leadership - effective classroom management			√
Confident - not shy			√
Loyalty			√
Personality traits (in good health)			√
Being as a model for pupils – inspiring			√

5. 1. 3 Initial domain vs. Klassen's framework

The previous two steps produced a list of 56 attributes: 31 from the official documents and 25 from the interviews. Before developing the questionnaire to explore the importance of these attributes with a large sample of teachers, school

principals, and supervisors, two analysis processes were completed. Firstly, those attributes that were classed as similar were aggregated – this produced a second list of 43 attributes. Secondly, the attributes were compared with those in the three UK domains identified by Klassen et al. (2014b), namely: empathy and communication; resilience and adaptability; and planning and organisation.

The 43 attributes found from the Omani context were put into a matrix, along with the attributes of the three domains found by Klassen, as seen in Appendix 5. This process resulted into the following:

- Some of the attributes found in Oman (18) matched the definitions of the three domains in the UK. However, the first of these ('empathy and communication') was changed to 'communication skills', as a result of the interview process. Some of the participants rejected the combination of 'empathy' and 'communication' in terms of the relationship between teacher and pupils. This argument will be discussed further in the discussion chapter.
- The remaining attributes (25) seemed not to match the three domains, hence they were further analysed.
- Some of the remaining attributes related to teacher motivation and enthusiasm, whereas others were associated with the ethics and morals of teaching. Hence, two new domains were established: 'enthusiasm and motivation' (nine attributes) and 'professional ethics' (11 attributes). Analysis of the official documents and the interviews revealed the importance of these two domains for the quality of teachers in Oman. The literature also supported those findings (see the Discussion chapter).
- Five attributes were removed: self-assessment, being in a good health, care for appearance, strong personality, and professional appearance. These did not fit into any of the five domains, and some could be developed or measured by the current tools (interview, medical check).

The completion of this step resulted in an initial framework, with five key non-cognitive domains and 29 attributes. The importance of this initial framework was tested in the next step by distribution of a questionnaire to a large sample of teachers, school principals, and supervisors.

5. 1. 4 Results from the questionnaire

The respondents to the questionnaire were 181 working teachers, school principals, and supervisors from three educational governorates. Of these, 58% were female and the mean work experience duration was 16.4 years (*SD* = 7.98). The questionnaire was distributed in an electronic format, and all of the questions had to be answered in order to save the data. Thus, there were no missing values. Table 5.3 shows the distributions of the participants by gender, job, governorate, and work experience.

Table 5.3 Participants' distribution by gender, job, governorate and experience

		Frequency	%
Gender	Female	104	57.5
	Male	77	42.5
	Total	181	100.0
Job	School Principal	71	39.2
	Supervisor	60	33.1
	Teacher	50	27.6
Governorate	Batinah South	81	44.8
	Sharqya North	42	23.2
	Muscat	58	32.0
Experience	≤ 5 years	17	9.4
	6-10 years	18	9.9
	11-15 years	47	26.0
	16-20 years	53	29.3
	> 20 years	46	25.4

The aim here was to measure the participants' rating of the level of importance of the five domain and their related attributes to the effectiveness of the school teacher and the ITEP applicant, in a ten-point scale (1 = not important, 10 = very important). Therefore, the descriptive statistics were calculated, namely the means and standard deviations (*SDs*) of the responses. To investigate the difference between the responses according to each independent variable (gender, job,

governorate, and experience), mean and standard deviation were also calculated and analysed for each.

Firstly, Table 5.4 reveals that the five domains were generally seen as important for the teacher 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 ITEP applicants, ranging from 8.92 for 'resilience and adaptability' (*SD* = 1.71) to 9.24 for 'enthusiasm and motivation' (*SD* = 1.65).

Table 5.4 Means and standard deviations of all Responses

		Communication skills	Organisation & planning	Resilience & adaptability	Enthusiasm & motivation	Professional ethics
Effective Teacher	Mean	9.48	9.46	9.21	9.41	9.53
	SD.	1.48	1.50	1.60	1.56	1.54
Effective Applicant	Mean	9.02	8.96	8.92	9.24	9.20
	SD.	1.73	1.77	1.71	1.64	1.73

Regarding male and female perceptions of the importance of the five domains for teachers and applicants, the data in Table 5.5 shows that they were highly rated by both. The females' ratings were higher for all variables (with a mean above nine), the mean for the male participants was less than nine (ranging between 8.48 and 8.82) on the importance of the domains for the applicants. The values of the standard deviations (*SDs*) were smaller for females than for males which indicated less variance in responses between the female participants.

Table 5.5 Means and standard deviations of the perceptions by gender.

			Communication skills	Organisation & planning	Resilience and adaptability	Enthusiasm & motivation	Professional ethics
Effective Teacher	Female	Mean	9.70	9.73	9.51	9.73	9.86
		SD.	.98	.94	1.08	.96	.91
	Male	Mean	9.18	9.09	8.81	8.97	9.09
		SD.	1.93	1.98	2.04	2.05	2.04
Effective Applicant	Female	Mean	9.30	9.31	9.23	9.56	9.51
		SD.	1.45	1.47	1.48	1.28	1.31
	Male	Mean	8.64	8.48	8.51	8.82	8.79
		SD.	1.99	2.03	1.92	1.97	2.12

To investigate if there was a statistically significant difference between males and females, a Mann-Whitney test was used as the responses were non-normally distributed. The data in Table 5.6 indicate that, for working teachers, males and females differ significantly on 'organisation and planning' ($p = .048$), 'resilience and adaptability' ($p = .017$), 'enthusiasm and motivation' ($p = .003$), and 'professional ethics' ($p < .001$), but not 'communication skills' ($p = .209$). The males and females' responses regarding new applicants indicated significant differences on the all five domains.

Table 5.6 Mann-Whitney Test for Gender

		Communication skills	Organisation& planning	Resilience & adaptability	Enthusiasm & motivation	Professional ethics
Effective Teacher	Mann-Whitney U	3706	3519	3295	3247	3220
	Wilcoxon W	6709	6522	6297	6250	6223
	Z	-1.26	-1.98	-2.39	-2.97	-3.96
	Asymp. Sig. (2-tailed)	.209	.048	.017	.003	.000
Effective Applicant	Mann-Whitney U	3240	2993	3004	3055	3238
	Wilcoxon W	6243	5997	6008	6058	6241
	Z	-2.50	-3.25	-3.13	-3.36	-2.71
	Asymp. Sig. (2-tailed)	.012	.001	.002	.001	.007

Difference in responses by job were also examined. The data in Table 5.7 show that the mean of the five domains was above nine for all three groups (teachers, supervisors, and school principals), except in the cases of 'communication', 'organisation and planning', and 'resilience and adaptability', as seen by supervisors of the new applicants; and the case of 'resilience and adaptability' for effective teacher as seen by the teachers.

Table 5.7 Means and standard deviations of the perceptions by job.

			Communication Skills	Organisation & Planning	Resilience & Adaptability	Enthusiasm & Motivation	Professional Ethics
Effective Teacher	Teacher	Mean	9.32	9.26	8.86	9.20	9.38
		SD.	1.73	1.83	1.86	1.82	1.79
	Supervisor	Mean	9.42	9.42	9.10	9.32	9.47
		SD.	1.71	1.58	1.79	1.74	1.71
	Sch. Principal	Mean	9.65	9.63	9.55	9.63	9.69
		SD.	1.02	1.14	1.09	1.14	1.15
Effective Applicant	Teacher	Mean	9.20	9.20	9.10	9.18	9.40
		SD.	1.51	1.63	1.53	1.92	1.53
	Supervisor	Mean	8.85	8.63	8.58	9.12	8.88
		SD.	2.03	2.12	2.10	1.82	2.12
	Sch. Principal	Mean	9.03	9.06	9.08	9.39	9.34
		SD.	1.60	1.51	1.42	1.25	1.46

In order to test the significance of the differences between the three job groups, a Kruskal-Wallis test was used. The results in Table 5.8 show that the three groups did not differ significantly on all domains for working teachers and new applicants, except in the case of the ‘resilience and adaptability’ domain for working teachers (p = .011).

Table 5.8 Kruskal-Wallis Test for Job

		Communication Skills	Organisation & Planning	Resilience and Adaptability	Enthusiasm & Motivation	Professional Ethics
Effective Teacher	Chi-Square	.878	2.09	8.96	4.78	2.82
	df	2	2	2	2	2
	Asymp. Sig.	.645	.351	.011	.092	.244
Effective Applicant	Chi-Square	.609	3.54	3.01	.812	3.60
	df	2	2	2	2	2
	Asymp. Sig.	.738	.170	.222	.666	.165

Similarly, differences between governorates were calculated as shown in Table 5.9. The means of the five domains were nine and above for the responses from Muscat and Batinah South. The means ranged from 8.26 to 9.12 for the Sharqya North governorate, with high disagreement in all domains (*SD* > 2.10).

Table 5.9 Means and standard deviations of the perceptions by Governorate.

			Communication Skills	Organisation & Planning	Resilience and Adaptability	Enthusiasm & Motivation	Professional Ethics
Effective Teacher	Muscat	Mean	9.74	9.69	9.59	9.67	9.71
		SD.	.89	1.14	1.12	1.07	1.09
	Sharqya N.	Mean	9.02	9.00	8.67	8.93	9.12
		SD.	2.20	2.04	2.16	2.21	2.21
	Batinah S.	Mean	9.53	9.53	9.22	9.47	9.62
		SD.	1.31	1.37	1.48	1.41	1.37
Effective Applicant	Muscat	Mean	9.26	9.28	9.21	9.33	9.29
		SD.	1.41	1.37	1.27	1.62	1.53
	Sharqya N.	Mean	8.67	8.26	8.40	8.86	8.81
		SD.	2.36	2.39	2.45	2.19	2.42
	Batinah S.	Mean	9.02	9.09	8.99	9.38	9.35
		SD.	1.54	1.57	1.47	1.30	1.40

A non-parametric test was used to check if the differences between the three governorates were significant. The data in Table 5.10 show that the difference was significant for ‘organisation and planning’ (p = .03) and ‘resilience and adaptability’

(p = .004) for working teachers. The responses by the governorate for the applicants differed significantly only on 'organisation and planning' (p = .025).

Table 5.10 Kruskal-Wallis Test for Governorate

		Communication Skills	Organisation & Planning	Resilience and Adaptability	Enthusiasm & Motivation	Professional Ethics
Effective Teacher	Chi-Square	3.02	7.02	11.17	4.06	2.70
	Df	2	2	2	2	2
	Asymp. Sig.	.221	.030	.004	.131	.259
Effective Applicant	Chi-Square	2.07	7.40	2.34	1.03	.913
	Df	2	2	2	2	2
	Asymp. Sig.	.356	.025	.311	.597	.634

Responses were also analysed according to participants' experience. The means of the responses ranged from 8.33 to 9.77, as shown in Table 5.11.

Table 5.11 Means and standard deviations of the perceptions by experiences

			Communication Skills	Organisation & Planning	Resilience and Adaptability	Enthusiasm & Motivation	Professional Ethics
Effective Teacher	≤5 years	Mean	9.53	9.47	9.18	9.41	9.65
		SD.	1.46	1.51	1.55	1.46	1.46
	6-10 years	Mean	9.06	9.06	8.83	9.06	9.17
		SD.	2.10	2.41	2.23	2.24	2.26
	11-15 years	Mean	9.66	9.53	9.19	9.55	9.62
		SD.	1.01	1.23	1.41	1.25	1.23
	16-20 years	Mean	9.70	9.66	9.49	9.60	9.77
		SD.	1.03	1.04	1.15	1.12	.99
	> 20 years	Mean	9.20	9.30	9.07	9.17	9.26
		SD.	1.96	1.76	1.93	1.97	1.98
Effective Applicant	≤5 years	Mean	8.88	8.82	8.82	9.29	8.94
		SD.	1.80	1.94	1.63	1.49	2.08
	6-10 years	Mean	8.61	8.61	8.33	8.61	8.56
		SD.	2.09	2.17	2.14	2.23	2.20
	11-15 years	Mean	9.40	9.45	9.32	9.47	9.55
		SD.	1.04	1.16	1.05	1.56	1.27
	16-20 years	Mean	9.15	8.92	9.04	9.47	9.28
		SD.	1.69	1.77	1.74	1.48	1.52
	> 20 years	Mean	8.67	8.67	8.65	8.98	9.11
		SD.	2.10	2.01	2.00	1.69	1.99

Table 5.12 shows that there were no significant differences between the experience groups, as shown by a Kruskal-Wallis test.

Table 5.12 Kruskal-Wallis Test for Experience

		Communication Skills	Organisation & Planning	Resilience and Adaptability	Enthusiasm & Motivation	Professional Ethics
Effective Teacher	Chi-Square	2.29	.168	1.89	1.32	1.71
	Df	4	4	4	4	4
	Asymp. Sig.	.682	.997	.756	.858	.788
Effective Applicant	Chi-Square	2.81	5.29	4.12	6.09	4.21
	Df	4	4	4	4	4
	Asymp. Sig.	.589	.259	.389	.193	.378

The 29 attributes were all seen as highly important for working teachers and applicants. As shown in Table 5.13, on a scale of 1-10, the mean was above nine for all attributes for teachers, and ranged from 8.91 to 9.3 for applicants. For each attribute, the difference between the two means was too small, ranging from 0.07 for ‘seeks help when necessary’ to 0.45 for ‘good attitude towards pupils with learning difficulties’.

Table 5.13 Mean and Standard Deviation for the Sub-Domains

	Effective Teacher		Effective Applicant	
	M	*SD*	*M*	*SD*
Humanistic in relation to others.	9.22	1.54	8.97	1.68
Shows a concern and understanding for pupils' needs	9.37	1.36	8.97	1.58
Believes about the pupils' ability to learn	9.40	1.34	9.03	1.57
Good attitude towards pupils with learning difficulties	9.38	1.35	8.93	1.78
Collaborative	9.40	1.35	9.19	1.43
Uses appropriate communication style to suit recipients	9.40	1.35	9.04	1.63
Exhibits active listening	9.34	1.38	9.15	1.59
Good in managing competing priorities	9.13	1.42	8.96	1.65
Displays good time management skills	9.28	1.42	9.03	1.61
Displays good organisation skills	9.26	1.43	9.06	1.59
Good planning skills	9.25	1.45	9.02	1.65
Good classroom management	9.30	1.34	8.93	1.74
Demonstrates the capability to remain resilient under stress	9.23	1.43	8.97	1.57
Comfortable with challenges to own knowledge	9.11	1.41	8.91	1.63
Not disabled by remarks and feedback	9.17	1.43	9.05	1.51
Uses appropriate coping strategies	9.14	1.46	9.06	1.46
Demonstrates high confidence	9.29	1.37	9.15	1.48
Seeks help when necessary	9.13	1.45	9.06	1.58
Commitment to the job roles	9.42	1.32	9.20	1.59
Shows strong and reliable concern to be a teacher	9.47	1.35	9.23	1.48
Aware of national and job loyalty	9.44	1.38	9.30	1.55
Seeks professional development	9.34	1.34	9.12	1.62
Takes pleasure in doing teaching tasks	9.29	1.45	9.13	1.69
Shows high consideration to the Islamic, Omani and professional ethics	9.49	1.30	9.30	1.61
A good model for pupils	9.53	1.44	9.27	1.62
Accepts taking responsibility	9.42	1.30	9.23	1.52
Trustworthy	9.47	1.32	9.24	1.55
Treats others fairly	9.48	1.32	9.18	1.51
Demonstrates respect for pupils and colleagues	9.52	1.29	9.30	1.50

5. 1. 5 Summary of results in Phase one

The aim of this phase was to build the SJT specifications by finding the key non-cognitive attributes that are necessary for prospective teachers in Oman. This section presents the findings of the steps used in this phase. The findings are discussed in line with the literature in the following discussion chapter.

Firstly, data from the official documents and the interview process, compared to the work of Klassen in the UK, contributed to an initial framework of five non-cognitive domains necessary for prospective teachers in Oman. All five domains, and their attributes, were seen as highly important for new teachers and ITEP applicants in Oman. Responses differed significantly according to the gender, except for those

concerning ‘communication skills’. There were no significant differences according to experience. There was significant alignment with participants’ jobs and views on the ‘resilience and adaptability’ domain of new teachers. Responses by governorate differed significantly in the ‘organisation and planning’ domain for both teacher and applicant, and on ‘resilience and adaptability’ for teachers. Moreover, compared to the domains found in the UK by Klassen et al. (2014b), two more domains were seen as necessary for prospective teachers in Oman; namely, ‘professional ethics’ and ‘enthusiasm and motivation’.

The five non-cognitive domains and their definitions comprise the framework for the SJTs in Oman for Phase two:

Communication skills – Candidate is humane in relation to others and demonstrates active listening. Candidate is responsive to pupils’ needs, and able to adapt style of communication to suit recipients.

Organisation and planning – Candidate has the ability to manage competing priorities and display time management skills effectively. Demonstrates good organisation and planning skills.

Resilience and adaptability – Candidate shows the capability to remain resilient under stress and challenges to own knowledge. Demonstrates adaptability and the confidence to make decisions independently, and seeks help when necessary.

Professional ethics – Candidate shows high consideration to the Islamic, Omani, and professional ethics. Demonstrates respect for pupils and colleagues, and treats others fairly. Accepts responsibility and is trustworthy.

Enthusiasm and motivation – Candidate is aware of national and job loyalty, and shows strong and reliable commitment to being a teacher. Takes pleasure in teaching tasks, and seeks professional development.

5. 2 Results of Phase two: The development of SJTs

As illustrated in the methodology, the process of developing the SJTs comprised five main steps: collecting the incidents, building the response format, revising the items, building the answer key, and establishing the scoring key. In this section, the researcher addresses the main results of two of the steps: the revision of the items by the expert teachers ($N = 8$) (step three), and the building of the answer key by groups

of teachers (N = 108) (step four). In addition, the researcher's observations on the development process, in terms of its strength and weakness, are presented in the last point.

5. 2. 1 SJT items and the five domains

In the third step, the items were revised by a group of expert teachers (N = 8). The teachers were asked to distribute the items to the five domains according to the content in each item (the stem and the responses). The distribution of the 53 items, according to the format and the domain they claimed to measure, is presented in Table 5.14.

Table 5.14 SJTs' items by the five domains

Domain	Items' no. (Ranking)	Items' no. (select best three)	Total
Communication (C)	3 – 4 – 5 – 7 – 8 – 10	3 – 4 – 6 – 7 – 9 – 19 – 32 – 37	14
Resilience & Adaptability (R&A)	1	1 – 5 – 13	4
Planning & Organisation (P&O)		2 – 8	2
Enthusiasm & Motivation (E&M)	14 – 15	38	3
Professional Ethics (PE)	13	17 – 20 – 21 – 22 – 31 – 34	7
C / R&A	6 – 9	11 – 15 – 28	5
C / P&O	2		1
C / E&M	12	12	2
C / PE	11	14 – 16 – 18 – 23 – 24 – 25 – 26 – 29 – 33	10
R&A / P&O		36	1
R&A / PE		30	1
E&M / PE		35	1
C / R&A / PE		10 – 27	2
Total	15	38	53

The results show that 30 items were assumed to measure a certain domain, while 23 items could explain two or more domains. Most of the items (n = 14) claimed to

measure the teachers' 'communication skills', 10 items measured a combination of 'communication skills' and 'professional ethics', and seven items measured 'professional ethics'. Only two items measured 'planning and organisation'. The multidimensional nature of the SJTs - meaning that each item can measure multi constructs - is also explored by using the factor analysis procedure in the next phase (Section 5.3.4).

5. 2. 2 Building the answer key

In step four, of the development process, the effectiveness of the 53 items was tested by a group of teachers (N = 108). The participants were asked how a new teacher should respond to each item. The answers for each option were summed up and the percentages calculated per gender. Table 5.15 presents examples of the teachers' responses to some of the items in each format.

Responses for item 13 in the 'ranking' format, for example, show that both males and females expressed strong agreement with the best ranking order. They believed that B was the best option, C second best, A third, E fourth, and option D the least effective option. On the other hand, there was no consensus in the answers to item number 3. In the 'select best three' items, the responses for items number 3 and 5, for instance, were likely agreed that options A, C and E were the best.

Table 5.15 Examples of teachers' responses (%) to some items

Part One: 'Ranking' (1= best option, 5= less appropriate)

Item.		A					B					C					D					E				
		1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
3	F	38	19	38	4		8	4	27	35	27	19	54	23	4			0	12	38	50	35	23	0	19	23
	M	25	46	29	0		14	7	11	29	39	43	14	32	11			14	14	36	36	18	18	14	25	25
	T	31	33	33	2		11	6	19	31	33	31	33	28	7			7	13	37	43	26	20	7	22	24
13	F	0	40	44	12	4	96	4	0			0	50	42	8	0			0	4	96	4	8	17	71	
	M	4	19	37	11	30	93	4	4			4	52	26	11	7			11	26	63	0	26	22	52	
	T	2	29	40	12	17	94	4	2			2	51	33	10	4			6	16	78	2	18	20	61	

Part Two: 'Select best three'

Item		A	B	C	D	E	F
3	F	33	3	16	9	33	5
	M	29	6	14	14	29	8
	T	31	4	15	12	31	7
5	F	29	18	24	4	18	7
	M	30	13	30	6	17	5
	T	29	16	27	5	18	6

Generally, there was not strong agreement between the participants. Hence, the results are as follows:

- Firstly, for the 'ranking' questions, there were 15 items with five responses for each. Only item 13 produced data showing strong agreement (participants agreed on the first response, the second, and so on). Hence, the answer to this item was directly built from the responses. On the other hand, item 3, in this format, had the lowest agreement of the responses. Thus, it was removed and not included in the piloting phase. For the other ranking questions, the participants agreed about the most and least appropriate responses, but there was less consensus on the options in between. For those, a second round of analysis was conducted by the researcher to find the most appropriate answers. Besides the numeric figures, the options were weighted according to factors such as the gaps in choices of males and females, and similarities with other items.

- For the 'select the best three' questions, there were 38 items, with six options for each. It was straightforward to find the answer key for 17 of the items because the data show a good agreement between males and females on the question of the best three options. On the other hand, the agreement was not high for seven of the items. Thus, the researcher sought the better three options, considering other factors related to the nature of the options and the incidents (for example, in some incidents, the option of seeking assistance was considered proper for a new teacher). Finally, and to reduce the number of the items in this format, 14 items were deleted and not included in the next phase. Those items were removed because of a lack of consensus on the answers, a large gap between male and female respondents, and/or high numbers of missing responses.

5. 2. 3 Limitation of the development of SJTs in Oman

For the process of developing the SJT for Oman, the critical incident method was used to collect the items, and the judgment of expert teachers to build the answer key for the developed test. Although this process was driven by evidence from the literature, it had some limitations.

The guidance for collecting the incidents, in the first steps, was to develop each item to measure a certain domain. However, the produced items were multi-dimensional in nature, and each item targeted more than one domain, as seen in Table

5.14. This finding was not limited to this study, but rather assumed to be the nature of the SJT (Lievens, 2006; Patterson et al., 2015b; Whetzel & McDaniel, 2009). As a result, the 35-item SJT was considered a general measurement of the five non-cognitive domains.

In addition, and due to time and cost limitations, the participants in this phase were recruited from only one educational governorate. Further studies should have a wider range of participants from other regions, and include other stakeholders (college tutors, supervisors, and so on). It would also be useful to use other methods, such as observations, to collect incidents. For better development of the answer key, the SJT items might require further analysis using a concordance panel review or a workshop with a group of experts.

The above limitations might have an impact on the reliability of the test. However, it could be argued that this study – which appears to be the first on the development of SJTs in the Omani context – produced significant results. The participants in the development process showed a strong interest and positive attitude towards the nature of the test and its future implications, not just in selection, but also in training and professional development. In addition, the initially collected items (88), in the first step, produced the first bank item of the SJT in Oman. These items could be a starting point for future joint research on SJTs for screening, selecting, or training teachers, not only in Oman, but in any similar Arabic context.

5. 3 Results of Phase three

In Phase three, the 38-item SJT was piloted to a sample of new entrance students in an ITEP in Oman (N = 171). This section presents the results of the pilot study, looking at the missing data, descriptive statistics, reliability, item analysis, and participants' feedback.

5. 3. 1 Missing data treatment

After revising the test papers, 27 were found to have no answers for one or more of the SJT items. For those, the data were counted missing and removed. As a consequence, the data of 144 participants (56.3% female) were analysed.

5. 3. 2 Descriptive statistics

Table 5.16 shows the mean, maximum, minimum, and standard deviations (*SDs*) of the participants' performance in the SJTs. The SJT total scores ranged from 284 to 478 points, with a mean of 415 and *SD* = 36.4. The mean made approximately 73% of the total possible scores (the total possible was 568), which could reflect the difficulty level of the test. The high percentage indicates mean scores close to the total possible score, and, therefore, reflects low difficulty (and vice versa). The standard deviations also indicate the variation in the scores from the mean, with a higher standard deviation meaning a larger spread of values.

Looking at the scores for each type, the mean score for the 'select the best three' was 208 (of 288 total), giving a difficulty of 72%. The 'ranking' items had a mean of 207 of 280, thus 74% difficulty. The standard deviations for both types were almost the same.

A consideration of the scores by gender show that females scored better than males (female: *M* = 433, *SD* = 26.4; male: *M* = 392, *SD* = 35.0). An independent-samples t-test was used to investigate any statistically significant difference between males and females in total scores. The data indicate that scores were significantly higher for females than for males, $t\ (112) = 7.65$, $p < .001$.

Table 5.16 Descriptive statistics for the SJTs' scores by gender and questions' format

Type of Items		Female	Male	Total
Part One (Select best three)	*N*	81	63	144
	Mean	216	198	208
	Minimum	164	124	124
	Maximum	244	236	244
	Std. Deviation	16	21.3	20.5
Part Two (Ranking)	*N*	81	63	144
	Mean	217	194	207
	Minimum	150	157	150
	Maximum	243	233	243
	Std. Deviation	15.5	20.1	20.9
SJTs Score (Total)	*N*	81	63	144
	Mean	433	392	415
	Minimum	314	284	284
	Maximum	478	453	478
	Std. Deviation	26.4	35	36.4

The distribution of SJT total scores, in Figure 5.1, shows that the scores were slightly left-skewed, as the left tail is a little longer. This indicates that most of the participants obtained higher scores on the test. The skewness is much clearer with the female participants, as shown in Figure 5.2.

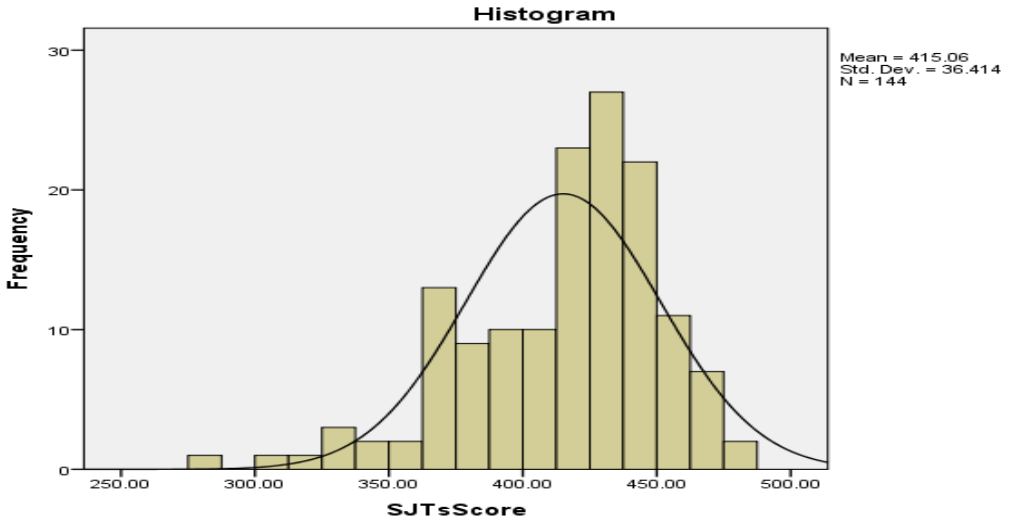


Figure 5.1 The distribution of the SJT total scores.

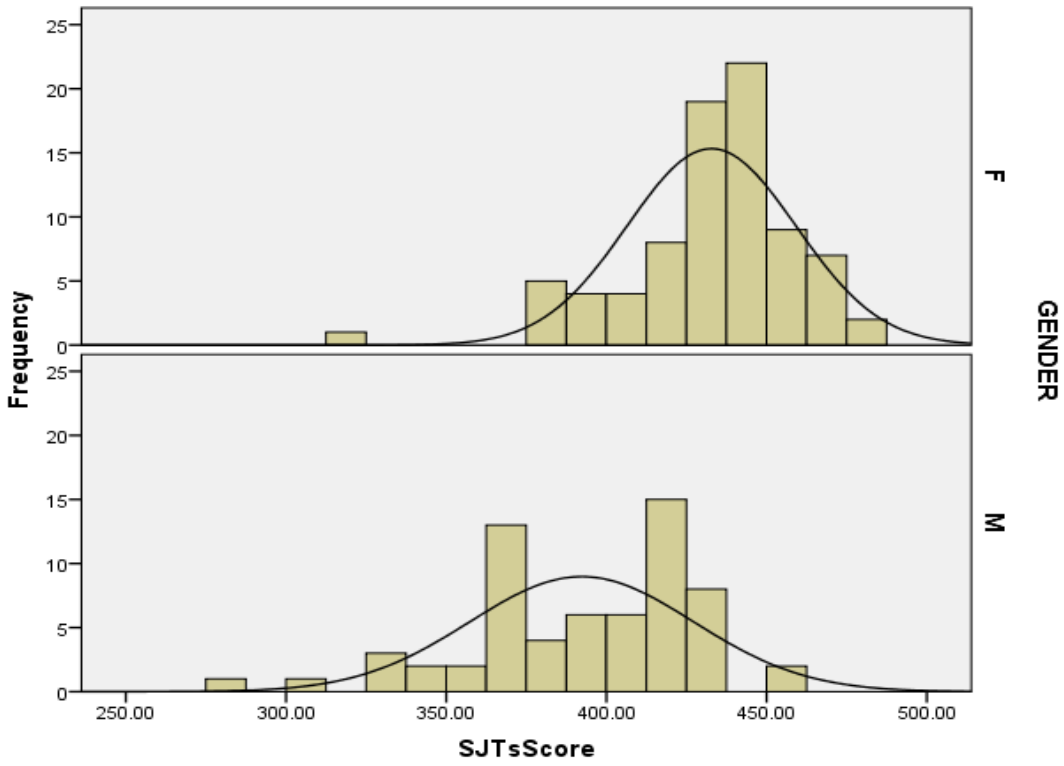


Figure 5.2 The distribution of the SJT scores by gender.

5. 3. 3 Reliability

Most studies of SJTs assess the reliability in terms of internal consistency by using the Cronbach's alpha coefficients. A meta-analysis by McDaniel et al. (2001) finds that Cronbach's alpha coefficients range from $\alpha = 0.43 – 0.94$. Patterson et al. (2015b) state that the internal consistency of the SJT used in a medical and dental context is approximate to – or exceeds – $\alpha = 0.7$. Here, a value of .7 and above is assumed as a standard of good reliability (Cohen, Manion & Morrison, 2013).

The results of the pilot study show that the internal consistency of the pilot SJT was $\alpha = .81$ for the 38 items. For the 'select the best three' (24 items), the internal consistency was $\alpha = .65$, and $\alpha = .78$ for the 'ranking' (14 items). This indicates a good reliability for the piloting test, compared with previous findings in the literature and the assumed standard.

5. 3. 4 Item analysis

Anastasi and Urbin (1997) state that item analysis is a good method for both evaluating and shorting the test. This process can be done qualitatively, by looking to the contents of the items, and quantitatively, by examining the statistical properties of the items. Some of the statistical techniques used in the SJTs are 'item-partial correlation' and 'item's difficulty' (Klassen et al., 2017b; Patterson et al., 2015b). Item-partial correlation, or item-total correlation, is the correlation between the item and the total scores without the item. As a rule of thumb, the item with a partial correlation of .3 and above is considered a good item (Streiner et al., 2015). The item's difficulty, as another statistical method, is the proportion between the actual score and the total possible score of the item. According to Kaplan and Saccuzzo (2009), a range of .3 and .7 in difficulty of the test items can better reflect the differences between the test-takers. However, in some contexts, a higher proportion of difficulty is required.

In the pilot study, the 38 items were analysed using two methods. Firstly, data were analysed quantitively, by looking to the statistical figures including 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.

Table 5.17 shows the analysis of the 'ranking' items. The difficulty of the items ranged from .67 to .8. The difference in difficulty by gender was small, from .0 to .2. The data for partial correlation were .3 and above for all items except two: 0.1 for item 1 and 0.2 for item 10. However, the data show that the deletion of either of those items made no significant improvement to the Cronbach's alpha (it was .78 for the 'ranking' items). Hence, the decision was to include all 14 'ranking' items in the next phase.

Table 5.17 Analysis of the 'ranking' items in the pilot study

Item	Domain*	Difficu -lty	Difficulty by gender ((F/M) – 1)	Corrected Item-Total Correlation	Cronbach's Alpha if Item	Inc. in Phase Four
1	R&A	0.70	0.0	.102	.784	1
2	C / P&O	0.74	0.1	.277	.771	2
3	C	0.78	0.1	.438	.757	3
4	C	0.71	0.1	.381	.762	4
5	C / R&A	0.71	0.1	.472	.753	5
6	C	0.79	0.2	.536	.747	6
7	C	0.71	0.1	.391	.761	7
8	C / R&A	0.70	0.1	.359	.764	8
9	C	0.78	0.2	.537	.746	9
10	C / PE	0.67	0.1	.236	.775	10
11	C / E&M	0.77	0.1	.460	.755	11
12	PE	0.80	0.1	.371	.763	12
13	E&M	0.77	0.1	.367	.763	13
14	E&M	0.70	0.2	.478	.753	14

*Domains: R&A (Resilience & Adaptability), C (Communication skills), P&O (Planning & Organisation), PE (Professional Ethics), E&M (Enthusiasm & Motivation).

For the ‘select the best three’, there were 24 items. The data in Table 5.18 show that the difficulty of the items ranged from .59 to .91. The difficulty by gender showed a difference of .3 in items 6 and 16, and between .0 and .2 for the rest. In addition, the partial correlation had a range between .03 (item 22) and .37 (item 13). There were 11 items with a partial correlation of less than .3. However, the Cronbach’s alpha did not change significantly when an item was deleted. Hence, the decision to omit some of the items was made by looking to the content of the items (see the notes in Table 5.18, for more details). By the end, nine of the items were deleted (1, 5, 9, 10, 16, 17, 18, 22, and 24) and 15 items were reserved for use in the next phase.

Table 5.18 Analysis of the ‘select best three’ items in the pilot study

Item	Domain	Difficulty	Difficulty by gender ((F/M) – 1)	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted	Inc. in Phase Four	Notes
1	P&O	0.69	-0.1	.057	.658	X	Low partial correlation
2	R&A	0.73	0.1	.180	.646	1	
3	C	0.75	0.2	.286	.635	2	
4	P&O	0.60	0.1	.103	.656	3	
5	C / E&M	0.72	0.0	.071	.657	X	Low partial correlation
6	R&A	0.69	0.3	.259	.638	4	
7	C / PE	0.64	0.1	.046	.658	5	Low partial correlation but incident was good for prospective teacher
8	C	0.77	0.1	.251	.639	6	
9	PE	0.73	0.0	.231	.642	X	Seemed inappropriate for females (pupils behaved in unethical situation)
10	PE	0.70	-0.1	.083	.657	X	Low partial correlation
11	PE	0.81	0.1	.275	.637	7	
12	C / PE	0.77	0.1	.307	.634	8	

Table 5.18 *cont.* Analysis of the 'select best three' items in the pilot study

Item	Domain	Difficulty	Difficulty by gender ((F/M) – 1)	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted	Inc. in Phase Four	Notes
13	C / PE	0.87	0.2	.372	.627	9	
14	C / PE	0.78	0.2	.312	.632	10	
15	C / PE / R&A	0.78	0.1	.281	.636	11	
16	C / R&A	0.61	0.3	.335	.628	X	Difficulty and gender gap
17	C / PE	0.63	0.1	.133	.651	X	Difficulty & Low partial correlation
18	C	0.70	0.1	.344	.631	X	options of the incident
19	E&M / PE	0.91	0.1	.265	.640	12	
20	C	0.68	-0.1	.173	.647	13	Good because communication with Special Education Needs' pupils
21	E&M	0.81	0.0	.190	.646	14	Keep because of the domain
22	R&A	0.62	0.0	.030	.659	X	Lowest partial correlation
23	C	0.76	0.2	.305	.633	15	
24	C	0.59	0.1	.303	.634	X	The difficulty

Finally, it is worth saying that a factor analysis method was also explored for the 38 items. This statistical technique helped to summarise the items into smaller factors or components and to reduce the number of items. Pallant (2010) details two approaches: exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). CFA is more complex and sophisticated and used to test specific hypotheses, whilst the EFA is used in the early stages of research to explore the structure of the items.

In general, the factor analysis process has three steps. The first starts by checking the suitability of the data for factor analysis. This step can be done by looking to the sample size and the strength of the relation between the variables. There is little consensus on the best sample size, but generally, the bigger the better: a sample of 150 and above is assumed suitable. On the other hand, the strength of the relationship can be assessed by looking to the correlation matrix for correlations of at least .3 and above. Besides these two criteria, there are two statistical measures in the SPSS to test the suitability of data for factor analysis: the Bartlett's test (should be significant, $p < .05$) and the Kaiser-Meyer-Olkin (KMO) measure (.6 and above is assumed good) (Pallant, 2010).

The second step, once the data have been confirmed as suitable, is to determine the factor extraction technique. This is a statistical treatment to extract the items according to a number of factors. In the SPSS, there are some techniques available, such as principal components, principal axis factoring, image factoring, maximal likelihood factoring, alpha factoring, unweighted least squares, and generalised least squares. The most commonly used is the principal components. After running the extraction technique, a number of factors will come out. To determine the best number of factors, two outputs from the extraction method can be analysed. Firstly, the Kaiser's criterion can be used by looking to the 'total variance explained' table – as a rule of thumb, taking factors which equal 1 or above. The second output looks to scree plot diagram, and takes factors at the point where the curve changes direction. A third statistical method can be used as an advance technique to calculate the number of factors, which is the parallel analysis using a certain formula (Pallant, 2010).

Finally, and for better interpretation of the results, the last step is to rotate the factors. This can help to find a simple structure, with each variable loading strongly

with one component. There are two approaches to rotation. The first is orthogonal, which assumes that the variables are uncorrelated (independent). In SPSS, this comprises three mathematical techniques: Varimax, Quartimax, and Equamax. The second approach is oblique, with two formulas in SPSS – Direct Oblimin and Promax – and assumes that the variables are correlated. The same results are largely produced by both approaches, and might help to conduct both (Pallant, 2010).

The factor analysis for 38-item SJT used in this phase was conducted using the principal component analysis (PCA) with Oblimin rotation, and SPSS version 24. The analysis was conducted separately for the items in each format (the 'select the best three' and the 'ranking'), and also for the total number of items. The results are presented in Appendix 11.

For the 38 total items, the suitability of the factor analysis was firstly assessed using three techniques. The correlation matrix reveals a correlation of .3 between many items. In addition, the Kasier-Meyer-Olkin value is .648 (the recommended value is .6), and the Bartlett's test is significant, at ($p < .001$). Thus, the three results illustrate that the data are suitable for factor analysis. The extraction technique, using the PCA, reveals the presence of 14 components, explaining about 63% of the variance. The scree plot shows a clear break after the first component, which means that one component explains more of the variance than the remaining components. To interpret the component, the Oblimin rotation was conducted. The data in the 'component matrix' table show that most of the items load on the first component. These findings are similar to the analysis of the items in each format. In general, the data show no clear factor structure for the SJT, which supports the multi-dimensional nature of the items, and this is supported by similar findings in the literature (Kasten & Freund, 2015; Schmitt & Chan, 2006).

5. 3. 5 Participants' feedback

Of the participants included in the analysis process ($N = 144$), 92.4% provided a written response to the open-ended question at the end of the test ('Kindly, give your point of view on the test in terms of its suitability for use in the future within the admission procedures of students who wish to enrol on the ITEP, to ensure the selection of the best future teachers. Also add any other comments you have on the test'). The content analysis began with reading the answers to develop a general idea

of how applicants responded. According to the answers, responses were then categorised into themes. Fifteen responses were found to be irrelevant and, hence, removed. The themes were classified as follows:

- The test was positively described by 86.4% of the participants (excellent idea, very good/good, good preparation for entering the training program, enjoyable, highlights my abilities, helps to imagine the future job, helps the college, and rational items and responses).
- Of those who provided positive statements, 46.1% (n = 47) clearly stated that the test was suitable for use in the admission procedures.
- Ten participants claimed that the test was not suitable for admission selection because 'it is not effective', 'little honesty in answers', 'suitable in higher years but not at the beginning', 'should not be compulsory', and 'the variety in the personalities of the applicants'.
- Twelve participants wrote about the difficulty of the test. Of those, nine said the test was moderate, whilst the others said that it was easy.
- Seven participants believed that the test was very long, and the answers might be chosen randomly because of that.
- Other views were included, such as 'some items and responses need to be reviewed and developed' (n = 5), 'need to add open-ended questions in the options' (n = 2), 'better to use single-response questions' (n = 1), and 'better to use visual incidents (video format)' (n = 1).

5. 3. 6 Summary of the pilot study

In general, the analysis of the pilot study indicated that the developed SJT had a good internal consistency. The distribution of the data was close to normal distribution and, hence, able to differentiate between the participants. The items were analysed statistically and qualitatively. In addition, participants gave good feedback on the test and its suitability for the selection process. Some indicated the importance of reviewing the quality of the options and suggested the number of items should be reduced. These results were promising for further exploration of the appropriateness and validity of the SJT for Oman, in the next phase.

5. 4 Results of Phase four

In Phase four, the SJT and three other criterion measures (the Big Five Inventory (BFI), students' academic performance (GPA), and interview scores) were used with a sample of new students in an ITEP in Oman (N = 142). The applicants' reaction to the SJT was measured using a close feedback measure and an open-ended question. In this section, the results from Phase four are presented and analysed. First, the missing data treatment is illustrated. Secondly, descriptive statistics are calculated and presented to understand the distribution of the data. The difference in scores between the participants according to 'gender' and 'year of study' are then analysed. The findings on the reliability and validity of the SJT are analysed individually. Finally, data from the participants' feedback are explored.

5. 4. 1 Missing Data Treatment

In Phase four, the targeted sample were new entrants to the ITEP (students at first or second year, maximum) who were doing their foundation programme and not studying educational courses in teaching or practising in schools. Participants who were not on their first or second year (n = 18) were removed. Thus, data from 124 participants were analysed. The missing data for the 124 participants in each measure are shown on Table 5.19.

Table 5.19 Valid and missing data in Phase Four

	Valid		Missing	
	N	%	*N*	%
SJT Total scores	122	98.4	2	1.6
Neuroticism	105	84.7	19	15.3
Extraversion	111	89.5	13	10.5
Openness	110	88.7	14	11.3
Agreeableness	111	89.5	13	10.5
Conscientiousness	113	91.1	11	8.9
Interview scores	91	73.4	33	26.6
GPA	123	99.2	1	0.8

For the SJT, two participants were counted as missing because they left more than one item unanswered (one left six items, and the other left 14 items). There were

three other participants who left one item unanswered in the SJT; and for those, the missing item was replaced by zero.

For the other criterion measures, the missing data ranged from 11 to 19 for the five factors of the BFI. The participants who missed one or more item in any of the five factors were counted as missing. On the other hand, data for the interview scores and GPA were obtained through the admission unit at the ITEP. An email was sent, including an Excel sheet with the students' college number, for all of the participants. Later, the students' interview scores and GPA were received by email, and the data were analysed. The analysis reveals that 33 scores in the interview and one GPA were missed. The reason, as explained by the sender, is that some students did not attend the interview sessions, and the missing GPA score was for a fresh student in the first semester. The missing data in both measures was counted as missing too.

During the analysis process, the 'exclude cases pairwise' option was chosen, where appropriate, to deal with missing data in all measures. This option allowed the maximising of the number of inputs in the statistical calculations.

5. 4. 2 Descriptive statistics and data distribution

Here, descriptive statistics and data distribution were, firstly, analysed for the participants' scores in the SJTs. Data for the other measures were then analysed in terms of mean, standard deviation (SD), and test of distribution.

Firstly, Table 5.20 shows that the SJTs' total scores ranged from 250 to 404, with a mean of 360 and standard deviation of 26.6. The mean makes approximately 78% of the total possible scores (total possible score was 460), which indicates the difficulty of the test. The difficulty of the 'select the best three' items was approximately 77%, and 79% for the 'ranking'.

Table 5.20 Descriptive statistics of the SJT scores

	Part One 'Select best Three'	Part Tow 'Ranking'	SJTs' total scores
N	122	122	122
Minimum	52	168	250
Maximum	164	244	404
Mean	139.31	220.75	360.06
Std. Deviation	20.49	13.56	26.64
Skewness	-2.500	-1.43	-1.76
Kurtosis	7.57	3.38	3.88
Std. Error Skewness	.219	.219	.219
Std. Error Kurtosis	.435	.435	.435

Checking the normality of the SJT score distribution can help to understand the ability of the SJT to differentiate between the test-takers; and it can also be used as an indicator of the appropriate statistical tests (parametric and non-parametric tests). One can test for normality by looking to (i) the skewness and kurtosis z-values (dividing skewness and kurtosis by their standard errors – the result should be between 1.96 and -1.96 for the normal distribution), (ii) the Shapiro-Wilk test p-value in the normality test (should be above .05), and/or (iii) histogram, normal Q-Q plots and box plots.

The skewness and kurtosis values in Table 5.21 indicate that SJTs scores are slightly negatively skewed and have a slightly higher peak than in a normal distribution. Moreover, the *p*-value in a Shapiro-Wilk's test on Table 5.21 was less than .05, which statistically indicates non-normal distribution. However, the histogram of the SJTs, in Figure 5.3, indicates that the scores were, visually, a close to normal distribution.

Table 5.21 Tests of Normality of SJT

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	df	Sig.
SJT Total scores	.164	122	.000	.846	122	.000

a. Lilliefors Significance Correction

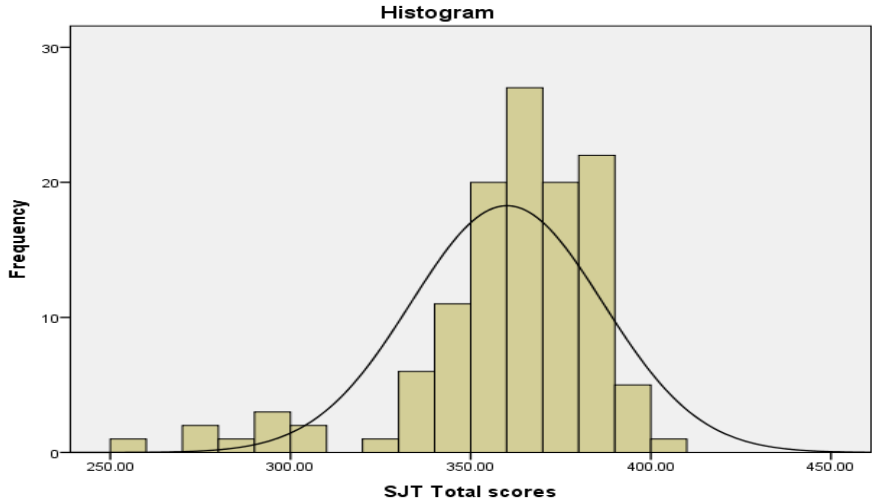


Figure 5.3 Histogram of the SJT scores.

Moving on from the participants' SJT scores, the descriptive statistics for the data in the other measures were gathered, as shown in Table 5.22. The results show that, for the five factors of the BFI, the participants scored higher for 'conscientiousness', with a mean of 47.8 (*SD* = 5.62), whilst 'neuroticism' had the lowest mean (*M* = 30.6, *SD* = 5.39). The mean for the interview scores was 20.7 (of 24 points), and the *SD* was 2.70. The GPA mean was 2.77 (of four points) and *SD* = .62.

Table 5.22 Descriptive statistics of the criterion measures

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis
Neuroticism	105	17	51	30.61	5.39	.629	1.17
Extraversion	111	24	53	42.91	4.80	-.379	1.31
Openness	110	28	51	39.22	4.49	.006	-.114
Agreeableness	111	31	57	45.31	4.99	-.281	.176
Conscientiousness	113	27	57	47.78	5.62	-.780	.994
Interview scores	91	13	24	20.66	2.701	-.574	-.087
GPA	123	1.12	4.00	2.77	.622	-.538	-.081

The distribution of the data in the criterion measures was, statistically, normally distributed for 'openness' ($p = .758$) and 'agreeableness' ($p = .487$), as seen in Table 5.23, though not for the other measures.

Table 5.23 Tests of Normality of criterion measures

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	df	Sig.
Neuroticism	.084	105	.065	.966	105	.009
Extraversion	.088	111	.035	.970	111	.012
Openness	.076	110	.131	.992	110	.758
Agreeableness	.074	111	.172	.989	111	.487
Conscientiousness	.117	113	.001	.958	113	.001
Interview scores	.115	91	.005	.927	91	.000
GPA	.077	123	.067	.971	123	.010

a. Lilliefors Significance Correction

5. 4. 3 Group differences

SJTs total scores in two independent variables (i.e. 'gender' and 'year of study') were tested. Firstly, by looking to the mean and the standard deviations in Table 5.24, females scored better than males in the SJTs and showed less variability (female: *M* = 365, *SD* = 21.9; male: *M* = 339, *SD* = 32.9). In contrast, participants on their first and second year had, almost, the same mean and variance on SJTs' scores.

Table 5.24 SJTs scores by gender and year of study

		N	Mean	Std. Deviation	Std. Error Mean
Sex	Female	97	365	21.9	2.23
	Male	25	339	32.9	6.58
Year of Study	1st year	64	360	26.4	3.29
	2nd year	58	360	27.2	3.57

To test the significance of the difference by gender, an independent sample test was used. The data in Table 5.25 show that the scores were significantly higher for females than for males, t (30) = 3.77, p = .001. The magnitude of the differences in the means in terms of *SD* units (the strength of the relationship) was calculated using an online calculator (http://www.socscistatistics.com/effectsize/Default3.aspxusing) and found to be very high (Cohen's d = .94). The mean difference between males and females was 26.18. The 95% confidence interval for the estimated population mean difference was between 11.98 and 40.38.

Table 5.25 SJTs' Independent Samples Test by Gender

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
SJT Total scores	Equal variances assumed	6.5	.01	4.76	120	.000	26.2	5.50	15.29	37.08
	Equal variances not assumed			3.77	29.7	.001	26.2	6.95	11.98	40.38

5. 4. 4 SJTs' reliability

The reliability of the SJT was tested in terms of its internal consistency using a Cronbach's coefficient alpha. The results in Table 5.26 show that the internal consistency of the SJT was α = .75 for the 29 items. It was .80 for the 'select the best three' items and .55 for the 'ranking' items.

Table 5.26 Reliability of the SJT

	No. of Items	Cronbach's Alpha (N = 122)
Part One (Select best three)	15	.799
Part Two (Ranking)	14	.547
SJT Total	29	.748

5. 4. 5 Validity of the SJTs – correlation between measures

Validity simply means the degree to which an instrument measures what it has been developed to measure (Punch, 2013). This can be explored by correlating results from the developed instrument with other criterion measures. Here, Pearson correlation coefficient was used to test the relationship between the participants' performance in the SJT and the other measures (BFI, interview scores, and GPA). Because some of the measures were not normally distributed, Spearman's rho correlation was also obtained. Results were compared and presented when a difference was noted. For Cohen (1988), correlation coefficient can range from -1 to 1. The value indicates the strength of the relation (a correlation of (0) means no relationship, where (1) is a perfect relationship), whilst the sign shows the direction (positive or negative relationship). As a rule of thumb, the value of .5 and above is counted as large, from .3 to .49 is medium, and from .1 to .29 is small. The correlation between the SJTs and the measures is shown in Table 5.27 and explained next.

Table 5.27 SJTs' correlation with the other measures

		SJT Total scores	Neuroticism	Extraversion	Openness	Agreeableness	Conscientiousness	Interview scores	GPA
SJTs Total scores	Pearson Correlation	1	-.161	.108	.180	.195*	.294**	-.169	.306**
	Sig. (2-tailed)		.103	.264	.063	.043	.002	.113	.001
	N	122	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).

5. 4. 5. 1 ***Correlation with factors in the BFI***

The data in Table 5.27 show a positive significant correlation for 'conscientiousness' (r = 0.29, p = .002, n = 111) and 'agreeableness' (r = 0.20, p = .043, n = 109). It was also positive for 'extraversion' (r = 0.11, p = .264, n = 109) and 'openness' (r = 0.18, p = .063, n = 108), though not statistically significant. SJT had a negative non-significant correlation with 'neuroticism' (r = -0.16, p = .103, n = 104). By squaring the correlation coefficients, approximately 2-9% of the variation in the SJT scores can be explained by the scores in the BFI factors. In other words, more than 91% of the variations were unexplained by reference to the performance in the two tests. Almost the same findings were found using Spearman's rho correlation (see Appendix 12).

The correlation between the SJTs and the BFI for males and females were analysed by splitting the sample by sex. The results in Table 5.28 indicate that the correlation is higher for males, except in the case of 'neuroticism'. The significance of the different correlations between males and females (z value) was also tested using an online calculator (http://vassarstats.net/rdiff.htmal). The calculation of the z value allows an assessment of the likelihood that the difference in sex could be due to chance. The difference is not statistically significant if the obtained z value is between -1.96 and +1.96. The results indicate that the correlation coefficients were not statistically significant (neuroticism z = .19, extraversion z = -1.02, openness z = -0.95, agreeableness z = -0.8, conscientiousness z = -1.03).

Table 5.28 SJT and BFI correlation by sex

		Neuroticism		Extraversion		Openness		Agreeableness		Consciousness	
		Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
JT Total scores	Pearson Correlation	-.209	-.158	.049	.315	.076	.323	.136	.336	.221*	.447*
	Sig. (2-tailed)	.055	.517	.649	.190	.480	.178	.202	.148	.038	.033
	N	85	19	90	19	89	19	89	20	88	23

5. 4. 5. 2 ***Correlation with the Interview scores***

The Pearson correlation in Table 5.27 illustrates a negative non-significant correlation between score in the SJT and the participant's performance in interview (r = -0.17, p = .113, n = 89). The negative correlation was significant when using Spearman's correlation (r = -.26, p = .014). Correlation by sex, in Appendix 12, illustrates that females' scores in the SJT have a negative correlation with their performance in interview (r = -0.24, n = 68), whereas males had a medium positive correlation (r = 0.34, n = 21). However, this difference in correlation was not statistically significant (z = -0.42)

5. 4. 5. 3 ***Correlation with the GPA***

A positive significant correlation was found between the SJTs and the participants' GPA, as seen in Table 5.27, where r = .31, p = .001, n = 121. The finding is the same for females (r = 0.23, n = 97), but a negative correlation was found for males (r = - 0.02, n = 24). The difference in sex was not significant (z = 0.89).

5. 4. 6 Participants' feedback

The participants' reactions to the developed SJT were explored using two measures. First, the participants were asked to express their level of agreement with seven statements on the SJT, assessing its content and use as a tool of measurement. The first three statements were related to the content of the SJTs (relevance, difficulty, and fairness), whilst the last four statements described the potential of the SJT for use in the admission process. There were five options for each statement; where 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, and 5 = strongly agree. In addition, there was an open-ended question that invited any further comments. A total of 140 participants (105 females, 37 males) completed the applicants' feedback about the SJTs, and 67 participants answered the open-ended question. Data from all of the participants, regardless the year of the study, were analysed and presented next.

First, responses to the seven statements were analysed in terms of mean and *SD*, as shown in Table 5.29. The data indicate a strong agreement with all of the statements. The mean ranged from 3.82 to 4.50, in the scale of five options. The strongest agreement was with the first statement: 'Overall, the content of the SJT was

clearly relevant to those applying for the ITEP to become teachers'. The weakest agreement was with statement number six: 'The SJT is an appropriate method that can be used as part of the selection process for candidates applying for the ITEP to become teachers'.

Table 5.29 Mean and standard deviation of Participants' Feedback on the SJT

	N	Mean	Std. Deviation
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
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
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
The SJT will help to differentiate between candidates applying for the ITEPs to be teachers in the future	140	3.99	.944
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	1.008
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
The SJT is able to measure the non-cognitive attributes that are necessary for teachers	140	3.86	.918

A further analysis looked at the percentage of participant responses for each option, as seen in Figure 5.4. In general, there was strong agreement with those statements related to the content of the SJT, which was reduced for the statements on the SJT as an admission tool. More than 91% of the participants agreed or strongly agreed that the content of the SJT was clearly relevant. The percentage saying 'agree' or 'strongly agree' was approximately 82% for the 'difficulty' of the content, and 75% for 'fairness'. In contrast, about 77% and 68% agreed or strongly agreed, respectively, that the SJTs is 'able to differentiate between the candidates' and 'able to measure the non-cognitive attributes'; whilst approximately the same percentage (61%) expressed agreement on the 'fairness' and 'appropriateness' of the SJT as a selection method. Finally, the rate of disagreement (disagree or strongly disagree) was

approximately 5%, on average, for the 'relevance', 'difficulty', and 'fairness' of the content of the SJT, and approximately 7%, on average, for the 'appropriateness of the SJT as a selection method', 'the ability to differentiate between candidates', and 'the ability to measure the non-cognitive attributes'. However, about 12% disagreed, or strongly disagreed, that SJTs were a fair method of selection.

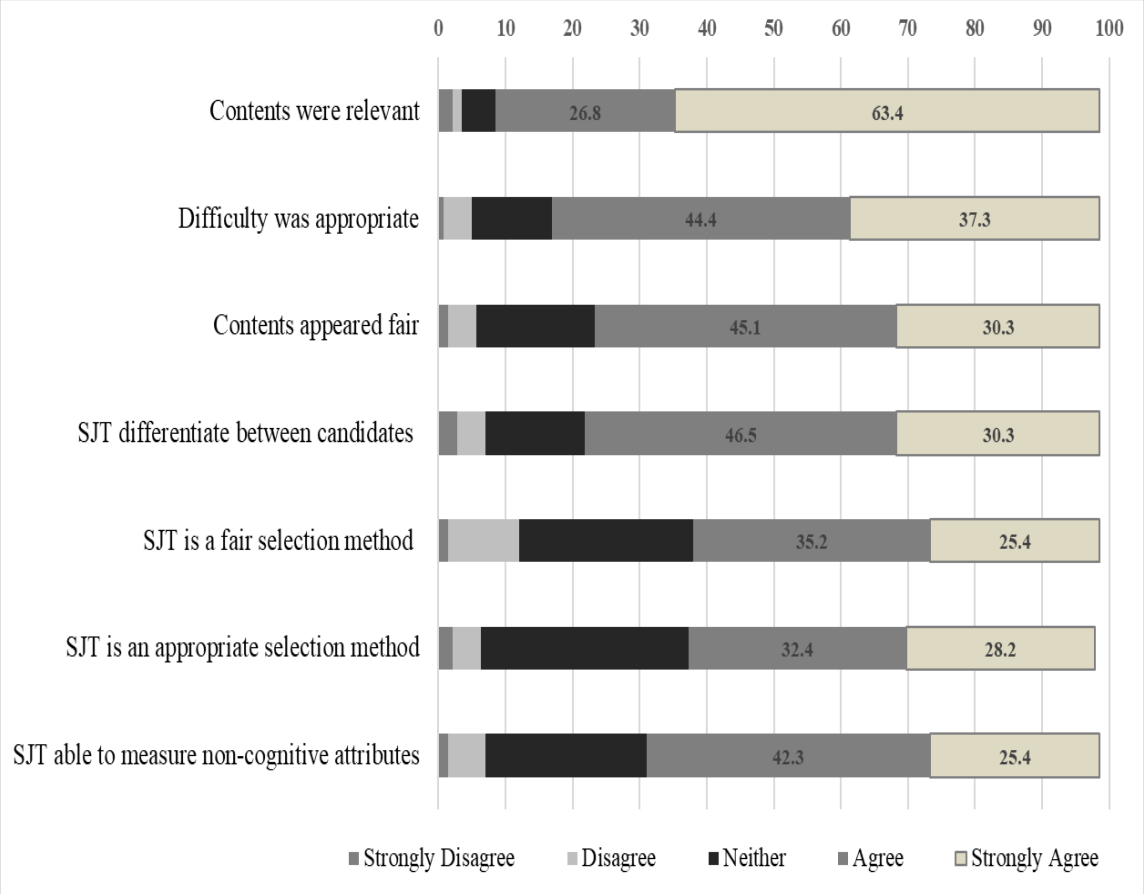


Figure 5.4 Participants' Feedback on the SJT (%).

The difference in the responses given by males and females was also explored, as seen in Table 5.30. The difference between the means of each statement ranged from .03 for statement five, to .33 for statement six.

Table 5.30 Mean and standard deviation of each question by gender

		M	*SD.*
Overall, the content of the SJT was clearly relevant to those	F	4.56	.784
applying for the ITEPs to be teachers in the future	M	4.31	.963
Overall, the level of difficulty of the SJT was appropriate for	F	4.22	.843
those applying for the ITEPs to be teachers in the future	M	3.94	.838
Overall, the content of the SJT appeared to be fair for those	F	3.98	.899
applying for the ITEPs to be teachers in the future	M	4.06	.873
The SJT will help to differentiate between candidates	F	4.05	.913
applying for the ITEPs to be teachers in the future	M	3.80	1.02
The SJT is a fair method that can be used as part of the	F	3.74	1.00
selection process for candidates applying for the ITEPs to be	M	3.71	1.05
teachers in the future			
The SJT is an appropriate method that can be used as part of	F	3.90	.990
the selection process for candidates applying for the ITEPs to	M	3.57	.884
be teachers in the future			
The SJT is able to measure the non-cognitive attributes that	F	3.83	.904
are necessary for teachers	M	3.94	.968

The participants' comments about the SJT on the open-ended question were also analysed ('Please kindly give any other comments you have about this test in terms of either its suitability for future use within the admission procedures for students wishing to join the ITEP, or in comparison to current selection tools [secondary school results, the admission interview, and so on], and/or any other observations you have'). There were 67 responses: 11 from males and 56 from females. The comments were grouped into four categories: the test in general, items and responses, the appropriateness for admission, and other comments.

Most of the comments (about 73%) were positive. Some of the comments state that SJT in general: 'good test', 'excellent', 'very useful', 'clarifies the nature of teaching as a profession', 'makes me more interested in teaching', and 'motivated'. Other positive comments concerned the contents. They were seen as 'suitable', 'realistic', and 'come over them in schools as pupils'. 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, some comments challenged the suitability of the SJTs to be used as a selection method: ‘good, but not for the admission onto the ITEP – better for recruitment’. In addition, some stated that ‘a few items need more clarification’. The fake ability of the test was also mentioned in some comments. The next chapter will discuss these comments.

5. 5 Summary of the main findings

This chapter presents the results from the four phases of the study. Regarding the aim and research questions, two main findings can be summarised. First, five key non-cognitive domains, with 29 attributes, were identified as necessary for the effectiveness of teachers in Oman. It was also seen as crucial to measure the domains and their related attributes in the applicants to the ITEP in Oman during the admission process. Secondly, the results of the pilot and the implementation phases show that the developed SJT has good reliability in terms of internal consistency. In addition, the correlation between the SJTs and the criterion measures shows that the SJT has a positive and significant correlation with two of the five factors in the personality test (the BFI); namely, conscientiousness and agreeableness. The SJT also reveals a positive and significant correlation with the students’ academic performance; but a negative non-significant correlation with the students’ interview scores. Finally, the developed SJT was favourable and accepted by a large percentage of the participants. The next chapter will relate these findings to the literature and the context in Oman.

Chapter 6 Discussion

The findings from the previous chapter will be discussed and interpreted here, according to the two aims of the present study: (a) finding the key non-cognitive attributes necessary for prospective teachers in Oman, and (b) exploring the reliability, validity and applicants' reactions to the use of situational judgment tests (SJTs) in the admission process for initial teacher education programs (ITEPs) in Oman. The discussion of each begins with the presentation of the main findings, and then links these to the literature and the context in Oman, in order to build upon and extend the existing knowledge. To conclude, a summary is given of the key findings and the extent to which the aim of the study has been accomplished.

6. 1 Key non-cognitive attributes of prospective teachers in Oman

For developing the SJTs, this study began by finding the key non-cognitive attributes considered important for prospective teachers in Oman, based on the three domains of attributes found in the UK by Klassen et al. (2014b). An explorative method was used, including a review of official documents, semi-structured interviews (N = 8), and an online questionnaire (N = 181; 58% females). The participants were school principals, teachers' supervisors, and working teachers from three educational governorates in Oman, as well as two academicians working in an ITEP.

As well as the three domains identified in the UK (empathy and communication, resilience and adaptability, and planning and organisation), two new domains were found: professional ethics, and enthusiasm and motivation. The first domain was changed to 'communication skills' from 'empathy and communication'. The results of the questionnaire (N = 181) show that in a 10-point scale, the 'professional ethics' domain got the highest rating for 'effective teacher' (M = 9.53), whilst 'enthusiasm and motivation' was seen as very important for 'an effective applicant' (M = 9.24). Appendix 13 illustrates the definition of each domain.

Here, the researcher discusses the findings of the five domains from four perspectives: (a) how the findings in Oman differ than those in the UK; (b) how the five domains relate to teachers' effectiveness; (c) the use of the five domains in the

selection for the ITEP; and, finally, (d) the development of a theoretical framework for the effectiveness of prospective teachers in Oman. For each, the discussion is supported with evidence from the literature and the context in Oman.

6. 1. 1 Non-cognitive attributes in Oman and the UK: the role of culture

In Chapter 3, culture was defined as the behavioural patterns shared by a group of people which distinguishes it from other groups (Tseng, 2001). The different thoughts about the meaning of work are influenced by the cultural values of a society (Schwartz, 1999), and individuals' non-cognitive attributes are greatly affected by cultural factors (Zhou, 2016). According to Hofsted's model (Hofsted, 2001), countries can be classified in terms of cultural differences, according to five main dimensions: power distance, uncertainty avoidance, individualism/collectivism; masculinity/femininity; and, finally, long-term/short-term orientation. Building on that framework, the researcher considers the similarities and differences between the findings of this study and that in the UK (Klassen et al., 2014b, 2017b). Firstly, the results of the two studies are compared and then seek to understand the findings on the role of culture. Although that is not the aim of this study, it shows the importance of context when exploring non-cognitive attributes for developing the SJTs (or other purposes) in different countries.

Firstly, the UK study used an inductive approach, including observation, interviews, and a focus group. The study produced 13 attributes, classified into three domains, and targeted the selection of primary teachers into the ITEP. The same inductive approach was used in this study, with different methods (analysis of official documents, interviews, and questionnaire). However, the target teachers here were lower and upper secondary teachers (teachers for grades 5 to 12). The results of the two studies share similarities with some theoretical and empirical studies in the related literature. The two studies had similar findings that 'communication skills' ('empathy and communication' in the UK study), 'resilience and adaptability', and 'planning and organisation', were all seen as needing to be tested during the ITEP selection process. In their study, Klassen et al. (2017b) state that the three domains and their related attributes align with other models of teacher effectiveness (Pianta and Hamre's CLASS framework, 2009)- the similarities with the literature are discussed for each domain, in detail, at the next section. However, the findings also indicate some differences that match the needs in the Omani context. There are two

main differences with the UK study. The first is the exchange of the ‘empathy and communication’ domain to ‘communication skills’. The second relates to the findings of the two new domains in Oman: ‘professional ethics’ and ‘enthusiasm and motivation’.

Firstly, the participants in the interview step (N = 8) had some concerns about the Arabic translation of the two words ‘empathy’ and ‘communication’ together. Although they believed that empathy was an important attribute, they were worried that ‘empathy’ with students in certain circumstances could negatively affect ‘communication’. Hence, they preferred ‘communication skills’, as this is widely used in the education system in Oman, whilst ‘empathy’ could be implicitly understood, in the definition of the domain, as an aspect of ‘humanity in relation to others’. This concern might be looked as a simple language matter. However, even in contexts where the English language is used, ‘empathy’ must be cautiously interpreted. Konrath and Grynberg (2013) believe that empathy comes with a few ‘thorns’ and one must be aware of its limitation. In addition, Barr (2011) states that the complex student-teacher relationship needs more than simple empathy. He recommends that teacher-training programmes focus more on training future teachers to practise their empathic capacities. McAllister and Irvine (2002) make a very similar recommendation. Hence, the use of ‘communication skills’ in place of ‘empathy and communication’ in the Omani context is not just a matter of language. Using this terminology in this study could, in fact, reduce any possible ambiguity that could occur when developing the SJT items.

Secondly, the current study identifies two new domains (professional ethics, and enthusiasm and motivation) as important for selecting prospective teachers in Oman, in addition to the three domains found in the UK. The two new domains found in Oman align with the national educational policies and the current practices for selecting and evaluating teachers. For example, the interview for the ITEP candidates includes a statement to measure candidates' commitment to becoming teachers (showing care for academic specialisation) and also their consideration of Islamic and Omani values (demonstrating appreciation of these). Moreover, the assessment criteria used to measure the annual performance of Omani teachers have standards related to commitment to teaching and teachers' discipline, as those values are clearly included in the job description of teachers in Oman.

Despite the new domains matching the current educational practices in Oman, the findings can be attributed to differences in cultural dimensions. In a cross-cultural study, including Finland and Malawi, the findings in Oman and the UK are explained using Hofsted's model. The dimension of power distance was relatively high in Oman compared to England and, as a consequence, produced the attributes in the 'professional ethics' domain. The participants in Oman highlighted the importance of shared community values, whilst the participants in England did not. Similarly, the difference in cultural factors resulted in new non-cognitive domains to match the context in Finland and Malawi ('cooperation and fostering of community' domain in Finland, and 'integrity and community relationships' in Malawi) (Klassen et al., 2018).

Other studies also show the influence of cultural factors on defining the characteristics of teachers' effectiveness. Gao and Liu (2013) explore personality traits of effective teachers, as represented in the narratives of American and Chinese preservice teachers. They find 12 salient personality traits of effective teachers, though the measure of importance differed for each group. The same finding on the role of culture has been revealed among selected Chinese and US teachers (Grant, Stronge & Xu, 2013; Liu & Meng, 2009), Finnish and Swedish teachers (Hemmi & Ryve, 2015), and preservice teachers in Singapore, Hong Kong, Taiwan, and the US (Jiang, 2016).

6. 1. 2 Non-cognitive attributes and teacher's effectiveness

At this point, the researcher discusses the relationship between each domain and teacher effectiveness. For each domain, the definition and the related attributes are presented and then the literature and the context in Oman examined.

Despite the difference in terminology, the 'communication skills' domain has the same definition in Oman as in the UK study: 'Candidate is humanistic in relation to others and demonstrates active listening. Candidate is responsive to pupils' needs and able to adapt the style of communication to suit recipients'. Seven attributes shaped this domain: humanistic in relation to others, shows a concern and understanding for pupils' needs, believes in pupils' ability to learn, good attitude towards pupils with learning difficulties, collaborative, uses an appropriate communication style to suit recipients, and exhibits active listening. Results from the

2013 Teaching and Learning International Survey (TALIS) show that good relationships/communication between teacher and student influence teachers' job satisfaction (OECD, 2014). Likewise, Kington, Reed, and Sammons (2014) found that good relationships with pupils contribute to teacher effectiveness. In addition, emotion, in relationships with pupils, has some influence on social learning and the professional identity of student teachers (Timoštšuk & Ugaste, 2012). Tettegah and Anderson (2007) propose that developing empathy and empathic listening is important for preservice teachers before becoming classroom teachers. Furthermore, teachers' empathy has a significant effect on students' motivation (Waxman, 1983). In Oman, the official job description for teachers has clear items relating to this domain, such as 'taking care of pupils, including pupils with special needs'. Teachers are also evaluated annually on their relation to the school, peers, pupils, and parents.

The second domain is 'resilience and adaptability'. This is defined as, 'Candidate shows the ability to remain resilient under stress and challenges to own knowledge; demonstrates adaptability and the confidence to make decisions independently, and seeks help when necessary'. It has six attributes: demonstrates the capability to remain resilient under stress, comfortable with challenges to own knowledge, not disabled by remarks and feedback, uses appropriate coping strategies, demonstrates high confidence, and seeks help when necessary. The importance of resilience in teacher effectiveness was examined by Gu and Day (2007) in a four-year research project. Their findings show that the interaction between teachers' sense of efficacy, and professional and personal identities contributes strongly to the strength of their resilience. Hong (2012) studied the influence of resilience on teacher attrition rates, exploring the differences between leavers and stayers in the profession. The results indicate that teachers, both leavers and stayers, had intrinsic interest in the profession and shared similar in-job challenges. However, leavers were weaker in self-efficacy beliefs and put greater loads on themselves. Teaching adaptively is also required to respond quickly to the variation among learners (Corno, 2008). 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.

Thirdly, the 'organisation and planning' domain is defined as, 'Candidate has the ability to manage competing priorities and display time management skills

effectively; demonstrates good organisation and planning skills'. This includes skills in managing competing priorities, time management, organisation, planning, and classroom management. In Pianta and Hamre's CLASS framework (2009), classroom organisation is one of the most important dimensions of teacher effectiveness. Sammons et al. (2016) have also found that well-understood routines of classroom management help classes to work smoothly. Omani teachers are asked, in their job description, to prepare and present an annual timeline for their duties, and they are evaluated according to their plans.

The new domain of 'enthusiasm and motivation' is defined as, 'Candidate is aware of national and job loyalty; shows strong and reliable concern to be a teacher; takes pleasure in doing teaching tasks; and seeks professional development'. It comprises the attributes of 'commitment to the job roles, shows strong and reliable concern to be a teacher, aware of national and job loyalty, seeks professional development, and takes pleasure in doing teaching tasks'. In spite of the lack of consensus on a definition, the literature supports the importance of commitment, enthusiasm, and motivation for teacher effectiveness. Hobson et al. (2009) state that prior commitment to the profession is very important. However, Dennis et al. (2015) show that commitment to the workplace is a multi-dimensional concept. With a sample of Hong Kong teachers (N = 857), the study reveals that different types of commitment produce differential effects on employees' job satisfaction, psychological well-being, and desire to stay in their current profession and organisation. Regarding teacher 'enthusiasm', Keller et al. (2014) reveal that this positively predicts students' interest. Finally, teacher effectiveness depends on teachers' ability, and motivation retained even after years of experience (Kyriacou, 2007). However, studies distinguish between different types of motivation (intrinsic, extrinsic, and altruistic) and are guided by a number of theoretical models (goal theory, expectancy-value theory) (Darmody & Smyth, 2016; McGeown et al., 2015). From the work of Watt and Richardson (2007), the model of Factors Influencing Teaching Choice (FIT-Choice) is built. This is used in theoretical and empirical studies to understand teaching candidates' motivation for choosing teaching as a career. A validation study of the scale, conducted in two Australian institutions, shows that factors such as 'intrinsic value, social utility value, and perceived teaching ability emerge as the highest rated influences on the choice of a teaching career; followed by positive prior

teaching and learning experiences and personal utility value' (p.196); whereas choosing teaching as 'a fallback career was rated very low as a motivation for entering the profession' (p.197) (Watt & Richardson, 2007). In addition, research on 'motivation' argues that this attribute is an aspect of an individual's character that is difficult to change (Bieri & Schuler, 2011), remaining remarkably stable over time (Praetorius et al., 2017).

Studies in Oman show that motivation is a significant matter for current working teachers and has a negative impact on students' achievement at college (Al Harthy et al., 2013; Chapman et al., 2012; Issan et al., 2011; Klassen et al., 2011; Zayed et al., 2011). The job description and annual appraisal documents indicate – and there is perhaps a worldwide consensus on this – that a teacher's role is to raise pupils' motivation. Thus, one might ask how this can be achieved if the teacher lacks motivation! Therefore, the emphasis of this new domain in the Omani context reflects the need for enthusiastic and motivated teachers.

The second new domain deemed important for (prospective) teachers in Oman is 'professional ethics'. The study defines this as, 'Candidate shows strong consideration for Islamic, Omani, and professional ethics; demonstrates respect for pupils and colleagues; treats others fairly; accepts responsibility; and is trustworthy. This domain includes the following attributes: shows consideration for Islamic, Omani, and professional ethics; a good model for pupils; accepts responsibility; trustworthy; treats others fairly; and demonstrates respect for pupils and colleagues. In the literature, 'professional ethics' come in different forms, such as integrity and moral fitness. Campbell (2003) highlights the complexity in the relationship between the two terms, 'morals' and 'ethics', despite the assumption of their shared conceptual orientation. Regardless of the difference in terms, he supports the argument that ethics lies at the heart of teacher professionalism: teachers should have the ability to distinguish between right and wrong, especially when dealing with others. In addition, different studies support the importance of 'professional ethics' in teacher effectiveness and when selecting prospective teachers. In Casey and Childs (2007), Sockett (1993) argues that a teacher has moral and ethical obligations to the students and the community, and, therefore, should show good attitudes, morality, and ethics. Lumpkin (2008) defines attributes related to teacher integrity. He believes that teachers must be viewed as moral role models who do the right thing, even when no one is looking,

providing academic programmes of quality and positive educational experiences, and are trustworthy, honest, and respectful. In addition, Jacobowitz (1994) states that teacher preparation programmes must seek out and select candidates who express commitment to their assumed moral and ethical responsibilities. In a study of 440 undergraduate students, Meriac (2015) finds several dimensions of work ethic related to academic motivation and academic performance. This shows that ethics are important in academic settings as well as in work settings.

The terminology 'professional ethics' has been used here as a translation of the Arabic term '*Akhlaqyaat AlMihnah*', which is widely used in the education system in Oman. This is the first statement in the teacher's job description. It is also included in the Candidate's Proficiencies of the College of Education (CoE) at Sultan Qaboos University (SQU). Despite its wide use in the educational context in Oman, 'professional ethics' as yet has no clear definition. However, this domain reflects the shared values of religious and national beliefs (Klassen et al., 2018). In a published speech, Her Excellency the Minister of Education in Oman said that the concept of ethics in the educational profession goes beyond the simple moral dimensions of right and wrong. She assumed that teachers should follow a set of rules and foundations that act as a code of ethics for the profession. She listed a number of ethics, including discipline in work, initiative, self-education, development of individual abilities, enjoyment of the sense of innovation and working in the spirit of the same team, spreading the spirit of cooperation in work, objectivity, and giving priority to the public interest (Resalat Al Tarbyaa, 2011). In addition, Al-Ani, Al-Sulaimani, Al-Aharthi, Al-Munthiri, and Al-Seyabi (2018) explores the definition by conducting semi-structured interviews with 49 school educators. The results show that educators view the ethics of the teaching profession as a collection of good dispositions and values such as sincerity, care, and professional consciousness.

As a theoretical framework, the Arab Bureau of Education for the Gulf States established a deceleration for ethics in the educational professions. It has five areas and 20 items. The first theme looks at teaching as a religious message. The second and the third areas describe the type of relationship between teacher, pupils, and society in general. The fourth theme states that the teacher is self-monitored. Finally, the fifth theme describes the role of teacher in the school-home relationship (Resalat Al Tarbyaa, 2011). In addition, the attributes of the concept of 'professional ethics' in

Oman emphasise Islamic values. According to the Basic Law of the State in Oman, issued in 1996, Islam is the religion of the state and the basis of legalisation. In addition, one of the aims of the philosophy of education in Oman is to develop citizens to follow the principals and values of Islam (OEC, 2017b). Al-Ani and Ismail (2015) find that, from a sample of 161 schools, the theme of 'Islamic values and ethics' is reflected in 55.3% of the schools' mission statements. Although, there is no clear and agreed list of Islamic values, Muslims rely on the Holy Qur'an and Hadith as the source of Islamic Values (Al-Ani, 2014).

As a backdrop to the above, the researcher discusses the importance of the five domains and their related attributes for teacher effectiveness, with evidence from the literature and the Omani context. The results highlight the importance of the domains for the teaching quality and student performance. The next section discusses the importance of the domain for the selection of prospective teachers.

6. 1. 3 Non-cognitive attributes and the selection for the ITEP

The findings of Phase one reveal that the participants (N = 181) also believed in the importance of the five domains for applicants to the ITEP. On the 10-point scale, the mean ranged from 8.92 for 'resilience and adaptability' to 9.24 for 'enthusiasm and motivation'. None of the five domains can be neglected in the screening of applicants. This indicates participants' high perspective of the standards of teaching in Oman. Regardless of their experience, teachers and applicants to the preparation programmes should perform the five domains and their related non-cognitive attributes in their teaching and learning practices.

The selection process for the teaching profession and for the ITEP in other countries share some of the attributes found in this study. In the UK, for example, the Department for Education (DoE) ITEP guidance mandates that attention be paid to non-cognitive attributes, such as ethics and values (Klassen & Kim, 2017a). Communication and motivation are also measured at the Zurich University of Teacher Education in Switzerland (Bieri & Schuler, 2011). In Taiwan, teacher education programmes select according to eight criteria, including character and moral conduct, values and attitudes towards education, and motivations and enthusiasm for teaching (Wang & Fwu, 2007). Furthermore, integrity and career interest are tested when screening trainee teachers (Hashim et al., 2013). Away from education, in a systemic

review of selection in healthcare, the most frequent domains assessed were communication skills, teamwork/collaboration, and ethical/moral judgment (Callwood et al., 2018).

Currently, as illustrated in Chapter 2, the selection process for the ITEP in Oman gives limited attention to the non-cognitive attributes of the candidates. However, for an ITEP, a 10-15-minute interview is conducted with new students during their induction week. The applicant is evaluated against seven items: care for academic specialisation; enjoying working with students with special education needs; attitudes towards, and appreciation of Islamic and Omani values; awareness of the research culture; technological skills; language and communication; and, finally, general professional appearance. Some of these items can be related to at least three of the five domains (enthusiasm and motivation, professional ethics, and communication skills). However, a participant at the interview stage of this study claimed that the outcomes of the interview with new students have little influence on the decision regarding acceptance onto the ITEP.

To conclude, the above discussion about the importance of the five domains and the non-cognitive attributes reveals the limitations of the current practices for selecting prospective teachers in Oman. This asserts the need to enhance the selection method to give the same attention to non-cognitive attributes as given to cognitive abilities. One aspect of the solution is the introduction of new measures, such as the SJT. The next section discusses the findings related to the properties of the SJT developed in this study. However, before this, the researcher will discuss an adapted theoretical framework for the development of prospective teachers in Oman, highlighting the role of the non-cognitive domains found in this study.

6. 1. 4 The development of prospective teachers in Oman: a proposed framework

The researcher will now attempt to develop a framework for the potential contribution of our findings to the development of prospective teachers in Oman. In addition, the proposed framework might help in future research on teacher effectiveness in Oman. Some related literature will be discussed, and then the proposed framework introduced.

In an early review, Doyle (1977) differentiates between three research paradigms in teacher effectiveness. Firstly, the process-product paradigm focuses on

understanding the relations between teacher behaviour and student learning outcomes. The second paradigm is the mediating process which tries to identify the variations in outcomes of student activities during the learning process, and in turn, the instructional conditions. Finally, the classroom ecology paradigm focuses on the relationship between environmental demands and human responses in classroom settings. Over the years, different studies have explored the components that influence teacher quality and, therefore, student outcomes. For example, a model, developed by Hamer et al. (2013), looks at teacher–student interaction as a core driver of student learning. The model proposes that three major domains are important for student learning; namely, emotional support, classroom organisation, and instructional support. Creemers and Kyriakides (2006) suggest a dynamic model that describes more observable factors of teachers' instructions as related to student outcomes. The eight factors included in the model are orientation, structuring, questioning, teaching-modelling, applications, management of time, teacher role in making the classroom a learning environment, and classroom assessment. However, some research of teacher effectiveness has been criticised for its lack of the link to other parts of the education system (Muijs et al., 2014).

Here, the researcher refers to the model of Kunter et al. (2013a), adapted by Klassen and Kim (2017c), for the development of effectiveness in prospective teachers. The model is closely related to the aim of the study, in a number of different ways. It is, firstly, assumed that individuals differ in their capacities gained during the professional development course, whilst also being influenced by certain entry personal characteristics. Thus, the model combines two perspectives on teacher education. The first view is of individual aptitude for teaching, which assumes that success in teaching relies on stable personal characteristics and, thus, determining these characteristics before entry into the teaching is important for teacher success. Secondly, the qualification hypothesis states that teacher education is the important factor in determining success for a teacher. In addition, the model takes into consideration the influence of the context (policy and culture) on the development of teacher effectiveness. Building on that, the researcher develops a model for the context of Oman, as seen in Figure 6.1.

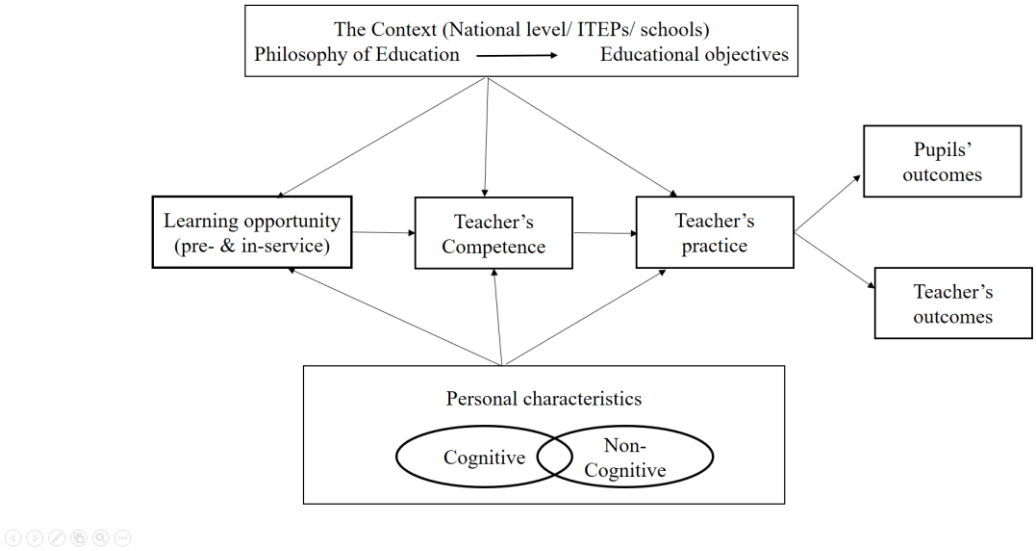


Figure 6.1 Model of the development of teacher effectiveness (developed from Kunter et al., 2013a; Klassen & Kim, 2017c).

According to this model, teachers differ in their success – as measured by pupil and teacher outcomes – because of the variety of teaching practices. This is also influenced by the diversity of the professional competences which are seen as an outcome of the available learning opportunities. This process of development of teacher effectiveness is affected by two main factors: context and personal characteristics. In Oman, the components of the context are identified in the philosophy of education, which takes into consideration the political and cultural factors of the country. However, the different levels of context (national, ITEP, schools) are shaped by the educational objectives of each. On the other hand, the findings of this study put a floor for the components of the non-cognitive dimension of the personal characteristics. Further studies are needed to clarify the components of the other boxes in the developed framework.

6. 2 Exploring the properties of the developed SJTs in Oman

In this section, the researcher discusses the results of implementing the developed SJT in terms of (a) reliability (internal consistency), (b) validity (correlation with the criterion measures), and (c) the applicants' reaction to the content and the proposed use of the test as a selection measure. These three properties are considered part of the principles necessary for psychological testing (Kaplan & Saccuzzo, 2009). In addition, the same approach was used in similar studies for developing the SJT (e.g. Klassen et al. 2014b, 2017b; Patterson et al. 2000, 2008, 2012a, 2013b). The subsections below discuss the main findings.

6. 2. 1 Distribution and subgroup differences of the SJTs

Whetzel, McDaniel and Nguyen (2008) state the importance of testing the potential for mean subgroup differences when using the SJT. A large subgroup difference can indicate a greater likelihood of discrimination between applicants. Thus, before addressing reliability, validity, and applicants' reactions, this point discusses the findings related to the distribution of the SJT scores and the differences between males and females in performance. This allows us to explore the difficulty of the test and its ability to differentiate between the applicants (fairness).

Results at both the piloting phase (N = 144) and the implementation phase (N = 124) indicate that the distribution of SJT scores were slightly negatively skewed, as more participants obtained high scores. The high scores might be explained by the ease of the test for the participants. However, its difficulty was about 73% in the pilot study and 78% in the implementation phase. Another possible explanation is the ability of the participants to perform well in the test. In Oman, the entrance to ITEP is highly competitive, especially for females, thus the entrants are considered the best of the applicants. Therefore, the researcher assumes that the negatively skewed distribution is preferred here, as it is able to identify candidates with poorer performance.

The gender differences showed that the scores were significantly higher for females than for males. At the implementation phase, the magnitude of the difference in the means was very high (Cohen's d = .94), in favour of females. The better performance of the female participants is generally similar to that seen in other studies on the SJT (Lievens et al., 2008; Whetzel et al., 2008). In their review, Patterson et al.

(2012b) indicate that females score slightly higher on the SJT than males, with a mean score of 0.1 *SD* (in Nguyen et al., 2005) and a Cohen's *d* of 0.27 (in O'Connell, Hartman, McDaniel, Grubb & Lawrence, 2007). The Cohen's *d* coefficient was reported in the meta-analysis by Whetzel, McDaniel, and Nguyen (2008), where they found that the female advantage in SJT performance was small ($d = 0.11$).

The magnitude of the differences in performance on the SJT between males and females is higher in Oman than seen in other studies. Generally, female students in Oman outperform males in school achievements (OMoE, 2012) and college courses (Islam & Al-Ghassani, 2015). The results of international studies reveal that Oman had the highest average mathematics achievement by gender for students in grade 8, with girls scoring higher than males (TIMSS, 2015). In addition, the large difference between males and females in the SJT found in this study might be attributable to the context of the school system in Oman, where teachers work in single-sex schools. This gender separation might be an additional factor in the differences in responses to the SJT items. However, such an assumption needs further investigation.

6. 2. 2 Reliability of the developed SJT in Oman

As discussed in Chapter 4, internal consistency was measured in this study to assess the level of reliability. The findings in the pilot study (Phase three) show that the internal consistency for the 38-item SJT was $\alpha = .81$ ($\alpha = .65$ for the 'select the best three' (24 items) and $\alpha = .78$ for the 'ranking' (14 items). The reduction of the items in the implementation study (Phase four) to 29 items reduced the internal consistency of the SJT to $\alpha = .75$ (it was .80 for the 'select the best three' items and .55 for the 'ranking' items).

As a rule of thumb, a reliability of .7 is a minimum for a good test (Anastasi & Urbina, 1997; Kaplan & Saccuzzo, 2009; Kline, 2000). In addition, meta-analysis studies on the reliability of the SJT found means of $\alpha = .46$ and $\alpha = .61$ (Catano et al., 2012; Kasten & Freund, 2015). Therefore, our findings illustrate good reliability for the developed SJT in Oman. The decrease in reliability between the implementation phase and the pilot study aligns also with the findings by Kasten and Freund (2015) on the influence of the number of items on the reliability of the test.

6. 2. 3 Exploring the validity of developed SJTs in Oman

Unlike reliability, validity is not straightforward to measure through one coefficient (Kline, 2000). In its simplest definition, validity is measuring the extent to which the new test measures that which it is supposed to measure (Anastasi & Urbina, 1997). This is done by collecting evidence to justify the conclusion of the test results. Generally, there are three types of evidence: construct-related, criterion-related, and content-related (Kaplan & Saccuzzo, 2009). The correlation between the tests can be cited as evidence of the extent to which they measure the same general area of behaviour. However, a better conclusion of a test's validity should be supported by a combination of logical argument and empirical evidence (Shepard, 1993). Hughes (2017) defines validity with reference to the 'accuracy' of the test (measures what it purports to) and the 'appropriateness' of the test (how useful it is for a given purpose in a given situation).

Relating to the previous definition(s), literature identifies challenges of assessing the validity of SJTs because of its heterogeneous nature, as a one item that can target many constructs or performance dimensions. Moreover, the test-takers' responses might be affected by a combination of their cognitive abilities, personality, and experience (Patterson et al., 2016b). However, in their meta-analysis, McDaniel and Nguyen (2001) believe that the SJT has the largest correlation with the general mental ability and with three personality factors; namely, emotional stability, conscientiousness, and agreeableness.

In the current study, the SJT was developed to measure the non-cognitive attributes of applicants for a better selection of prospective teachers. Hence, the validity of the test can be measured by tracing the participants once in their profession, and correlating their performance with their results on the SJT. This is a long-term goal and will be the aim of future research. However, here, the researcher explores the validity of the SJT by correlating the participants' scores with their performance, using three criteria: personality, using the Big Five Inventory (BFI); cognitive abilities, using the students' latest GPA; and current selection measure, using the interview scores. The findings are as follows.

6. 2. 3. 1 ***Correlation with the Big Five Inventory (BFI)***

The literature shows that the BFI, or 'Five-Factor model', is a widely used measure of personality (Borghans et al., 2008; Heckman & Kautz, 2012) and a valid predictor of job performance (Barrick & Mount, 1991; Rothmann & Coetze, 2003). It has been used in testing the validity of the SJT, and results show that the SJT correlates with three of the factors; namely, conscientiousness, emotional stability, and agreeableness (McDaniel & Nguyen, 2001).

Generally, the results indicate a small correlation between the participants' scores in the SJT and their scores in the Five-Factor Model. Specifically, the SJT correlates positively and significantly with 'conscientiousness' (r = 0.29, p = .002) and 'agreeableness' (r = 0.20, p = .043). The SJT also has a positive correlation with 'extraversion' (r = 0.11, p = .264) and 'openness' (r = 0.18, p = .063), though this is not statistically significant; whilst SJT has a negative non-significant correlation with 'neuroticism' (r = -0.16, p = .103). The difference in correlation between males and females was not statistically significant for the five factors. This result has similarities with the findings of Chan and Schmitt (2002), where SJT shows a significant but weak correlation with Big Five personality traits (r ranged from .19 to .29). In addition, the findings generally align with the meta-analysis of McDaniel et al. (2007). In that meta-analysis, the estimated mean correlations between the SJT and the Big Five are .25 for 'agreeableness', .27 for 'conscientiousness', .22 for 'emotional stability', .14 for 'extraversion', and .13 for 'openness to experience'.

The significant correlation between 'conscientiousness' and 'agreeableness' could be explained by their components and how they relate to the five domains shaping the SJT. 'Conscientiousness', in the BFI, contains items related to self-control, active processes of planning, organising, a hardworking nature, taking responsibility, and applying moral principles. Thus, it shares some attributes with at least three of the five domains: 'enthusiasm and motivation', 'planning and organisation', and 'professional ethics'. 'Agreeableness', on the other hand, includes sympathy with others, being eager to help, and being unselfish. Thus, it shares attributes with 'communication skills' and 'professional ethics'. However, this assumption of similarity between the attributes measured by the SJTs and the factors from the BFI needs further research. In addition, further research should investigate the differences associated with type of situation in the SJT. Kell et al. (2010) note that

emotionally stable and conscientious actions are more effective in task situations, whilst open and agreeable actions are more effective in interpersonal situations.

6. 2. 3. 2 *Correlation with interview*

Although both the developed SJT and the interview are assumed to be testing the non-cognitive attributes of the participants, the results illustrate a negative non-significant correlation between the scores (r = -0.17, p = .113, n = 89). The negative correlation is significant when using Spearman's correlation (r = -.26, p = .014). Correlation by sex reveals a negative correlation for females (r = -0.24, n = 68) and a medium positive correlation for males (r = 0.34, n = 21). However, this difference in correlation was not statistically significant. It is difficult to explain the gender difference, due to the large difference in the sample sizes for males (N = 21) and females (N = 68).

The finding of a negative correlation contradicted the findings of Klassen et al. (2017b) in the UK. The SJT in their study has a correlation of .29 with the overall interview score. A correlation of r = 0.52 between applicant scores on the SJT and in a multiple-mini interview in selection for postgraduate training in medical schools was also found (Patterson et al., 2012b). However, the correlation between the two measures could be affected by the purpose of each. In Patterson et al. (2012a), the SJT scores had a positive correlation with the management, leadership, and professionalism interview, but not with the clinical skills interview. Although the interview process used in the ITEP in Oman targets the personality of the applicant, validity has not – to our knowledge – been measured. In addition, the feasibility of the interview step in the admission process is not clear. While the document from the ITEP states that the interview is part of the selection process, acceptance in higher education programs in Oman is generally made from the outside, by the national admission centre (HEAC).

6. 2. 3. 3 *Correlation between the SJT and cognitive ability (GPA)*

Understanding of the relationship between non-cognitive attributes and academic performance has increased over recent decades (McAbee & Oswald, 2013). The literature shows that personality traits are significant predictors of academic achievement in university (Chamorro-Premuzic & Furnham, 2004). Schwager, Hülsheger, Bridgeman and Lang (2015) find that non-ability-related factors – such as

conscientiousness, motivation, or adaptability – can play a larger role in determining whether or not students complete their studies within the designated timeframe. In Oman, Al-Harthy and Aldhafri (2014) reveal that the variables of task-value and self-efficacy correlate significantly with students' GPAs. In addition, a study of 1,511 Omani students investigated predictors of first-year university GPA. The results show that general education diploma (GED) score, gender, overall performance on the foundation placement tests, type of college, extrinsic goal orientation, university readiness, and critical thinking are collectively statistically significant predictors of GPA (Alkhausi et al., 2015).

Based on previous evidence of the importance of the non-cognitive attributes of students' academic achievement, the correlation between the developed SJT and the students' academic achievement, in terms of their GPA, was tested. The results show a medium positive significant correlation between the SJT and the participants' GPA (r = .31, p = .001, n = 121). This aligns with previous studies on the relation between the non-cognitive attributes, measured by the SJT, and the students' academic achievement, measured by their GPA. Lievens (2013) proposes a video-based SJT to measure interpersonal behaviour, which has significant added value over cognitive tests for predicting interpersonal GPA and doctor performance. Furthermore, in some related systemic review and meta-analysis studies, the SJT has a correlation with cognitive abilities (Patterson et al., 2012b). This correlation is equal at 0.46 in the McDaniel et al. (2001) meta-analysis study. However, with a sample of employees, Chan and Schmitt (2002) find that the SJT was uncorrelated with cognitive ability (r = –.02). This variability in findings was examined by McDaniel et al. (2001). The results indicate that the SJT based on a job analysis, and that with more detailed questions, were more closely correlated with cognitive ability.

6. 2. 4 Applicants' reactions

To explore the validity of a psychometric measure, it is required to consider (a) the relevance of the content to the targeted construct(s), and (b) the appropriateness of the measure for a given purpose (Hughes, 2017). The evidence was collected using the participants' feedback. The feedback had seven items: the first three statements were related to the content of the SJT (relevance, difficulty, and fairness), and the last four statements described the potential of the SJT for use in the admission process. In

addition to the seven-item feedback, the participants were asked to share any further responses to the test in an open-ended question.

The applicants saw the content of the SJT as relevant, of appropriate difficulty, and fair, by margins of 91%, 82%, and 75%, respectively. Specifically, the majority agreed or strongly agreed that the content of the SJT was relevant and fair for those applying for the ITEP, and its level of difficulty was appropriate. In the open-ended question, the majority of the comments were positive about the content of the SJT. For example, participants wrote that it was, in general, 'a good test', 'excellent', 'very useful', and that it 'clarified the nature of teaching as a profession', and 'made me more interested in teaching', and 'motivated'. Other positive comments said that it was 'suitable' and 'realistic'. However, some of the comments indicated that 'a few items need better clarification'.

Secondly, responses to the four statements on the potential of the SJT for the admission process were also good, but with less agreement than in the previous section. Approximately 77% and 68% agreed or strongly agreed, respectively, that the SJT is 'able to differentiate between the candidates' and 'able to measure the non-cognitive attributes'. However, only 61% expressed agreement with the two statements related to the 'fairness' and 'appropriateness' of the SJT as a selection method. However, in the open-ended question, some comments stated the appropriateness of the SJT for selecting candidates for the ITEP, saying it 'must be implemented', and it is 'better than the interview', 'better than just seeing results in secondary school', and should be 'included alongside the current tools'. Some comments challenged the suitability of the SJT as a selection method. For instance, one comment said, 'Good, but not for the admission onto the ITEP – better for recruitment'.

In addition, some comments concerned the fakeability of the test. Several studies concern the issue of dishonesty in non-cognitive measures. Results indicate two potential moderators of faking behaviour: the ability to fake and the opportunity to do so (Douglas, McDaniel & Snell, 1996; McFarlan & Ryan, 2000). Peeters and Lievens (2005) state that faking has a negative effect on criterion-related validity and the incremental validity. However, the degree of faking can vary according to the response format. Nguyen, Biderman, and McDaniel (2005) argue that the knowledge

response format (what you should do? what is the correct thing to do? how effective is the behaviour?) is more resistant to faking.

In general, the applicants' reactions to the developed SJT, as content and as a selection tool, were positive. This aligns with the study conducted by Klassen et al. (2014b). The results in the UK show that most applicants (76.7%) found the content and format of the pilot selection tool to be favourable. The SJT was seen as clearly relevant by 79% of participants, and the level of difficulty appropriate by 74% (Klassen et al., 2017b). A project developing an SJT for teacher selection in Australia reveals also the same positive reactions. Agreement with the relevance of the content was 91%, fairness was 94%, and appropriate level of difficulty was 98.5%; though most participants were neutral as to whether the tool would be fair and appropriate as a selection method (Durksen & Klassen, 2018). In the medical admission process, the SJT was also received positively by candidates (Patterson et al., 2012b).

6. 3 Summary

In this chapter, the findings of the data analysis in relation to the context in Oman and the related literature are discussed. The discussion can be summarised as follows.

- Five non-cognitive domains were found to be essential for the selection of applicants into the ITEP. These domains reflect the cultural factors in the Omani context and align with the requirements of (a) the official documents for teacher preparation and expected performance, and (b) the current practices in Omani schools.
- The distribution of SJT scores was close to the normal distribution, and therefore allows differentiation between the candidates. However, the difference in scores by gender was high, compared to previous studies. This might be attributable to the nature of the single-sex schools in Oman. One implication of this finding is that the gender issue should be taken into consideration when using the SJT in the selection process – by, for example, using different forms of the test or a different scoring key for each gender.
- The developed SJT has a good internal consistency, using the alpha Cronbach coefficient. However, research on the reliability of the SJT suggests the use of other tests, such as the test-retest reliability. This was not possible in the current study, and is therefore a recommendation for future research.

- The overall correlation between the BFI and the GPA suggests that it is worthwhile to assess the predictive validity of the SJT in Oman in future research. However, one concern regarding the correlation with students' academic performance is the reliability of the current students' GPA, as one study found that the distribution of grades, both at the departmental and college levels, was higher than expected (Hassan et al., 2009).
- The negative correlation with the interview scores suggests that the SJT and the interview measure different constructs. However, there are some limitations of the interview process for the ITEP in terms of reliability and/or feasibility. Thus, further investigation is required.
- Consistent with previous studies, the developed SJT was accepted favourably by a large proportion of the participants.

To conclude, the initial findings on the psychometric properties of the developed SJT in Oman are encouraging for further research, especially in terms of the validity of the test. Although research argues that SJT can be a useful and valid complement to traditional student admission tests, even in an operational high-stakes context (Lievens et al., 2005), such a conclusion needs further investigation in Oman. In the next chapter, the conclusion of the study will be presented, alongside recommendations for future research and explanations of related policy implications.

Chapter 7 Conclusion

This final chapter concludes the study and comprises five sections. The next section gives a summary of the study and the main findings. The limitations of the study are presented in the second section. Section three highlights the contribution of the study to the existing knowledge and the context in Oman. In sections four and five, respectively, the recommendations for the policymakers and suggestions for further research are presented.

7. 1 Summary of the study

Improving the effectiveness of teachers is an important step towards improving the quality of the education system as a whole, and therefore the quality of outcomes. In addition to benefitting from improved qualifications, better training, fair promotions, and efficient evaluations, teacher effectiveness could be increased by changes to the selection process. Studies show that the ability to predict prospective teachers' performance is relevant to students' scores and teachers' attrition rates. In their study, Boyd, Grossman, Lankford, Loeb and Wyckoff (2008) note that less effective first-year teachers have higher attrition rates than do more effective teachers. Furthermore, Atteberry et al. (2013) show that, on average, initial performance is predictive of future performance: the top fifth of teachers remain the top fifth of teachers, and so on.

Despite its importance, the selection process, in practice, has two main challenges. First, the educational system focuses on selection at the recruitment stage, before the job, giving less attention to selection in the early stages – that is, at the entrance to the initial teacher education programmes (ITEPs). Secondly, most selection practices focus largely on the cognitive aspects, either looking to the candidates' performance on previous examinations or setting admission tests (Casey & Childs, 2007; Ingvarson et al., 2013). The non-cognitive attributes of the applicants are given less attention during the admission process, and mostly not considered in connection with the acceptance decisions. However, the researcher refers in the discussion chapter to the Kunter et al. (2013a) model in order to understand the role of non-cognitive attributes in the professional development of prospective teachers. The

model explains the influence of personal characteristics (cognitive, non-cognitive, and background) into the difference in effectiveness among teachers.

In some admission practices, the secondary view of the role of non-cognitive attributes into the ITEP can be referred to some controversial issues, such as whether teachers are 'born or made', stability/malleability of non-cognitive attributes, and the ability of the selection measures to predict future performance. Although these aspects can be applied to both, they have been identified as more pertinent to non-measures than to cognitive. However, evidence for the importance of non-cognitive attributes and the limited time available to prepare teachers for the ITEP (Dolan, 2012; Fantilli & McDougall, 2009; Jacobowitz, 1994) indicate an urgent need to understand these attributes during the admission process. In addition, the limitations of the current selection measures for non-cognitive attributes (Klassen & Kim, 2017c) motivate the ITEP to seek a better tool. Evidence-based research on the use of situational judgment tests (SJTs) to select applicants for medical schools shows that they have good levels of reliability, predictive validity, and incremental validity for testing professional attributes (Patterson et al., 2012b).

Although the education system in Oman has undergone considerable development, teacher quality remains a concern (OMoE, 2012). Despite the current implications and the proposed initiatives to improve the effectiveness of the schools' teachers, the selection process for ITEP applicants remains the same, with its focus on students' performance in secondary school (grade 12). These challenges around teacher quality in Oman make it worthwhile to explore the use of the SJT during the admission process for better selection of prospective teachers. Therefore, this explorative study aims to develop the SJT for selecting ITEP applicants in Oman. The properties of the developed tool (reliability and validity) were tested by measuring the internal consistency and correlation of the SJT scores and other measures. Furthermore, the applicants' reactions to the developed SJT were explored using an applicants' feedback measure ending with an open-ended question. The study was built on previous work conducted in the UK on developing the SJT for teacher selection (Klassen et al., 2014b, 2017b).

The results show that five key non-cognitive attributes should be measured during the ITEP admission process in Oman. As well as the three domains found in the UK (communication skills, resilience and adaptability, and organisation and

planning), two more domains were deemed important for the Omani context; namely: 'professional ethics' and 'enthusiasm and motivation'. The findings regarding these new domains align with the current educational policies in Oman, as seen in the relevant official documents (the philosophy of education, the teachers' job description, and so on). In addition, the difference between the Omani and UK contexts indicates the role of culture in defining the effectiveness of teachers (Klassen et al., 2018). Finally, the Kunter et al. (2013a) model is modified for the development of teacher effectiveness to suit the context in Oman as seen in Figure 7.1. The adapted model could support future research in this field.

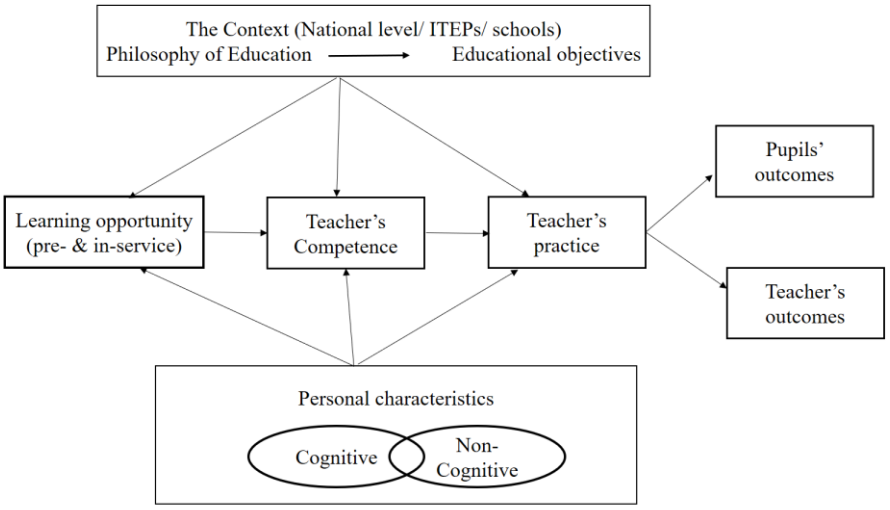


Figure 7.1 Model of the development of teacher effectiveness (developed from Kunter et al., 2013a; Klassen & Kim, 2017c).

The five resulting domains were used to develop the SJT for exploring the non-cognitive attributes of new entrants to the ITEP in Oman. To our knowledge, this is the first SJT to be developed in the Omani context. Data from the participants show that the developed SJT has a good internal consistency, using the Cronbach's alpha coefficient. Females performed significantly better than males with high effect size. The correlation between the SJT scores and other measures indicates that the SJT correlates significantly with two facets of the Big Five Inventory (BFI) personality measure; that is 'conscientiousness' and 'agreeableness'. In addition, the results show a medium positive significant correlation between the SJT and participants' GPA. Conversely, the SJT has a negative non-significant correlation with the scores of the interview currently used in the admission process. Furthermore, the participants' responses to the SJT as content and as a selection tool were positive. Therefore, the results were generally promising for further studies on the validity of the SJT, especially in high-stakes contexts.

7. 2 Limitations of the study

Despite the promising findings, the study inevitably has limitations. These limitations can be divided into two groups. The first is common to similar studies of the SJT, whilst the second is specific to the context of this research.

First, the multidimensional nature of the SJT items, with each item in more than one domain, poses a challenge for most studies in this field (Durksen & Klassen, 2017; Lievens, 2006; Patterson et al., 2015b; Whetzel & McDaniel, 2009). This challenge made it difficult to determine the correct responses to the situations, and thus affected the building of the answering and scoring keys. In addition, this multidimensional nature has an impact on the reliability of the SJT, measured by the Cronbach's alpha.

The second group of limitations is the participants and methods used in this study. For the development of the SJT, the participants were recruited using a non-random sampling procedure. They came from the same educational governorate, where the researcher works, and were all experienced teachers. Oman has 11 educational governorates, covering a variety of geographical and cultural diversity. Hence, further studies should accommodate this variation by introducing a wider range of participants, from other regions, and including other stakeholders in building

the SJT (college tutors, supervisors, students on the ITEP). In addition, and because it was not possible to conduct the implementation study during the selection stage, the participants in the implementation phase were not all fresh entrants. Some had already completed two or more semesters into the ITEP. This might have an effect on the correlation between the SJT and the other criterion measures, especially interview scores.

Finally, in Section 5.2.3 the researcher illustrated some of the limitations in the development process of SJTs in Oman. The development of the SJT could be strengthened by including other methods. Classroom observation, for instance, could be used to collect incidents. In addition, the data used for the building of the answer key might require further analysis by a concordance panel review or a workshop with a group of experts. However, this was not possible in this study due to the conditions of the participants and the researcher. Measuring the validity of the SJT requires outcome measures of teacher effectiveness in the profession. This is a long-term goal for future longitudinal studies.

7. 3 Contribution to knowledge

This study contributes to the existing knowledge on both the non-cognitive attributes of effective teachers and the use of SJTs in the ITEP selection process. Research on the non-cognitive attributes necessary for teacher effectiveness in Oman is scarce. Thus, the findings from Phase one offer a better understanding of the necessary attributes for Omani teachers. In addition to selection criteria, the five domains and their related attributes (see Appendix 13) contribute to education policy in Oman around teacher training, evaluation, and promotion. The five domains found in Oman will allow cross-cultural comparison studies of the critical attributes required for novice teachers. As a starting point, our findings have been compared to the findings of the critical non-cognitive attributes of novice teachers in three other countries: England, Finland, and Malawi (Klassen et al., 2018).

The current study contributes to the literature in the SJT in many ways. Firstly, studies in SJTs recommend the need to conduct further research in different cultures (Lievens et al., 2015). To our knowledge, this is the first study on developing the SJT for the selection of ITEP applicants in Arab countries in general, and Oman specifically. The Arabic version of the SJT, built on the basis of this study, makes a

significant contribution to the context in the Arab world. It could be used as a starting point for further collaborative studies.

Finally, the study contributes to the educational context in Oman. It illuminates the limitations of the current selection practices used for ITEP admissions in Oman. It reveals the importance of measuring the non-cognitive attributes of applicants in order to ensure high quality of new entrants. As a practical solution to the challenge of assessing the non-cognitive attributes, the SJT developed in this study could be considered an additional, promising selection measurement, benefiting from further studies and evaluation.

7. 4 Recommendations for policymakers in Oman

Many countries have policies in place designed to attract the candidates most likely to become highly qualified teachers (Tatto, 2008). The findings of this study therefore have implications for policymakers working with ITEP and those concerned with selecting prospective teachers in Oman.

The first recommendation is to establish a policy framework for the non-cognitive attributes necessary for teacher effectiveness in Oman. This policy should be developed by the stakeholders in both the MoE and the ITEPs. The policy will be used in the selection, training, and evaluation practices at the different educational levels. The framework of the five non-cognitive domains built on the basis of this study, as shown in Appendix 13, could be used as a starting point for further discussion.

The five domains share common attributes with Oman's philosophy of education. According to this philosophy, the aim of education is 'to develop citizens who demonstrate faith in Allah, follow the principles and values of Islam, practice loyalty to the country, understand current events and respond to these events in an appropriate manner, acquire scientific thinking skills and contribute to achieving sustainable development across all sectors of Omani society' (OEC, 2017b, p.11). Teachers are considered the cornerstone of the process of developing those skills and attributes in their students. The initial framework of the five domains could be discussed and modified by policymakers in the two main bodies of the MoE and the ITEP. Both bodies should make clear statements about which attributes should be (a) developed in schools, (b) tested during the admission process into the ITEP, (c)

developed during the training programme, (d) tested at the recruitment process, and (e) developed during the profession through training programmes. In addition, this policy framework should consider diversity among teachers, including gender and level of teaching.

The second recommendation is to place greater emphasis on non-cognitive attributes when selecting prospective teachers. The admission system in Finland, for example, is designed to afford an equal starting point to all students with a desire to be teachers, rather than accepting only the top upper-secondary school graduates. Selection for teacher education programmes in Finland focuses on finding those individuals with the right personality, advanced interpersonal skills, and right moral purpose of becoming lifelong educators (Sahlberg, 2015). Thus, policymakers at the Education Council in Oman should work collaboratively to design a comprehensive selection model that balances the needs for cognitive and non-cognitive attributes.

Admission to higher education in Oman is currently determined at the national level, through the Higher Education Admission Center (HEAC). Generally, students are screened and offered places according to their academic performance in the last grade of secondary school (grade 12). In some subjects, the successful candidates are set an admission test or an interview before being offered a final acceptance decision. For teacher training courses, the candidates are mainly accepted based on their school results, with the exception of art and physical education subjects, where they are set an aptitude test. Here, the researcher suggests that applicants to the ITEP, regardless of their subject, should be screened using a multi-stage model. In Chapter 3 (sub-section 3.3.5), two models proposed by Bowles et al. (2014) and Klassen and Kim (2017a) were presented. Both suggest a multi-stage model that allows screening of the applicants according to their background experience and their cognitive and non-cognitive attributes. The models are included in Appendix 3 and can be discussed further by the policymakers and adapted to the Omani context.

Thirdly, the current study introduces for policymakers in Oman a promising selection tool with positive reactions from applicants. However, the implementation of the SJT in a high-stakes selection process requires further investigation, especially in terms of the validity and feasibility of such a tool. A suggestion for further studies on the use of the SJT in Oman is presented in the next section.

Lastly, any proposed policy for enhancing the selection process for new entrants on the ITEP in Oman should be part of a comprehensive strategic plan to improve the quality and status of teaching as a profession. Related to this, the Education for All Global Monitoring Report 2013/2014 suggests four strategies to enhance the quality of teachers: improving teacher education, allocating teachers more fairly, and providing incentives in the form of appropriate salaries and attractive career paths (UNESCO, 2013). Thus, any policy proposed to improve the selection process should be integrated with such strategies to ensure better educational outcomes.

7. 5 Suggestions for future studies

This study aims to explore the use of the SJT to measure the non-cognitive attributes of new applicants to the ITEP in Oman. As noted earlier, this is the first study in Oman, to our knowledge, on developing the SJT to select prospective teachers. Therefore, further research is required. Here, the researcher suggests three potential groups of studies.

The first is related to non-cognitive attributes and teacher effectiveness in Oman. This group of studies would provide empirical evidence on the role of teachers' personal characteristics (cognitive attributes, non-cognitive attributes, background factors) in quality of teaching and student achievement in Oman. Secondly, the researcher suggests studying the efficiency of the current selection procedures for the teaching profession and for entrance to the ITEP in Oman. These studies could help to evaluate the current practices and clarify their strengths and weaknesses, thus contributing to improvement strategies. The last suggested group of studies concerns the validity and feasibility of implementing the SJT for teacher selection in Oman.

Specifically, the following studies are suggested:

- This study identifies the limited research on critical non-cognitive attributes related to teacher effectiveness in Oman. The initial findings of the current study could be further developed using different methods (classroom observation, for example) or samples from other educational regions.

- To understand the influence of teachers' non-cognitive attributes, it is necessary to further study their relationship with teacher effectiveness and pupil attainment at the school level. For example, a study could explore the role of teachers' professional ethics in student achievement in Oman.
- A longitudinal study could examine the validity of the current selection procedure for the ITEP in Oman. It could measure the relationship between performance at admission (interview, secondary school results, and so on) and performance in the profession (annual appraisal, peer/pupil reviews, and so on)
- Lastly, Lievens (2006) notes that the SJT being efficient in some cultures does not guarantee that it would be similarly so in other cultures. Hence, the researcher suggests a project comprising a number of studies of the feasibility and validity of the SJT in Oman. The project could develop different types of SJT (video-based, single-response, and so on), use different scoring methods, and testing at different stages (before the entrance to the ITEP, before the profession). The results of such studies could underpin better decisions on the inclusion of the SJT in the selection of prospective teachers in Oman.

Appendices

Appendix 1

Summary of studies on finding work-related non-cognitive attributes (Non-teaching context)

Study	Context	Method	Sample	Attributes in results
Casner-Lotto and Barrington (2006)	Skills required for new entrants into the U.S. workforce	Survey + interviews	N = 400 employees in four organisations in the US	professionalism/work ethic (including personal accountability, effective work habits, working productively with others, and time and workload management), communications, teamwork/collaboration, and critical thinking/problem solving.
Kim and Park (2013)	competencies required for training programmes of airline cabin crew members	Survey (questionnaire with a seven-point Likert-type scale)	N = 447 crew members	appearance and attitude, physical fitness, customer-oriented skills and company loyalty, knowledge of foreign cultures and languages, emotional intelligence, skills for in-flight services, past work experience, and interpersonal skills.
Patterson et al. (2000) and (2013b)	selection criteria for doctors entering training as general practitioners (GPs)	interviews with stakeholders, critical incidents focus groups, behavioural observation, and a validated questionnaire.	In (2013b): stakeholder consultation (n = 205) + a validation questionnaire (n = 1082) + an expert panel (n = 6)	empathy and perspective taking, communication skills, clinical knowledge and expertise, conceptual thinking and problem-solving, organisation and management of resources, professional integrity, coping with pressure, effective teamworking, respect for diversity and the law, learning and development of self and others, and leading for continuing improvement
Patterson, Ferguson, and Thomas (2008)	select for postgraduate medical training in three secondary care specialties	observation, focus groups, interviews, and reviews of research literature. + a validation questionnaire.		Empathy and sensitivity, Communication skills, Clinical knowledge and technical expertise, Conceptual thinking and problem solving, Organisation and planning, Professional integrity, Managing others, Team involvement, Legal, ethical and political awareness, Vigilance and situational awareness, Learning and personal development, Teaching, Coping with pressure, and Personal attributes.

Study	Context	Method	Sample	Attributes in results
Shultz and Zedeck (2011)	law school admission	interviewing groups of stakeholders and focus groups		Analysis and Reasoning, Creativity/Innovation, Problem Solving, Practical Judgment, Researching the Law, Fact Finding, Questioning and Interviewing, Influencing and Advocating, Writing, Speaking, Listening, Strategic Planning, Organizing and Managing One's Own Work, Organizing and Managing Others (Staff/Colleagues), Negotiation Skills, Able to See the World Through the Eyes of Others, Networking and Business Development, Providing Advice & Counsel & Building Relationships with Clients, developing Relationships within the Legal Profession, Evaluation, Development, and Mentoring, Passion and Engagement, Diligence, Integrity/Honesty, Stress Management, Community Involvement and Service, Self-Development
El-Baz & El-Sayegh, 2015	build a competency domain model for engineering managers	a review of related literature + interviews + a survey	60 practicing EMs in the UAE	In the 'People' domain: Effective Communication -Teams and Teamwork - Motivating Self and Others - Negotiation and Conflict Resolution - Vision and Strategic Thinking - Enthusiasm and Inspiration - Truthfulness and Integrity - Mentoring and Coaching Others.

Appendix 2

Summary of attributes related to teacher effectiveness

Attribute(s) related to teacher effectiveness	Focus/ Target sample	Study
pedagogical content knowledge, quality of instructions, classroom climate, classroom management, teacher beliefs, and professional behaviours	Define 'great teaching (A review)	Coe, Aloisi, Higgins, & Major, 2014
sensitivity to students' needs, knowledge of subject-matter content and pedagogy, and the ability to put that knowledge into practice	Measures of Effective Teaching (MET) project	Kane, McCaffrey, Miller, and Staiger (2013)
emotional support, classroom organisation, and instructional support	Working teachers	Pianta and Hamre (2009)
grit (a disposition toward perseverance and passion for long-term goals)	novice teachers	Robertson-Kraft & Duckworth, 2014
self-efficacy beliefs influence commitment to the teaching profession	preservice and in-service teachers (a meta-analysis study)	Chesnut & Burley, 2015
Teachers' enthusiasm	secondary teachers	Keller, Goetz, Becker, Morger, & Hensley, 2014
reflective practice, ongoing learning, engagement in research and innovation, collaboration, and commitment to school development	teacher competences in European policy	Caena, 2014
knowledge and instructional skills in teaching and learning; organisation/management skills; knowledge of diverse learners; effective collaboration with colleagues, parents, social services and the community; attitude to professional development; and development of ethical stand.	Investigate student teachers' readiness-for-the-job (sample: a student sample from four teacher education colleges)	Mohamed et al. (2017)
cognitive ability – conscientiousness – agreeableness - self-regulation – resilience – extraversion	Applicants to TEPs	Sautelle et al 2015
enthusiasm for the subject, the ability to communicate, and the ability to work with others	current students and applicants in a TEP	Turner & Turner, 1997

Attribute(s) related to teacher effectiveness	Focus/ Target sample	Study
The teacher as a person (caring - fairness and respect - interactions with students - enthusiasm - motivation - dedication to teaching - reflective practice) • Classroom management and organization • Organizing and orienting for instruction • Implementing instruction • Monitoring student progress and potential • Professionalism	A book: Qualities of effective teachers	Stronge, 2007
Knowledge: about the subject, pupils, curriculum, teaching methods, the influence on teaching and learning of other factors, and knowledge about one's own teaching skills - Decision-making: before, during and after a lesson - Action: behavior to foster pupil learning.	A book: Essential teaching skills	Kyriacou (2007)

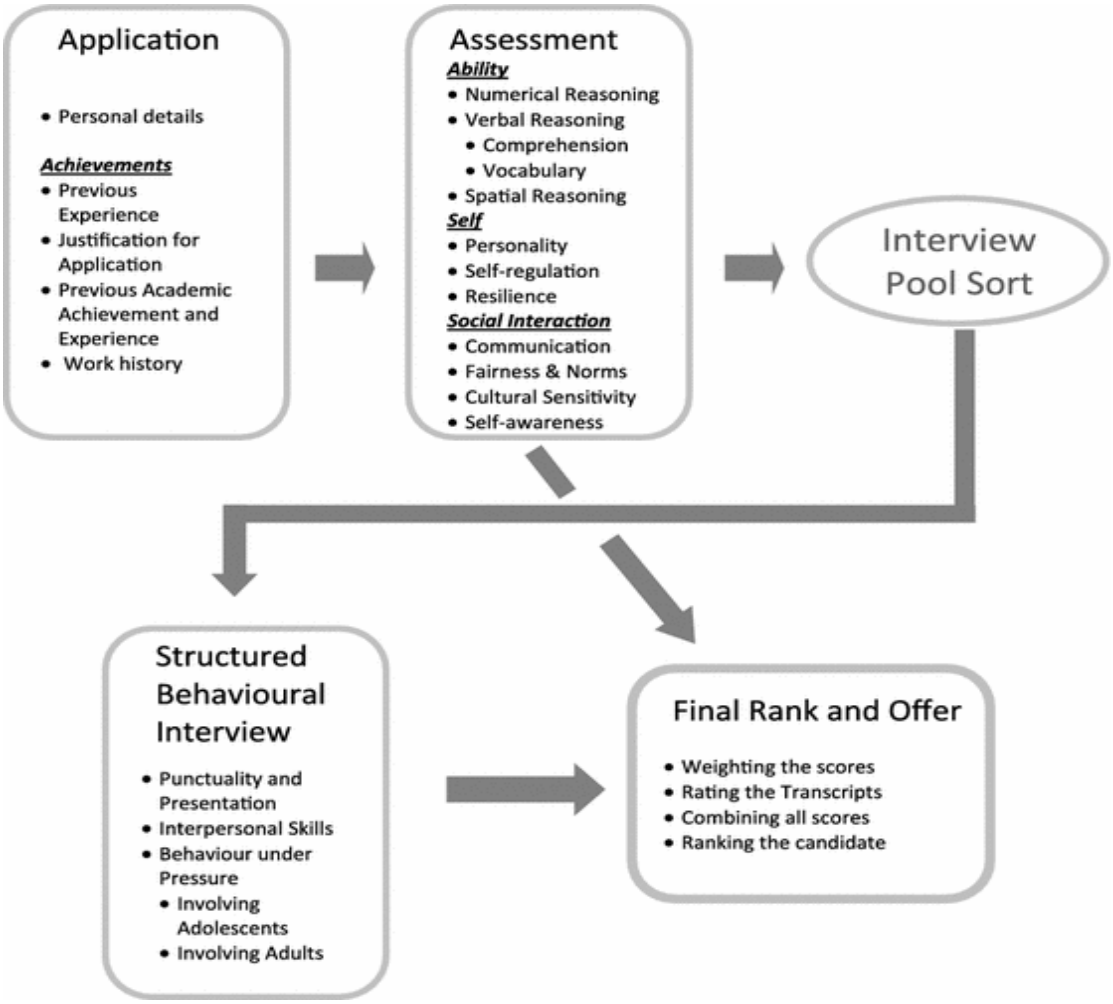
Appendix 3

Summary of attributes and measurements used for the admission into ITEPs + proposed models for selection method

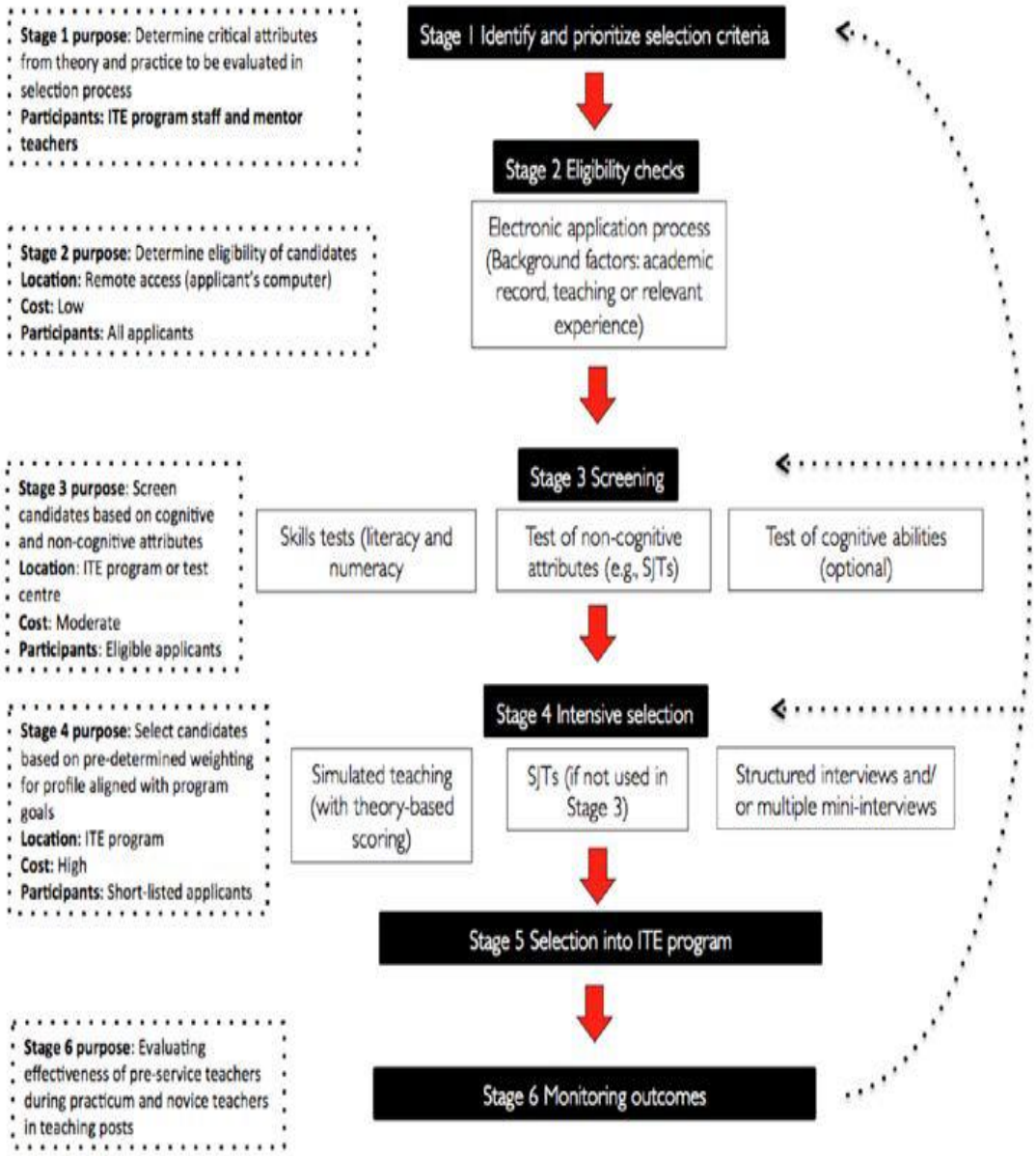
Study	Context/ Target sample	Attributes	Selection measures
Casey & Childs 2007	review the criteria utilized throughout North America to select prospective teachers into education programs	• content knowledge • pedagogical knowledge: what to do in a situation. • pedagogical skills: being able to do: communication skills. • attitudes: attitudes toward morality and ethics - self-efficacy.	• Grade point average. GPA • Written Profile (Responses to Questions About Relevant Experiences and Interest in Teaching) • Interview • Letters of Reference • Standardized Test Results • Performance in Pre-Requisite Education Courses
Hashim, Damio, & Hussin, 2013; Mat Kasim et al., 2012; Othman et al., 2008; Ramli et al., 2013	Malaysian Educators Selection Inventory	• personality traits: intellectual, analytical, persistence, extrovert, helping, achievement, assertive, leadership, autonomy, self-critic, honesty. • career interest: realistic, investigative, conventional, enterprising, social, artistic. • integrity: trustworthiness, honesty, wisdom. • emotional intelligence: self-awareness, emotional expression, aware others, resilience, interpersonal relationship, relationship quotient.	• different sets of criteria are used by different authorities • First, applicants are filtered according to their academic achievement. The candidates then sit the MEdSI as an entry examination, which evaluates the applicants on intrinsic qualities such as personality, interest in a teaching career, integrity, and emotional intelligence. Finally, applicants are interviewed
OMoE, 2012	Singapore	A CV to indicate his/her academic qualifications. An admission test in literacy Successful applicants are interviewed to evaluate their attitude, aptitude, and personality.	

Study	Context/ Target sample	Attributes	Selection measures
OMoE, 2012	Finland	Take a national admission test to measure their literacy, numeracy, and problem-solving skills. A university assessment test to measure abilities in processing information, thinking critically, and synthesising data. Successful applicants are interviewed to check their motivation to teach, motivation to learn, communication skills, and emotional intelligence	
Bieri & Schuler, 2011	the Zurich University of Teacher Education,	• communication: express clear thoughts, appropriate language, ask quotations, presentation methods. • cooperation: aware of others, support others, make compromises, takes an integration role. • assertiveness: convincing others. • motivation: commitment and involvement in teaching. • fact finding: knowledge and problem solving.	Assessment center
Wang & Fwu, 2007	criteria for selection in TE programmes in Taiwan	• academic ability • character and moral conduct • oral and written expression: • educational knowledge: • values/ attitudes for education • Motivation/enthusiasm for teaching • Psychological/personality aptitude • Social/ interpersonal skills	- For 'academic ability': academic records or subject-related tests. - For 'character and moral conduct': official records and recommendations letters. - 'Oral expressions': interviews, public speeches, and other classroom situations. - For 'written expressions': applicants' statements and other language tests. - In terms of 'general educational knowledge': written exams on educational issues, practices, and theories, - for 'attitudes and motivations': a combination of personal interviews, autobiographical statements, and recommendations. - For 'psychological aptitude': standardised tests, or personal interviews and recommendations. - for 'social/interpersonal skills': records of community service and leadership activities, written statements and recommendations.

A proposed model for ITEP selection (in Bowels et al., 2014)



A proposed model for ITEP selection (in Klassen & Kim, 2017a)



Appendix 4

The interview form[1]

UNIVERSITY of York

Informed Consent Form (Interview)

Dear/ Participant,

This study is conducted by Waleed Al Hashmi, a PhD student at the University of York, and aims to better understanding of the non-cognitive domains (such as motivation, self-control, resilience and communication) of effective teacher in Oman, and to apply the understanding to selection into teacher education programs, and in particular, at Sultan Qaboos University. I believe this research has the potential to make a significant contribution to knowledge and practice in education to improve teacher effectiveness by focusing on enhancing teacher selection processes.

I am writing to ask if you are able to take part in this study.

Your Participation:-
You will participate in a face-to-face interview for about 20- 30 minutes to help me understanding the key non-cognitive skills for effective teacher in Oman. The interview will take place in a location and at a time convenient to you.

If agreement is obtained, the interview will be recorded, if not, detailed notes will be taken. At the end of the interview I shall summarise the discussion and give you the opportunity to comment on the written record.

Your participation in this study is totally confidential. Your name will be protected and anonymised in the writing up process and in any future use of the data. The data will be kept for five years in a secure place accessible only to the researcher, and will be used for research.

Your participation is entirely voluntary and you can withdraw from the research at any time during your involvement or for up to seven days after that by emailing the researcher. If you do so, any data you have provided will be destroyed.

A proposal for this research was reviewed by the Department of Education Ethics Committee. If you have any concerns about the researcher or the way the research is being conducted, please contact the chair of the committee via email education-research-administrator@york.ac.uk

Thank you for taking the time to read this information.

Yours sincerely
Waleed Al Hashmi Wtaa500@york.ac.uk, Mobile Oman: oo96892800190, Mobile – U.K: 0044(0)7594863343

Your consent:-

I have agreed to participate in this study done by Waleed Al Hashmi on the basis of the information given above.

Name of participant: _____ Date: _____

[1] The original form was presented in Arabic.

Interview questions for college tutors:-

- The term 'non-cognitive domains' refers to the non-academic skills such as motivation, self-control, resilience and communication. What are the key non-cognitive domains you want to see in the student-teacher at the College of Education?
- From your opinion, do new entrants to the college of Education have the necessary non-cognitive competencies required for effective teacher? Please, give examples if possible.
- From your opinion, does the current selection procedure able to measure the non-cognitive competencies of new applicants? And why?
- How does the College of Education build and measure the non-cognitive domains of its' current students to ensure that they graduate as effective new teachers in the future?
- Any further comments…

Interview questions for supervisors/school principals:-

- The term 'non-cognitive domains' refers to the non-academic skills such as motivation, self-control, resilience and communication. What are the key non-cognitive domains you want to see in the new teacher in your school/ under your supervision?
- From your opinion, do new teachers have the necessary non-cognitive domains required for effective teacher? Please, give examples if possible.
- The current selection procedure in the entrance for initial teacher education programs focus mainly in the applicants' results in secondary school and some interviews. From your opinion, does this procedure able to measure the non-cognitive domains of new applicants?
- How do you measure the non-cognitive competencies of current teachers?
- Any further comments…

Appendix 5

Results of Phase 1 compared to the domains found in Klassen's work in the UK

Items (Klassen et al., 2014b & 2017b)	Other related items from Oman[1]	Summary (Questionnaire items)
Domain: 'Empathy & Communication' changed to 'Communication Skills'		
• Active listening • Open dialogue with both pupils and colleagues • Responsive to students' needs • Adapt the style of communication and nature of dialogue appropriately • Seek advice pro-actively • Responsive to professional feedback	• Care of pupils/ advise pupils for good attitudes (JD) • Good attitudes to work with SEN pupils (JD - ITEP) • Beliefs about students learning (In) • Raise pupils' motivation (CS) • Direct pupils for self-learning (CS) • Show humanity with pupils, peers and parents (In) • Good relationship with school, peers, pupils and parents (AA) • Cooperative / involved in school activities (JD) • Collaborates with schools, families and community (ITEP) • Social activities inside and outside school (AA) • Social worker - care of the community (In)	• Humanistic in relation to others. • Shows a concern and understanding for pupils' needs • Beliefs about the pupils' ability to learn • Good attitude towards pupils with learning difficulties • Collaborative • Uses appropriate communication style to suit recipients • Exhibits active listening
Domain: Organization & Planning		
• Manage competing priorities • Display time management skills effectively • Display organization skills effectively	• Planning/ annual and daily plans (JD – CS – AA - In) • Effective classroom management (CS – AA - In) • Leadership (In)	• Good in managing competing priorities • Displays good time management skills • Displays good organisation skills • Good planning skills • Good classroom management

[1] JD= Job Description CS= Classroom Supervision criteria AA= Annual Appraisal criteria ITEP= Documents from the ITEP In= Interview

Items (Klassen et al., 2014b & 2017b)	Other related items from Oman[1]	Summary (Questionnaire items)
Domain: Resilience & Adaptability		
• Demonstrate the capability to remain resilient under pressure • Demonstrates adaptability and an ability to change lessons and the sequence of lessons accordingly where required • Awareness of their level of competence • Confidence to seek assistance, as appropriate • Confidence to make decisions independently, as appropriate • Comfortable with challenges to own knowledge • Not disabled by constructive, critical feedback • Uses effective coping strategies	• Accept advice and feedback (AA) • Flexible (In) • Confident - not shy (In) • Patient (In)	• Demonstrates the capability to remain resilient under stress • Comfortable with challenges to own knowledge • Not disabled by remarks and feedback • Uses appropriate coping strategies • Demonstrates high confidence • Seeks help when necessary
New domains		
Enthusiasm and Motivation	• Strengthen the national and job loyalty (JD - In) • Demonstrate the spirit of the citizenship (In) • Commitment (committed to the job roles)/ discipline (JD - AA) • Show strong and reliable concern to be a teacher (SQU) • Show positive attitudes towards teaching/ proud of his/her career/ Show passionate about teaching (In) • Self-development (JD – AA – ITEP - In) • Motivation (In) • Enthusiasm (In) • Initiative (In)	• Commitment to the job roles • Shows strong and reliable concern to be a teacher • Aware of national and job loyalty • Seeks professional development • Takes pleasure in doing teaching tasks

Items (Klassen et al., 2014b & 2017b)	Other related items from Oman[1]	Summary (Questionnaire items)
Professional Ethics	• Observe Omani, Islamic and professional ethics/values in performing his/her professional tasks (ITEP) • Show high consideration to the Islamic and Omani values (ITEP) • Show religious faith & conscientious (In) • Develop positive attitudes and values (JD - CS) • Develop positive attitudes towards the profession and contributes effectively to it (ITEP) • Being as a model for pupils – inspiring (In) • Responsible (In) • Honest (In) • Reliable (In) • Fairness (In) • Wise (wisdom) (In)	• Shows high consideration to the Islamic, Omani and professional ethics • A good model for pupils • Accepts taking responsibility • Trustworthy • Treats others fairly • Demonstrates respect for pupils and colleagues
Others (excluded because can be learned, or not hard to measure by other selection tools)	• Self-assessment (JD – CS - In) • Personality traits (in good health) (In) • Care about his/her appearance (AA) • Strong personality (AA) • Has a professional appearance and behavior (ITEP)	

Appendix 6

The questionnaire[1]

12/8/2015 Non-Cognitive Domains of Effective Teacher in the Government Schools (5-12) in Oman

Non-Cognitive Domains of Effective Teacher in the Government Schools (5-12) in Oman

Dear/ the participant,

This study is conducted by Waleed Al Hashmi, a PhD student at the University of York, to increase understanding of the non-cognitive domains of effective teacher in government schools (5-12) in Oman, and to apply the understanding to selection into initial teacher education programs, and in particular, in the admission process at the College of Education at Sultan Qaboos University.
Non-cognitive domains in this study mean the non-academic attributes or soft competencies such as resilience, motivation and communication skills.
I believe this study has the potential to make a significant contribution to knowledge and practice in education to improve teacher effectiveness by focusing on enhancing teacher selection processes in early times, before enrolling to the College of Education.
I am writing to ask if you are able to take part in this study.

Your Participation:-
You are kindly asked to complete an online questionnaire telling your opinions of the key non-cognitive domains for effective teacher in government schools (5-12) in Oman.
It will take just about 5-10 minutes, but your answers are very crucial to the aim of the study.

Your participation in this study is totally confidential. Your answers will be protected and anonymised in the writing up process and in any future use of the data. The data will be kept for about five years in a secure place accessible only to the researcher, and will be used for research.

Your participation is entirely voluntary and you can withdraw from the research at any time during your involvement or for up to seven days after that by emailing the researcher. If you do so, any data you have provided will be destroyed.

A proposal for this research was reviewed by the Department of Education Ethics Committee at the University of York. If you have any concerns about the researcher or the way the research is being conducted, please contact the chair of the committee via email education-research-administrator@york.ac.uk.

Thank you for taking the time to read this information.

Yours sincerely
Waleed Al Hashmi Wtaa500@york.ac.uk, Mobile Oman: oo96892800190, Mobile – U.K: 0044(0)7594863343

*Required

1. **Your Consent:-** *
   Mark only one oval.
   - YES: I understand the information above and I agree to participate.
   - NO: I prefer not to participate in this study (please exit if you choose this).

PART ONE: Personal Information:

https://docs.google.com/a/york.ac.uk/forms/d/1haDQ4cQpGMkwXO2Y4nsy2K6yugAtWN0K-YMu2RZQOVedit 1/7

[1] The original questionnaire was distributed in Arabic and online format.

2. **Gender:-** *
Mark only one oval.

- ◯ Male
- ◯ Female

3. **Job Position:-** *
Mark only one oval.

- ◯ A Teacher/ Senior Teacher
- ◯ A Supervisor/ Senior Supervisor
- ◯ A School principal/ Principal's Assistant

4. **Date of Appointment:-** *

Example: 15 December 2012

5. **Governerate** *
Mark only one oval.

- ◯ Muscat
- ◯ AlBatinah South
- ◯ AlSharqya North

PART TWO: - Non-cognitive Domains of effective teacher:-

At this question, there are five non-cognitive domains of effective teacher in Oman found from the literature, official documents, and some exploratory interviews.
From your point of view and experience, please indicate how important these domains to be in:-

- ☐ - The teacher in government schools (5-12) in Oman.
- ☐ - The candidate applying for the College of Education to be a teacher in the future.

Where 1=not important, 10=very important

A) The importance to be in teacher in government schools (5-12).

6. *
Mark only one oval per row.

	10 very importnt	9	8	7	6	5	4	3	2	1 not important
Communication Skills	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
Orgainsation & Planning	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
Resilience and Adaptability	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
Enthusiasm & Motivation	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
Professional Ethics	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯

B) The importance to be in the candidate applying for the College of Education to be a teacher in the future..

7. *

Mark only one oval per row.

	10 very importnt	9	8	7	6	5	4	3	2	1 not important
Communication Skills										
Orgainsation & Planning										
Resilience and Adaptability										
Enthusiasm & Motivation										
Professional Ethics										

PART THREE: - Non-cognitive Sub-domains:-

Each of the previous non-cognitive domains consists of a group of sub-domains as follow:-

A) To what extent you agree that these sub-domains should be in teacher in government schools (5-12).

8. *

Mark only one oval per row.

	10 extremely agree	9	8	7	6	5	4	3	2	1 dissagree
Humanistic in relation to others.										
Shows a concern and understanding for pupils' needs										
Believes about the pupils' ability to learn										
Good attitude towards pupils with learning difficulties										
Collaborative										
Uses appropriate communication style to suit recipients										
Exhibits active listening										
Good in managing competing priorities										
Displays good time management skills										
Displays good organisation skills										
Good planning skills										
Good classroom managemnt										
Demonstrates the capability to remain resilient under stress										
Comfortable with challenges to own knowledge										
Not disabled by remarks and feedback										
Uses appropriate coping strategies										
Demonstrates high confidence										
Seeks help when necessary										
Commitment to the job roles										
Shows strong and reliable concern to be a teacher										
Aware of national										

and job loyalty	
seeks professional developmnt	
Takes pleasure in doing teaching tasks	
Shows high consideration to the Islamic, Omani and professional ethics	
A good model for pupils	
Accepts taking responsibility	
Trustwothy	
Treats others fairly	
Demonstrates respect for pupils and colleagues	

B) To what extent you agree that these sub-domains should be in the candidate applying for the College of Education to be a teacher in the future..

9. *

Mark only one oval per row.

	10 extremely agree	9	8	7	6	5	4	3	2	1 disagree
Humanistic in relation to others.										
Shows a concern and understanding for pupils' needs										
Believes about the pupils' ability to learn										
Good attitude towards pupils with learning difficulties										
Collaborative										
Uses appropriate communication style to suit recipients										
Exhibits active listening										
Good in managing competing priorities										
Displays good time management skills										
Displays good organisation skills										
Good planning skills										
Good classroom managemnt										
Demonstrates the capability to remain resilient under stress										
Comfortable with challenges to own knowledge										
Not disabled by remarks and feedback										
Uses appropriate coping strategies										
Demonstrates high confidence										
Seeks help when necessary										
Commitment to the job roles										
Shows strong and reliable concern to be a teacher										
Aware of national and job loyalty										
seeks professional										

developmnt									
Takes pleasure in doing teaching tasks									
Shows high consideration to the Islamic, Omani and professional ethics									
A good model for pupils									
Accepts taking responsibility									
Trustworthy									
Treats others fairly									
Demonstrates respect for pupils and colleagues									

Appendix 7

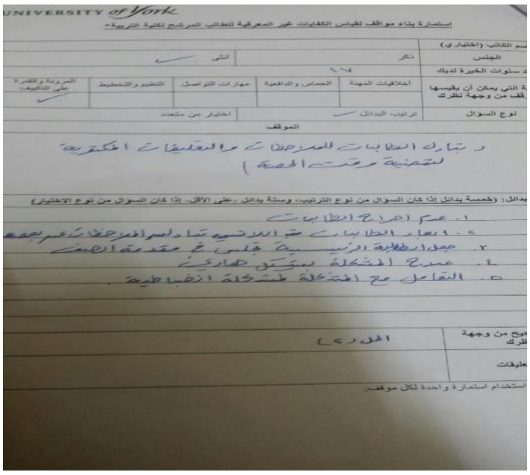
Incidents' collection Form[1]

<table>
<tr><td colspan="5"></td><td>Teacher's Name (optional)</td></tr>
<tr><td colspan="3">Female</td><td colspan="2">Male</td><td>Gender</td></tr>
<tr><td colspan="5"></td><td>Years of experience</td></tr>
<tr><td>Resilence & Adaptability</td><td>Planning & Organisation</td><td>Communication skills</td><td>Enthusiasm & Motivation</td><td>Professional Ethics</td><td rowspan="2">The targeted domain</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="3">Choose best three</td><td colspan="2">Ranking</td><td>Question type</td></tr>
<tr><td colspan="6">The incident</td></tr>
<tr><td colspan="6"></td></tr>
<tr><td colspan="6">The responses (5 for 'ranking' – 6 for 'select best three')</td></tr>
<tr><td colspan="6"></td></tr>
<tr><td colspan="6"></td></tr>
<tr><td colspan="6"></td></tr>
<tr><td colspan="6"></td></tr>
<tr><td colspan="6"></td></tr>
<tr><td colspan="6"></td></tr>
<tr><td colspan="5"></td><td>The Answer from your perspective</td></tr>
<tr><td colspan="5"></td><td>Any comments</td></tr>
</table>

[1] The original form was in Arabic.

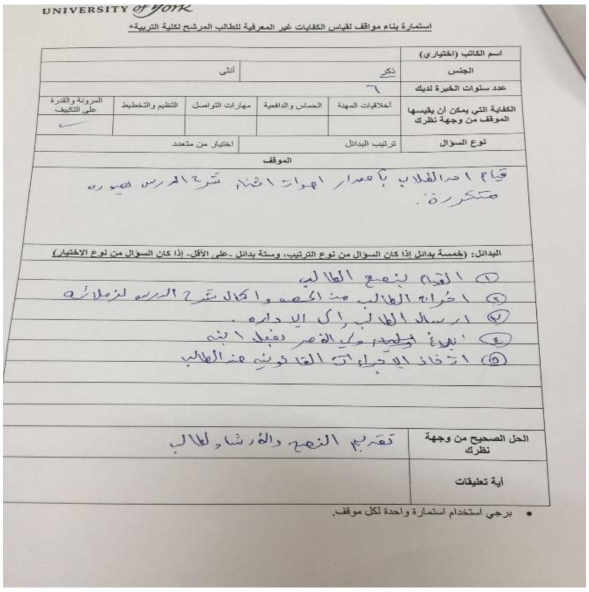
Appendix 8

Examples of the collected incidents from the Omani teachers



UNIVERSITY of York

استمارة بناء مواقف لقياس الكفايات غير المعرفية للطالب المرشح لكلية التربية*



UNIVERSITY of York

استمارة بناء مواقف لقياس الكفايات غير المعرفية للطالب المرشح لكلية التربية*

اسم الكاتب (اختياري)					
الجنس	ذكر		أنثى		
عدد سنوات الخبرة لديك	٦				
الكفاية التي يمكن أن يقيسها الموقف من وجهة نظرك	أخلاقيات المهنة	الحماس والدافعية	مهارات التواصل	التنظيم والتخطيط	المرونة والقدرة على التكيف ✓
نوع السؤال	ترتيب البدائل		اختيار من متعدد		

الموقف

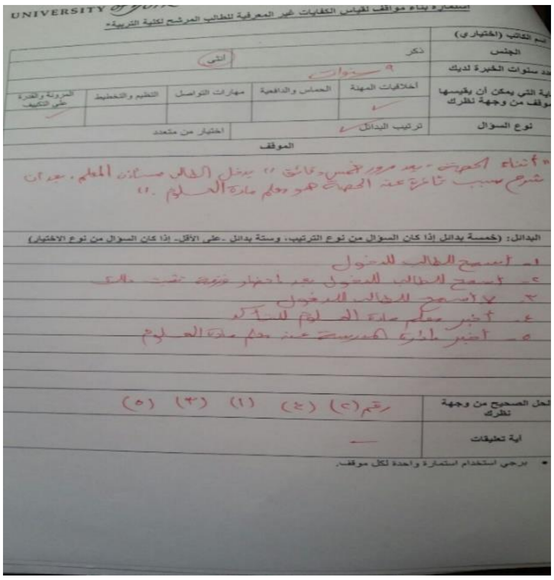
قيام احد الطلاب بإصدار اصوات اثناء شرح الدرس بصورة متكررة.

البدائل: (خمسة بدائل إذا كان السؤال من نوع الترتيب، وستة بدائل -على الأقل- إذا كان السؤال من نوع الاختيار)

1. القيام بنصح الطالب
2. اخراج الطالب من الحصة واكمال شرح الدرس لزملائه
3. ارسال الطالب الى الادارة.
4. [illegible]
5. اتخاذ الاجراءات القانونية ضد الطالب

الحل الصحيح من وجهة نظرك	تقديم النصح والارشاد للطالب
أية تعليقات	

- يرجى استخدام استمارة واحدة لكل موقف.



UNIVERSITY of York

استمارة بناء مواقف لقياس الكفايات غير المعرفية للطالب المرشح لكلية التربية*

الحل الصحيح من وجهة نظرك: رقم (٢) (٤) (١) (٣) (٥)

Appendix 9

The answer sheet used for the review of the collected items[1]

Your comments to the question in terms of: (write any suggested changes in the questions' paper)				The targeted domain (you may choose more than one)					The answer from your prospective (new teacher should do)	Item no.
Suitability for new teacher	The options	Suitability to the context in Oman	Clarity	Professional ethics	Enthusiasm & motivation	Resilience & adaptability	Organisation & planning	Communication skills		
										1
										2
										3
										4

[1] The original sheet was in Arabic.

Appendix 10

Official permissions for data collection in Oman

Sultanate of Oman
Ministry of Higher Education
Directorate General of Colleges
of Applied Sciences

سلطنة عمان
وزارة التعليم العالي
المديرية العامة لكليات العلوم التطبيقية

الرقم 6 / 5 / 8‏ ‎/ 1 ‎ -2016 ‎
التاريخ: 16 / 8 / 2016م

الدكتور / حمود بن عامر الوردي المحترم
عميد كلية العلوم التطبيقية بالرستاق
السلام عليكم ورحمة الله وبركاته ...

الموضوع: تسهيل مهمة الباحث / وليد بن طالب بن عبدالله الهاشمي

نود إفادة عنايتكم الكريمة بأن الفاضل / وليد بن طالب بن عبدالله الهاشمي — يقوم بإعداد دراسة رسالة الدكتوراه بجامعة يورك بالمملكة المتحدة بعنوان:

(إلى أي مدى يستطيع اختبار الحكم على المواقف (Situational Judgment Test) قياس الكفايات غير المعرفية اللازمة للطالب والمعلم خلال مرحلة قبول الطلبة في برنامج إعداد المعلمين في سلطنة عمان)

حيث تستهدف هذه الدراسة الطلاب الجدد المقبولين للدراسة بتخصصات التربوية في الكلية، وذلك خلال فترة الأسبوع التعريفي للعام الدراسي 2016 / 2017م.

عليه يرجى التكرم بتسهيل مهمة الباحث لإجراء هذه الدراسة، علماً بأن هذه البيانات والنتائج المستخلصة سوف تستخدم لأغراض البحث العلمي فقط. (مرفق أسئلة الاختبار للطلبة).

شاكرين لكم حسن تعاونكم الدائم.

وتفضلوا بقبول فائق الاحترام والتقدير

المدير العام
المدير العام المساعد للشؤون الأكاديمية

/ عزيزة بنت سويل الحربية
مديرة مساعدة مركز البحث العلمي والتطوير

نحو تعليم عال ذي جودة عالية يلبي متطلبات التنمية المستدامة
سلطنة عمان ص.ب: ٨٢ روي - الرمز البريدي: ١١٢ - هاتف ٢٤٣٤٠٥٨٠ / فاكس ٢٤٣٤٠٥٧٨
Sultanate of Oman, P.O.Box: 82 Ruwi, PC 112, Tel:24340580 / Fax 24340578 www.mohe.gov.om

---

Sultan Qaboos University
Deputy Vice Chancellor for Academic Affairs & Community Service

جامعة السلطان قابوس
نائب الرئيس للشؤون الأكاديمية وخدمة المجتمع

إستمارة وارد

٢٠١٥/٠٧/٢٧

رقـــم الوارد : ١٢٢١ / ٢٠١٥	رقـــم الرسالة : ٠
تاريخ الإستلام : ٢٠١٥/٠٧/٢٧	تاريخ الرسالة : ٢٠١٥/٠٧/٢٧
الجهة الوارد منها : رسائل الطلاب	
الموضوع : تسهيل مهمة بحثية لطالب دكتوراة بجامعة يورك بالمملكة المتحدة	
صفة الموضوع :	

ملخص الموضوع

يرجو الطالب/ وليد بن طالب بن عبدالله الهاشمي - طالب دكتوراة بجامعة يورك بالمملكة المتحدة ٢٠١٨-٢٠١٥م - التكرم بالإيعاز لمن يلزم بتوفير الدعم لدراسته البحثية لنيل درجة الدكتوراة بجامعة يورك بالمملكة المتحدة بعنوان: "إلى أي مدى يستطيع اختبار الحكم على المواقف (Situational Judgment Test) قياس الكفايات غير المعرفية اللازمة للمعلم العماني خلال مرحلة قبول الطلبة بكلية التربية بجامعة السلطان قابوس؟". مرفق

المـــــرفقات :

التعليمات

د. عميد كلية التربية:
للدراسة و الإفادة بالرأي.

مكتب عميد كلية التربية
الرقم : ٧٧٧/ ٢٠١٥
التاريخ : ٢٨/ ٧/ ٢٠١٥
وارد

صندوق البريد : ٥٠ الخوض – الرمز البريدي : ١٢٣ مسقط – سلطنة عُمان، هاتف : ٢٤١٤١٠٧٦ (٩٦٨+) ، فاكس : ٢٤٤١٣١٣٥ (٩٦٨+)
P.O. Box: 50 Al-Khod, Postal Code: 123 Muscat - Sultanate of Oman - Tel.: (+968) 24141076 - Fax: (+968) 24413135
E-mail: DVCAACS@squ.edu.om البريد الإلكتروني

Appendix 11

Factor Analysis of the 38-SJT items

METHOD = CORRELATION + KMO and Bartlett's
Extraction method: principle components
ROTATION OBLIMIN
MISSING PAIRWISE
/PRINT INITIAL CORRELATION KMO EXTRACTION ROTATION
/FORMAT SORT BLANK(.3)
/PLOT EIGEN
/CRITERIA MINEIGEN(1) ITERATE(25)
/EXTRACTION PC
/CRITERIA ITERATE(25) DELTA(0)

First: Factor analysis for 'choose best three' (24 items)

Correlation Matrix

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	1.000	.020	-.028	-.065	-.047	.040	-.066	.037	.000	.072	.032	-.042	.109	.092	.080	-.021	.037	.022	.016	.101	.076	-.057	.055	-.033
2	.020	1.000	.076	.031	.124	.123	-.061	.119	.107	.025	-.034	.053	.065	.188	.013	.284	.035	.028	-.009	.011	-.090	-.066	.104	.151
3	-.028	.076	1.000	.297	.069	.070	.029	.056	.114	.101	.147	.082	-.015	.062	.024	.106	.128	.254	.046	.083	.137	.120	.142	.066
4	-.065	.031	.297	1.000	.022	.022	.100	-.002	-.139	-.079	.061	.036	.039	.036	.019	-.045	.034	.126	.157	.022	.074	-.037	.015	.102
5	-.047	.124	.069	.022	1.000	.028	.142	.005	.069	.000	-.087	-.092	.030	-.036	.004	.036	-.087	.151	.020	.051	-.068	.022	.096	.132
6	.040	.123	.070	.022	.028	1.000	.125	.135	.075	-.113	.290	.135	.155	.113	.047	.241	.233	.038	-.083	-.097	-.019	.147	.139	.131
7	-.066	-.061	.029	.100	.142	.125	1.000	.136	-.103	-.127	.028	-.035	.019	.039	.083	.037	-.057	.200	.113	-.158	.005	.036	-.056	-.058
8	.037	.119	.056	-.002	.005	.135	.136	1.000	.101	.085	.130	.055	.061	.021	.077	.174	.042	.154	.139	.051	.001	.002	.212	.165
9	.000	.107	.114	-.139	.069	.075	-.103	.101	1.000	.153	.110	.194	.162	.090	.107	.181	.036	.090	.071	.106	-.008	.046	.052	.173
10	.072	.025	.101	-.079	.000	-.113	-.127	.085	.153	1.000	-.208	-.037	.166	.082	-.010	.135	.140	-.021	.010	.202	.121	-.101	.052	-.052
11	.032	-.034	.147	.061	-.087	.290	.028	.130	.110	-.208	1.000	.212	.110	.033	.249	.153	.118	.188	.172	.110	.145	-.028	.043	.120
12	-.042	.053	.082	.036	-.092	.135	-.035	.055	.194	-.037	.212	1.000	.344	.186	.118	.122	.027	.094	.129	.157	.208	.136	.154	.104
13	.109	.065	-.015	.039	.030	.155	.019	.061	.162	.166	.110	.344	1.000	.319	.305	.111	-.027	.244	.166	.116	.114	-.019	.125	.113
14	.092	.188	.062	.036	-.036	.113	.039	.021	.090	.082	.033	.186	.319	1.000	.134	.245	.042	.159	.099	-.005	.152	-.133	.166	.229

1	5	.080	.013	.024	.019	.004	.047	.083	.077	.107	-.010	.249	.118	.305	.134	1.000	.016	.086	.225	.151	.092	.133	-.018	.013	.264
1	6	-.021	.284	.106	-.045	.036	.241	.037	.174	.181	.135	.153	.122	.111	.245	.016	1.000	.083	.153	.175	.058	-.001	.063	.162	.111
1	7	.037	.035	.128	.034	-.087	.233	-.057	.042	.036	.140	.118	.027	-.027	.042	.086	.083	1.000	-.091	.032	-.041	.032	-.020	.130	.052
1	8	.022	.028	.254	.126	.151	.038	.200	.154	.090	-.021	.188	.094	.244	.159	.225	.153	-.091	1.000	.218	.062	.007	.150	.128	.088
1	9	.016	-.009	.046	.157	.020	-.083	.113	.139	.071	.010	.172	.129	.166	.099	.151	.175	.032	.218	1.000	.044	.029	.021	.146	.184
2	0	.101	.011	.083	.022	.051	-.097	-.158	.051	.106	.202	.110	.157	.116	-.005	.092	.058	-.041	.062	.044	1.000	.210	-.102	.129	.109
2	1	.076	-.090	.137	.074	-.068	-.019	.005	.001	-.008	.121	.145	.208	.114	.152	.133	-.001	.032	.007	.029	.210	1.000	-.079	.067	.183
2	2	-.057	-.066	.120	-.037	.022	.147	.036	.002	.046	-.101	-.028	.136	-.019	-.133	-.018	.063	-.020	.150	.021	-.102	-.079	1.000	.239	-.086
2	3	.055	.104	.142	.015	.096	.139	-.056	.212	.052	.052	.043	.154	.125	.166	.013	.162	.130	.128	.146	.129	.067	.239	1.000	.041
2	4	-.033	.151	.066	.102	.132	.131	-.058	.165	.173	-.052	.120	.104	.113	.229	.264	.111	.052	.088	.184	.109	.183	-.086	.041	1.000

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.585
Bartlett's Test of Sphericity	Approx. Chi-Square	430.879
	df	276
	Sig.	.000

Communalities

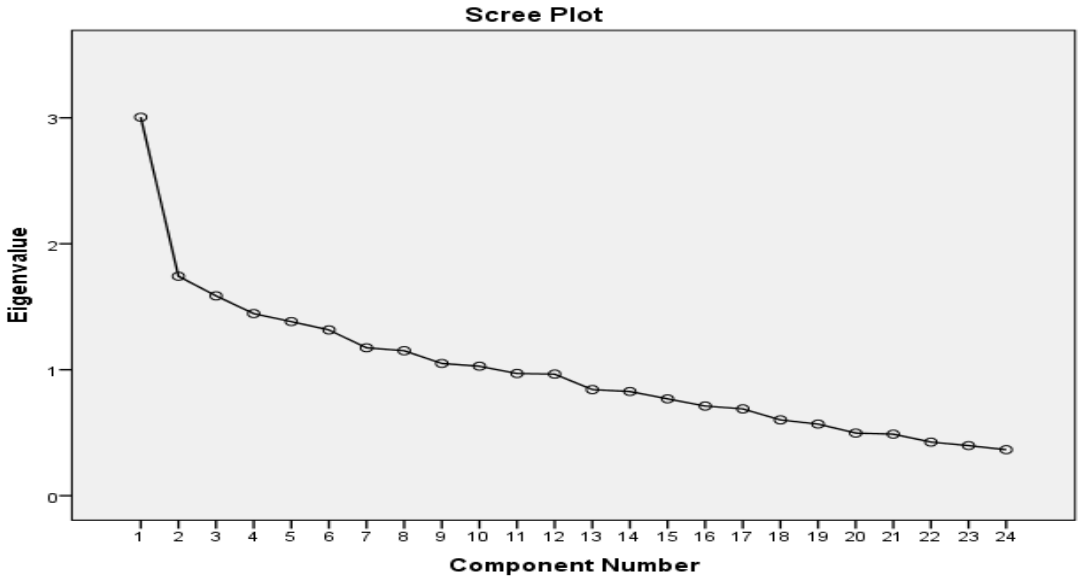
	Initial	Extraction
Choose3 Q.1 scores	1.000	.595
Choose3 Q.2 scores	1.000	.580
Choose3 Q.3 scores	1.000	.702
Choose3 Q.4 scores	1.000	.641
Choose3 Q.5 scores	1.000	.668
Choose3 Q.6 scores	1.000	.695
Choose3 Q.7 scores	1.000	.550
Choose3 Q.8 scores	1.000	.565
Choose3 Q.9 scores	1.000	.646
Choose3 Q.10 scores	1.000	.769
Choose3 Q.11 scores	1.000	.632
Choose3 Q.12 scores	1.000	.640
Choose3 Q.13 scores	1.000	.653
Choose3 Q.14 scores	1.000	.666
Choose3 Q.15 scores	1.000	.541
Choose3 Q.16 scores	1.000	.525
Choose3 Q.17 scores	1.000	.612
Choose3 Q.18 scores	1.000	.591
Choose3 Q.19 scores	1.000	.632
Choose3 Q.20 scores	1.000	.591
Choose3 Q.21 scores	1.000	.476
Choose3 Q.22 scores	1.000	.660
Choose3 Q.23 scores	1.000	.663
Choose3 Q.24 scores	1.000	.589

Extraction Method: Principal Component Analysis.

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.005	12.521	12.521	3.005	12.521	12.521
2	1.742	7.260	19.781	1.742	7.260	19.781
3	1.586	6.610	26.390	1.586	6.610	26.390
4	1.445	6.023	32.413	1.445	6.023	32.413
5	1.382	5.759	38.172	1.382	5.759	38.172
6	1.316	5.483	43.655	1.316	5.483	43.655
7	1.174	4.892	48.547	1.174	4.892	48.547
8	1.151	4.796	53.343	1.151	4.796	53.343
9	1.050	4.374	57.718	1.050	4.374	57.718
10	1.028	4.283	62.001	1.028	4.283	62.001
11	.970	4.043	66.044			
12	.965	4.022	70.066			
13	.842	3.510	73.576			
14	.827	3.447	77.023			
15	.769	3.203	80.226			
16	.712	2.967	83.193			
17	.689	2.870	86.063			
18	.602	2.507	88.571			
19	.569	2.369	90.939			
20	.497	2.072	93.011			
21	.489	2.036	95.048			
22	.426	1.773	96.821			
23	.398	1.657	98.478			
24	.365	1.522	100.000			

Extraction Method: Principal Component Analysis.



Component Matrix[a]

	Component									
	1	2	3	4	5	6	7	8	9	10
Choose3 Q.13 scores	.545					-.315				
Choose3 Q.12 scores	.486					-.324		-.382		
Choose3 Q.14 scores	.480				-.304		.321	-.424		
Choose3 Q.18 scores	.476	.327		.345						
Choose3 Q.16 scores	.475		.458							
Choose3 Q.24 scores	.456					.327	-.400			
Choose3 Q.15 scores	.445		-.376							
Choose3 Q.10 scores		-.546								.377
Choose3 Q.7 scores		.533					.321			
Choose3 Q.20 scores		-.463			.301					
Choose3 Q.2 scores			.476							
Choose3 Q.21 scores		-.332	-.385							
Choose3 Q.5 scores				.544					.483	
Choose3 Q.6 scores	.367	.326	.314	-.483					.303	
Choose3 Q.11 scores	.450			-.464						
Choose3 Q.17 scores				-.405		.394				
Choose3 Q.3 scores	.343				.584					
Choose3 Q.22 scores		.419			.370	-.515				
Choose3 Q.4 scores			-.348		.301	.460				
Choose3 Q.1 scores							.432		.405	
Choose3 Q.9 scores	.370						-.424			.400
Choose3 Q.8 scores	.359							.515		
Choose3 Q.19 scores	.400								-.547	
Choose3 Q.23 scores	.401		.309		.353					-.454

Extraction Method: Principal Component Analysis.
a. 10 components extracted.

Pattern Matrix[a]

a. Rotation failed to converge in 25 iterations. (Convergence = .000).

Second: Factor Analysis for ‘ranking’ (14 items)

Correlation Matrix

		Ranking Q.1 scores	Ranking Q.2 scores	Ranking Q.3 scores	Ranking Q.4 scores	Ranking Q.5 scores	Ranking Q.6 scores	Ranking Q.7 scores	Ranking Q.8 scores	Ranking Q.9 scores	Ranking Q.10 scores	Ranking Q.11 scores	Ranking Q.12 scores	Ranking Q.13 scores	Ranking Q.14 scores
Correlation	Ranking Q.1 scores	1.000	.173	.001	.124	.049	.052	.053	.129	.075	.036	-.044	.072	-.038	.022
	Ranking Q.2 scores	.173	1.000	.150	.181	.242	.053	.140	.250	.205	.108	.064	.076	.038	.178
	Ranking Q.3 scores	.001	.150	1.000	.254	.182	.323	.231	.164	.312	.174	.335	.188	.214	.307
	Ranking Q.4 scores	.124	.181	.254	1.000	.310	.256	.239	.191	.183	.074	.257	.195	.140	.123
	Ranking Q.5 scores	.049	.242	.182	.310	1.000	.299	.222	.148	.352	.138	.223	.307	.300	.275
	Ranking Q.6 scores	.052	.053	.323	.256	.299	1.000	.309	.273	.395	.150	.395	.292	.293	.328
	Ranking Q.7 scores	.053	.140	.231	.239	.222	.309	1.000	.139	.316	.147	.257	.210	.095	.193
	Ranking Q.8 scores	.129	.250	.164	.191	.148	.273	.139	1.000	.218	.135	.201	.058	.201	.271
	Ranking Q.9 scores	.075	.205	.312	.183	.352	.395	.316	.218	1.000	.179	.366	.197	.303	.341
	Ranking Q.10 scores	.036	.108	.174	.074	.138	.150	.147	.135	.179	1.000	.246	-.008	.095	.108
	Ranking Q.11 scores	-.044	.064	.335	.257	.223	.395	.257	.201	.366	.246	1.000	.276	.171	.215
	Ranking Q.12 scores	.072	.076	.188	.195	.307	.292	.210	.058	.197	-.008	.276	1.000	.205	.364
	Ranking Q.13 scores	-.038	.038	.214	.140	.300	.293	.095	.201	.303	.095	.171	.205	1.000	.347
	Ranking Q.14 scores	.022	.178	.307	.123	.275	.328	.193	.271	.341	.108	.215	.364	.347	1.000

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.822
Bartlett's Test of Sphericity	Approx. Chi-Square	324.708
	df	91
	Sig.	.000

Communalities

	Initial	Extraction
Ranking Q.1 scores	1.000	.508
Ranking Q.2 scores	1.000	.558
Ranking Q.3 scores	1.000	.383
Ranking Q.4 scores	1.000	.479
Ranking Q.5 scores	1.000	.428
Ranking Q.6 scores	1.000	.492
Ranking Q.7 scores	1.000	.444
Ranking Q.8 scores	1.000	.515
Ranking Q.9 scores	1.000	.460
Ranking Q.10 scores	1.000	.564
Ranking Q.11 scores	1.000	.571
Ranking Q.12 scores	1.000	.606
Ranking Q.13 scores	1.000	.584
Ranking Q.14 scores	1.000	.582

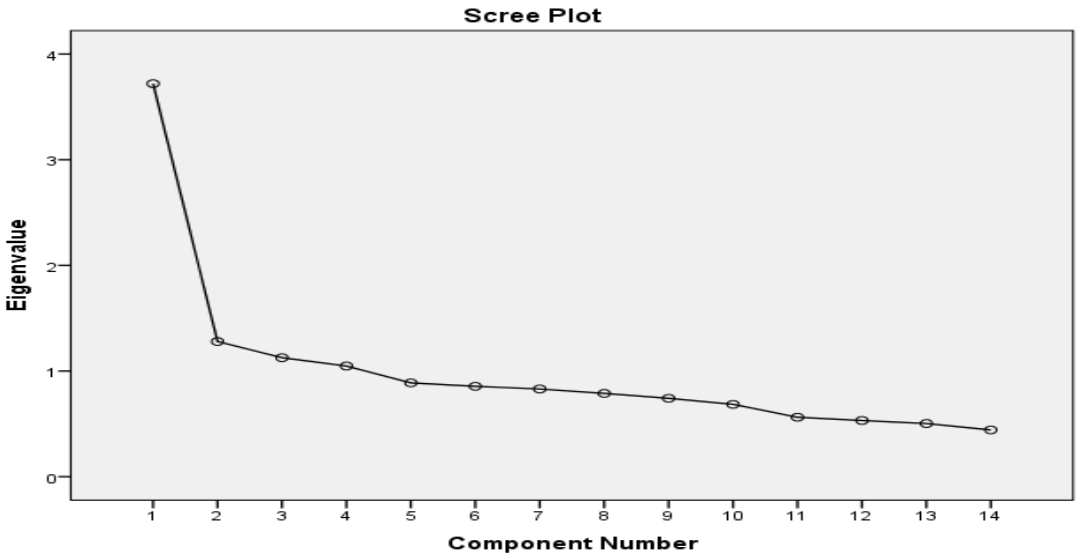
Extraction Method: Principal Component Analysis.

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings[a]
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	3.720	26.570	26.570	3.720	26.570	26.570	3.042
2	1.279	9.137	35.708	1.279	9.137	35.708	1.621
3	1.126	8.044	43.752	1.126	8.044	43.752	1.100
4	1.047	7.480	51.232	1.047	7.480	51.232	2.538
5	.888	6.342	57.574				
6	.855	6.109	63.683				
7	.830	5.927	69.610				
8	.788	5.629	75.239				
9	.742	5.299	80.538				
10	.685	4.891	85.430				
11	.562	4.016	89.446				
12	.532	3.800	93.246				
13	.503	3.593	96.839				
14	.443	3.161	100.000				

Extraction Method: Principal Component Analysis.

a. When components are correlated, sums of squared loadings cannot be added to obtain a total variance.



Component Matrix[a]

	Component			
	1	2	3	4
Ranking Q.6 scores	.672			
Ranking Q.9 scores	.665			
Ranking Q.14 scores	.605		-.318	.310
Ranking Q.11 scores	.598		.367	
Ranking Q.5 scores	.590			
Ranking Q.3 scores	.566			
Ranking Q.7 scores	.508			-.374
Ranking Q.12 scores	.501		-.449	-.333
Ranking Q.13 scores	.499			.413
Ranking Q.4 scores	.487			-.429
Ranking Q.8 scores	.452	.343		.437
Ranking Q.1 scores		.663		
Ranking Q.2 scores	.342	.642		
Ranking Q.10 scores	.319		.629	

Extraction Method: Principal Component Analysis.
a. 4 components extracted.

Pattern Matrix[a]

	Component			
	1	2	3	4
Ranking Q.11 scores	.742			
Ranking Q.7 scores	.678			
Ranking Q.4 scores	.551	.309		
Ranking Q.6 scores	.538			.312
Ranking Q.3 scores	.520			
Ranking Q.9 scores	.417			.370
Ranking Q.2 scores		.719		
Ranking Q.1 scores		.703		
Ranking Q.8 scores		.434	.347	.414
Ranking Q.10 scores	.399		.647	
Ranking Q.12 scores	.315		-.599	.300
Ranking Q.13 scores				.786
Ranking Q.14 scores				.732
Ranking Q.5 scores				.341

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.
a. Rotation converged in 14 iterations.

Structure Matrix

	Component			
	1	2	3	4
Ranking Q.11 scores	.718			
Ranking Q.7 scores	.644			
Ranking Q.6 scores	.636			.492
Ranking Q.3 scores	.574			.369
Ranking Q.4 scores	.564	.407		
Ranking Q.9 scores	.562			.529
Ranking Q.2 scores		.724		
Ranking Q.1 scores		.683		
Ranking Q.10 scores	.383		.637	
Ranking Q.12 scores	.427		-.595	.389
Ranking Q.13 scores				.754
Ranking Q.14 scores	.311			.750
Ranking Q.8 scores		.467	.346	.470
Ranking Q.5 scores	.453	.328		.458

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

Component Correlation Matrix

Component	1	2	3	4
1	1.000	.194	-.020	.350
2	.194	1.000	-.032	.113
3	-.020	-.032	1.000	.032
4	.350	.113	.032	1.000

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

Third: Factor Analysis: all item (38)

Correlation Matrix

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
1	1.000	.020	-.028	-.065	-.047	.040	-.066	.037	.000	.072	.032	-.042	.109	.092	.080	-.021	.037	.022	.016	.101	.076	-.057	.055	-.033	.129	-.022	.084	.023	-.002	-.159	.012	-.023	.086	-.153	.016	-.002	-.010	.004
2	.020	1.000	.076	.031	.124	.123	-.061	.119	.107	.025	-.034	.053	.065	.188	.013	.284	.035	.028	-.009	.011	-.090	-.066	.104	.151	.051	.028	.108	.148	.142	.260	.144	.034	.227	.135	.272	.087	.009	.209
3	-.028	.076	1.000	.297	.069	.070	.029	.056	.114	.101	.147	.082	-.015	.062	.024	.106	.128	.254	.046	.083	.137	.120	.142	.066	.087	.037	.106	.127	-.014	.192	.174	.175	.174	.087	.207	.160	.279	.263
4	-.065	.031	.297	1.000	.022	.022	.100	-.002	-.139	-.079	.061	.036	.039	.036	.019	-.045	.034	.126	.157	.022	.074	-.037	.015	.102	-.004	-.037	.097	.020	-.014	.036	-.039	.091	.048	.109	.210	-.027	.103	.048
5	-.047	.124	.069	.022	1.000	.028	.142	.005	.069	.000	-.087	-.092	.030	-.036	.004	.036	-.087	.151	.020	.051	-.068	.022	.096	.132	.177	.039	.131	.081	.139	.193	-.067	-.051	.178	.162	.057	.068	.024	-.042
6	.040	.123	.070	.022	.028	1.000	.125	.135	.075	-.113	.290	.135	.155	.113	.047	.241	.233	.038	-.083	-.097	-.019	.147	.139	.131	.049	.035	.103	.197	.312	.305	.061	.164	.274	.137	.238	.220	.194	.240
7	-.066	-.061	.029	.100	.142	.125	1.000	.136	-.103	-.127	.028	-.035	.019	.039	.083	.037	-.057	.200	.113	-.158	.005	.036	-.056	-.058	-.104	-.122	.079	-.014	-.052	.131	-.092	.063	-.018	-.068	.046	.002	.094	.050
8	.037	.119	.056	-.002	.005	.135	.136	1.000	.101	.085	.130	.055	.061	.021	.077	.174	.042	.154	.139	.051	.001	.002	.212	.165	-.018	.185	.052	.157	.187	.083	.096	.002	.231	.049	.094	.138	.002	.175
9	.000	.107	.114	-.139	.069	.075	-.103	.101	1.000	.153	.110	.194	.162	.090	.107	.181	.036	.090	.071	.106	-.008	.046	.052	.173	.005	.012	.013	.143	.216	.139	.147	.026	.168	.146	.040	-.022	.103	.089
10	.072	.025	.101	-.079	.000	-.113	-.127	.085	.153	1.000	-.208	-.037	.166	.082	-.010	.135	.140	-.021	.010	.202	.121	-.101	.052	-.052	.162	-.073	-.127	.032	-.025	-.058	-.021	-.020	.061	-.053	-.013	-.098	-.062	.091
11	.032	-.034	.147	.061	-.087	.290	.028	.130	.110	-.208	1.000	.212	.110	.033	.249	.153	.118	.188	.172	.110	.145	-.028	.043	.120	.036	-.070	.111	.092	.161	.112	.083	.219	.181	.074	.184	.210	.138	.016
12	-.042	.053	.082	.036	-.092	.135	-.035	.055	.194	-.037	.212	1.000	.344	.186	.118	.122	.027	.094	.129	.157	.208	.136	.154	.104	.146	.115	.089	.116	.149	.168	.105	.164	.229	.120	.145	.043	.232	.178
13	.109	.065	-.015	.039	.030	.155	.019	.061	.162	.166	.110	.344	1.000	.319	.305	.111	-.027	.244	.166	.116	.114	-.019	.125	.113	.048	.109	-.061	.020	.054	.081	.114	.144	.191	.081	.101	.061	.032	.205
14	.092	.188	.062	.036	-.036	.113	.039	.021	.090	.082	.033	.186	.319	1.000	.134	.245	.042	.159	.099	-.005	.152	-.133	.166	.229	-.099	.162	-.043	.047	.214	.138	.131	.074	.253	.138	.215	.042	.160	.220
15	.080	.013	.024	.019	.004	.047	.083	.077	.107	-.010	.249	.118	.305	.134	1.000	.016	.086	.225	.151	.092	.133	-.018	.013	.264	.070	.071	.208	-.021	.128	.133	.088	.125	.144	.080	.181	.185	.054	.099
16	-.021	.284	.106	-.045	.036	.241	.037	.174	.181	.135	.153	.122	.111	.245	.016	1.000	.083	.153	.175	.058	-.001	.063	.162	.111	-.020	.182	.103	.143	.223	.259	.022	.167	.325	.174	.273	.085	.205	.251
17	.037	.035	.128	.034	-.087	.233	-.057	.042	.036	.140	.118	.027	-.027	.042	.086	.083	1.000	-.091	.032	-.041	.032	-.020	.130	.052	.032	.058	.201	.015	.088	.099	.123	.136	.170	.108	.123	.025	.049	.082
18	.022	.028	.254	.126	.151	.038	.200	.154	.090	-.021	.188	.094	.244	.159	.225	.153	-.091	1.000	.218	.062	.007	.150	.128	.088	.030	.129	-.005	-.023	.118	.097	.011	.056	.083	.095	.109	.074	-.028	.043
19	.016	-.009	.046	.157	.020	-.083	.113	.139	.071	.010	.172	.129	.166	.099	.151	.175	.032	.218	1.000	.044	.029	.021	.146	.184	.010	.142	.088	.111	.160	.085	.069	.151	.168	.179	.210	.179	-.038	.080
20	.101	.011	.083	.022	.051	-.097	-.158	.051	.106	.202	.110	.157	.116	-.005	.092	.058	-.041	.062	.044	1.000	.210	-.102	.129	.109	.138	.011	.018	.088	.051	-.080	.037	.008	.033	.030	.101	-.001	-.054	-.061
21	.076	-.090	.137	.074	-.068	-.019	.005	.001	-.008	.121	.145	.208	.114	.152	.133	-.001	.032	.007	.029	.210	1.000	-.079	.067	.183	.066	.008	.047	.006	.156	-.011	.103	.087	.123	.019	.078	.033	.083	.161

2 2	-.057	-.066	.120	-.037	.022	.147	.036	.002	.046	-.101	-.028	.136	-.019	-.133	-.018	.063	-.020	.150	.021	-.102	-.079	1.000	.239	-.086	.119	.087	-.102	-.069	.031	.005	-.028	.121	.059	-.075	-.028	-.054	-.023	.009
2 3	.055	.104	.142	.015	.096	.139	-.056	.212	.052	.052	.043	.154	.125	.166	.013	.162	.130	.128	.146	.129	.067	.239	1.000	.041	.086	.046	-.041	-.135	.118	.272	.119	.050	.175	.053	.230	.192	.094	.242
2 4	-.033	.151	.066	.102	.132	.131	-.058	.165	.173	-.052	.120	.104	.113	.229	.264	.111	.052	.088	.184	.109	.183	-.086	.041	1.000	.111	.242	.230	.196	.289	.115	.148	.194	.170	.344	.185	.146	.200	.141
25	.129	.051	.087	-.004	.177	.049	-.104	-.018	.005	.162	.036	.146	.048	-.099	.070	-.020	.032	.030	.010	.138	.066	.119	.086	.111	1.000	.173	.001	.124	.049	.052	.053	.129	.075	.036	-.044	.072	-.038	.022
26	-.022	.028	.037	-.037	.039	.035	-.122	.185	.012	-.073	-.070	.115	.109	.162	.071	.182	.058	.129	.142	.011	.008	.087	.046	.242	.173	1.000	.150	.181	.242	.053	.140	.250	.205	.108	.064	.076	.038	.178
27	.084	.108	.106	.097	.131	.103	.079	.052	.013	-.127	.111	.089	-.061	-.043	.208	.103	.201	-.005	.088	.018	.047	-.102	-.041	.230	.001	.150	1.000	.254	.182	.323	.231	.164	.312	.174	.335	.188	.214	.307
28	.023	.148	.127	.020	.081	.197	-.014	.157	.143	.032	.092	.116	.020	.047	-.021	.143	.015	-.023	.111	.088	.006	-.069	-.135	.196	.124	.181	.254	1.000	.310	.256	.239	.191	.183	.074	.257	.195	.140	.123
29	-.002	.142	-.014	-.014	.139	.312	-.052	.187	.216	-.025	.161	.149	.054	.214	.128	.223	.088	.118	.160	.051	.156	.031	.118	.289	.049	.242	.182	.310	1.000	.299	.222	.148	.352	.138	.223	.307	.300	.275
30	-.159	.260	.192	.036	.193	.305	.131	.083	.139	-.058	.112	.168	.081	.138	.133	.259	.099	.097	.085	-.080	-.011	.005	.272	.115	.052	.053	.323	.256	.299	1.000	.309	.273	.395	.150	.395	.292	.293	.328
31	.012	.144	.174	-.039	-.067	.061	-.092	.096	.147	-.021	.083	.105	.114	.131	.088	.022	.123	.011	.069	.037	.103	-.028	.119	.148	.053	.140	.231	.239	.222	.309	1.000	.139	.316	.147	.257	.210	.095	.193
32	-.023	.034	.175	.091	-.051	.164	.063	.002	.026	-.020	.219	.164	.144	.074	.125	.167	.136	.056	.151	.008	.087	.121	.050	.194	.129	.250	.164	.191	.148	.273	.139	1.000	.218	.135	.201	.058	.201	.271
33	.086	.227	.174	.048	.178	.274	-.018	.231	.168	.061	.181	.229	.191	.253	.144	.325	.170	.083	.168	.033	.123	.059	.175	.170	.075	.205	.312	.183	.352	.395	.316	.218	1.000	.179	.366	.197	.303	.341
34	-.153	.135	.087	.109	.162	.137	-.068	.049	.146	-.053	.074	.120	.081	.138	.080	.174	.108	.095	.179	.030	.019	-.075	.053	.344	.036	.108	.174	.074	.138	.150	.147	.135	.179	1.000	.246	-.008	.095	.108
35	.016	.272	.207	.210	.057	.238	.046	.094	.040	-.013	.184	.145	.101	.215	.181	.273	.123	.109	.210	.101	.078	-.028	.230	.185	-.044	.064	.335	.257	.223	.395	.257	.201	.366	.246	1.000	.276	.171	.215
36	-.002	.087	.160	-.027	.068	.220	.002	.138	-.022	-.098	.210	.043	.061	.042	.185	.085	.025	.074	.179	-.001	.033	-.054	.192	.146	.072	.076	.188	.195	.307	.292	.210	.058	.197	-.008	.276	1.000	.205	.364
37	-.010	.009	.279	.103	.024	.194	.094	.002	.103	-.062	.138	.232	.032	.160	.054	.205	.049	-.028	-.038	-.054	.083	-.023	.094	.200	-.038	.038	.214	.140	.300	.293	.095	.201	.303	.095	.171	.205	1.000	.347
38	.004	.209	.263	.048	-.042	.240	.050	.175	.089	.091	.016	.178	.205	.220	.099	.251	.082	.043	.080	-.061	.161	.009	.242	.141	.022	.178	.307	.123	.275	.328	.193	.271	.341	.108	.215	.364	.347	1.000

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.648
Bartlett's Test of Sphericity	Approx. Chi-Square	1142.918
	df	703
	Sig.	.000

Communalities

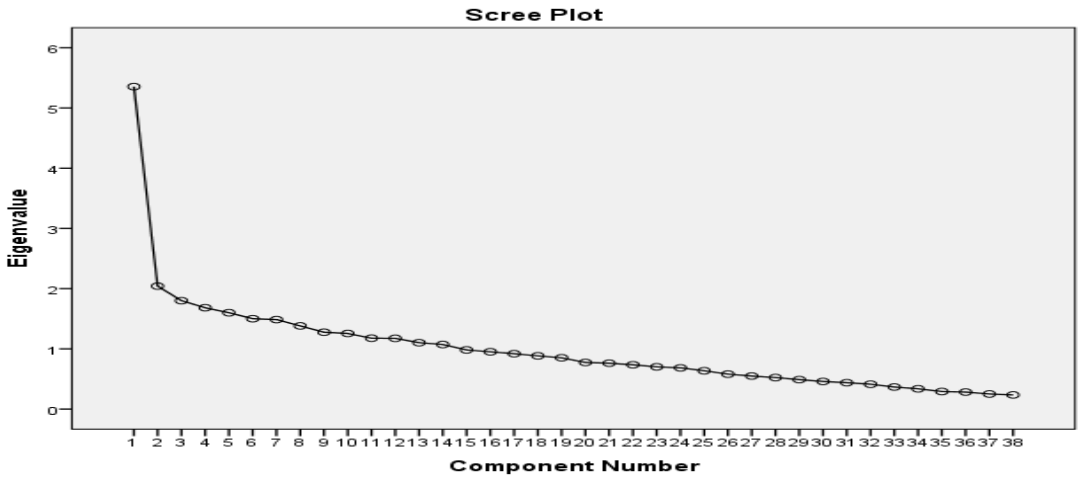
	Initial	Extraction
Choose3 Q.1 scores	1.000	.561
Choose3 Q.2 scores	1.000	.555
Choose3 Q.3 scores	1.000	.660
Choose3 Q.4 scores	1.000	.621
Choose3 Q.5 scores	1.000	.753
Choose3 Q.6 scores	1.000	.684
Choose3 Q.7 scores	1.000	.638
Choose3 Q.8 scores	1.000	.587
Choose3 Q.9 scores	1.000	.675
Choose3 Q.10 scores	1.000	.730
Choose3 Q.11 scores	1.000	.719
Choose3 Q.12 scores	1.000	.570
Choose3 Q.13 scores	1.000	.664
Choose3 Q.14 scores	1.000	.644
Choose3 Q.15 scores	1.000	.644
Choose3 Q.16 scores	1.000	.566
Choose3 Q.17 scores	1.000	.735
Choose3 Q.18 scores	1.000	.582
Choose3 Q.19 scores	1.000	.520
Choose3 Q.20 scores	1.000	.579
Choose3 Q.21 scores	1.000	.539
Choose3 Q.22 scores	1.000	.642
Choose3 Q.23 scores	1.000	.748
Choose3 Q.24 scores	1.000	.637
Ranking Q.1 scores	1.000	.655
Ranking Q.2 scores	1.000	.714
Ranking Q.3 scores	1.000	.590
Ranking Q.4 scores	1.000	.697
Ranking Q.5 scores	1.000	.594
Ranking Q.6 scores	1.000	.650
Ranking Q.7 scores	1.000	.575
Ranking Q.8 scores	1.000	.554
Ranking Q.9 scores	1.000	.488
Ranking Q.10 scores	1.000	.629
Ranking Q.11 scores	1.000	.627
Ranking Q.12 scores	1.000	.617
Ranking Q.13 scores	1.000	.632
Ranking Q.14 scores	1.000	.633

Extraction Method: Principal Component Analysis.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.355	14.092	14.092	5.355	14.092	14.092
2	2.042	5.373	19.465	2.042	5.373	19.465
3	1.800	4.736	24.201	1.800	4.736	24.201
4	1.683	4.428	28.629	1.683	4.428	28.629
5	1.600	4.210	32.839	1.600	4.210	32.839
6	1.501	3.949	36.788	1.501	3.949	36.788
7	1.486	3.912	40.700	1.486	3.912	40.700
8	1.381	3.634	44.334	1.381	3.634	44.334
9	1.276	3.359	47.692	1.276	3.359	47.692
10	1.256	3.305	50.997	1.256	3.305	50.997
11	1.177	3.098	54.095	1.177	3.098	54.095
12	1.175	3.092	57.188	1.175	3.092	57.188
13	1.103	2.903	60.090	1.103	2.903	60.090
14	1.073	2.823	62.914	1.073	2.823	62.914
15	.983	2.586	65.500			
16	.951	2.504	68.003			
17	.919	2.419	70.423			
18	.884	2.326	72.748			
19	.851	2.241	74.989			
20	.774	2.037	77.026			
21	.761	2.004	79.030			
22	.736	1.936	80.966			
23	.702	1.846	82.812			
24	.685	1.801	84.613			
25	.636	1.674	86.287			
26	.580	1.527	87.815			
27	.549	1.444	89.259			
28	.523	1.377	90.635			
29	.489	1.287	91.922			
30	.459	1.207	93.130			
31	.438	1.152	94.282			
32	.412	1.085	95.367			
33	.366	.963	96.330			
34	.337	.886	97.216			
35	.291	.767	97.983			
36	.282	.742	98.725			
37	.250	.658	99.383			
38	.234	.617	100.000			

Extraction Method: Principal Component Analysis.



	Component													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Ranking Q.9	.661													
Ranking Q.6	.609	-.354												
Ranking Q.11	.593													
Ranking Q.14	.569								-.372					
Ranking Q.5	.568						-.303							
Choose3 Q.16	.473			.357								.309		
Choose3 Q.24	.461			-.355	.315								.355	
Choose3 Q.6	.450													
Ranking Q.13	.449				-.307					-.405				
Ranking Q.3	.448	-.343		-.381										
Ranking Q.8	.429							-.320						
Ranking Q.7	.428												-.354	
Ranking Q.4	.405		-.308									.378	-.300	
Choose3 Q.12	.394	.312						-.303						
Choose3 Q.2	.338													
Choose3 Q.15	.332			-.313										.327
Choose3 Q.19	.321									.316				
Choose3 Q.13	.333	.530												
Choose3 Q.20		.501												
Choose3 Q.10		.444	-.366				.380					.313		
Choose3 Q.21		.350		-.317	-.333									
Choose3 Q.7			.546											
Choose3 Q.18			.533		.303									
Choose3 Q.23	.338			.472								-.378		
Choose3 Q.5					.554					-.375				
Ranking Q.1						.632					.317			
Choose3 Q.14	.388					-.530								
Choose3 Q.22				.474		.502								
Choose3 Q.3	.357						.554							
Choose3 Q.4			.383				.507							
Choose3 Q.1								.451			.367			
Ranking Q.12	.443							.450						
Ranking Q.10	.364				.365			-.399						
Ranking Q.2	.325								-.523					
Choose3 Q.11	.356		.343						.461					
Choose3 Q.9									.441					.371
Choose3 Q.17										.546	.303			.306
Choose3 Q.8	.309							.327				.373		

Extraction Method: Principal Component Analysis.

a. 14 components extracted.

Pattern Matrix[a]

a. Rotation failed to converge in 25 iterations. (Convergence = .001).

Appendix 12

SJTs' Spearman's rho correlation with the other measures

		SJT Total scores	Neuroticism	Extraversion	Openness	Agreeableness	Conscientiousness	Interview scores	GPA
SJTs Total scores	Pearson Correlation	1.000	-0.162	0.115	0.085	.218*	.250**	-.259*	.367**
	Sig. (2-tailed)		0.100	0.236	0.384	0.023	0.008	0.014	0.000
	N	122	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).

SJTs' Spearman's rho correlation with other measures by gender

		Neuroticism		Extraversion		Openness		Agreeableness		Consciousness		Interview scores		GPA	
		Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
JT Total scores	Pearson Correlation	-.252*	0.008	0.087	0.225	-0.005	0.309	0.185	0.360	.232*	0.231	-0.214	0.159	.237*	-0.137
	Sig. (2-tailed)	0.020	0.973	0.413	0.354	0.962	0.199	0.082	0.119	0.030	0.290	0.080	0.492	0.019	0.522
	N	85	19	90	19	89	19	89	20	88	23	68	21	97	24

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Appendix 13

Non-cognitive domains for teacher's effectiveness in Oman

Communication skills – Candidate is humane in relation to others and demonstrates active listening. Candidate is responsive to pupils' needs, and able to adapt style of communication to suit recipients.

Organisation and planning – Candidate has the ability to manage competing priorities and display time management skills effectively. Demonstrates good organisation and planning skills.

Resilience and adaptability – Candidate shows the capability to remain resilient under stress and challenges to own knowledge. Demonstrates adaptability and the confidence to make decisions independently, and seeks help when necessary.

Professional ethics – Candidate shows high consideration to the Islamic, Omani, and professional ethics. Demonstrates respect for pupils and colleagues, and treats others fairly. Accepts responsibility and is trustworthy.

Enthusiasm and motivation – Candidate is aware of national and job loyalty, and shows strong and reliable commitment to being a teacher. Takes pleasure in teaching tasks, and seeks professional development.

References

Affleck, P., Bowman, M., Wardman, M., Sinclair, S., & Adams, R. (2016). Can we improve on situational judgement tests?. *British dental journal*, *220*(1), 9.

Ahmed, H., Rhydderch, M., & Matthews, P. (2012). Can knowledge tests and situational judgement tests predict selection centre performance. *Medical education*, *46*(8), 777-784.

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

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 Ganbousy, S., Al Harthi, H., & Kazem, A. (2012). An Evaluation of Teacher Preparation Program in the College of Education, Sultan Qaboos University from the Perspectives of its Graduates. *Journal of Educational and Psychological Studies [JEPS], 6(1),* 99-113.

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 Jarida Al Rasmiyya. (2012- No. 984) Muscat Retrieved from http://www.ilo.org/dyn/natlex/natlex4.detail?p_lang=en&p_isn=94752.

Al Mahree, S. (2006). *The preparation of students in the colleges of education in Oman.*Unpublished thesis, Alazhar University, Egypt (Arabic).

Al Numani, S. (2006). *Standard for recruiting new teachers at MoE in Oman*. Unpublished Master thesis, Sultan Qaboos University. Muscat. Oman (Arabic)

Al Tobi, K. S. (2006). *The effectiveness of teacher recruitment and selection in Oman: an analysis of stakeholder perceptions.* Doctoral disertation, University of Leicester. UK.

Al-Ani, W. T. (2014). Core values matrix of the philosophy of basic education in Oman (PBEO). *Athens Journal of Education*, *1*(2), 167-181.

Al-Ani, W. T. (2017). Alternative education needs in Oman: accommodating learning diversity and meeting market demand. *International Journal of Adolescence and Youth*, *22*(3), 322-336.

Al-Ani, W. T., & Ismail, O. H. (2015). Can mission predict school performance? The case of basic education in Oman. *School Leadership & Management*, *35*(5), 459-476.

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-Ani, W., Al-Sulaimani, H., Al-Aharthi, A., Al-Munthiri, R., Al-Seyabi, F. (2018). Decision-Making in Light of Teaching Code of Ethics: Perceptions of Omani School Educators. *Journal of Educational and Psychological Studies [JEPS], 12*(2), 321-339.

Albelushi, A. (2004). Gender issues in teacher development: Career choice and commitment in Oman. *Engl Lang Teac Educat Develop*, *8*, 1-25.

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

Alhashmi, A., & Eissa, A. A. (2010). Perceptions of Arabic teachers of the educational component of the Arabic Education program in the College of Education, Sultan Qaboos University: An evaluative study. *Journal of Educational and Psychological Studies [JEPS], 4(1)*, 67-98.

Alkhausi, H. A., Al-Yahmadi, H., Al-Kalbani, M., Clayton, D., Al-Barwani, T., Al-Sulaimani, H., ... & Khan, M. A. (2015). Predictors of First-Year Sultan Qaboos University Students' Grade Point Average. *Journal of Education and Practice*, *6*(12), 121-127.

Al-Rawahi, N. Y. & Al Balushi, S. (2011). The effect of the teacher preparation program in the College of Education on student teachers' acquisition of professional competencies in relation to their attitudes towards the teaching profession. *Journal of Educational and Psychological Studies [JEPS]*, *5*(2), 54-75.

Al-Rawahi, N. Y. (2010). An Investigation into Omani Physical Education Teachers' Beliefs about Physical Education Teachers' Characteristics and their Roles in Omani Schools. *Journal of Educational and Psychological Studies [JEPS]*, *4*(2), 1-13.

AlRoya Newspaper, Issue No. 1929 (2016, August 3) Press release on Teachers. Retrieved August 30, 2017 from https://alroya.om/journal/36 (Arabic)

Alyahmadi, H., & Al-Kiyumi, A. (2014). The consequences of teacher evaluation on teacher professional development in oman. *International Journal of Education and Research*, *2*(4), 127-142.

Amzat, I. H., & Al-Neimi, R. K. R. (2014). Teachers' turnover and their job satisfaction at basic education school in some regions in Oman: Structural equation modelling approach. *International Journal of Management in Education*, *8*(1), 78-100.

Anastasi, A. & Urbin, S. (1997). *Psychological testing* (7th ed). Upper Saddle River, NJ, US: Prentice Hall/Pearson Education.

Anderson, A. & Herriot, P. (Ed). (1997) *International Handbook for Selection and Assessment*. England. John Wiley & Sons

Angrist, J. D., & Guryan, J. (2008). Does teacher testing raise teacher quality? Evidence from state certification requirements. *Economics of Education Review*, *27*(5), 483-503

Arthur Jr, W., Glaze, R. M., Jarrett, S. M., White, C. D., Schurig, I., & Taylor, J. E. (2014). Comparative evaluation of three situational judgment test response formats in terms of construct-related validity, subgroup differences, and susceptibility to response distortion. *Journal of Applied Psychology*, *99*(3), 535.

Arthur, W., Day, E. A., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology*, *56*(1), 125-153.

Atteberry, A., Loeb, S., & Wyckoff, J. (2013). *Do first impressions matter? Improvement in early career teacher effectiveness*. (No. 19096) National Burea of Economic Research Inc.

Barber, M., & Mourshed, M. (2007). *How the world's best-performing schools systems come out on top*: London: McKinsey & Company.

Barr, J. J. (2011). The relationship between teachers' empathy and perceptions of school culture. *Educational Studies*, *37*(3), 365-369.

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

Bauer, T. N., Truxillo, D. M., Sanchez, R. J., Craig, J. M., Ferrara, P., & Campion, M. A. (2001). Applicant reactions to selection: Development of the selection procedural justice scale (SPJS). *Personnel psychology*, *54*(2), 387-419.

Beauchamp, G., Clarke, L., Hulme, M., & Murray, J. (2015). Teacher education in the United Kingdom post devolution: Convergences and divergences. *Oxford Review of Education*, *41*(2), 154-170.

Becker, T. E. (2005). Development and validation of a situational judgment test of employee integrity. *International Journal of Selection and Assessment*, *13*(3), 225-232.

Belwal, R. Priyadarshi, P., & Al Fazari, M. H. (2017) Graduate attributes and employability skills: Graduates' perspectives on employers' expectations in Oman. *International Journal of Educational Management*, *31*(6), 814-827.

Bergman, M. E., Drasgow, F., Donovan, M. A., Henning, J. B., & Juraska, S. E. (2006). Scoring situational judgment tests: Once you get the data, your troubles begin. *International Journal of Selection and Assessment*, *14*(3), 223-235

Bieri, C., & Schuler, P. (2011). Cross-curricular competencies of student teachers: a selection model based on assessment centre admission tests and study success after the first year of teacher training. *Assessment & Evaluation in Higher Education*, *36*(4), 399-415

Biermann, A., Karbach, J., Spinath, F. M., & Brünken, R. (2015). Investigating effects of the quality of field experiences and personality on perceived teaching skills in German pre-service teachers for secondary schools. *Teaching and Teacher Education, 51,* 77-87

Bleidorn, W., Klimstra, T. A., Denissen, J. J., Rentfrow, P. J., Potter, J., & Gosling, S. D. (2013). Personality maturation around the world: A cross-cultural examination of social-investment theory. *Psychological Science*, *24*(12), 2530-2540.

Borghans, L., Duckworth, A. L., Heckman, J. J., & Ter Weel, B. (2008). The economics and psychology of personality traits. *Journal of Human Resources*, *43*(4), 972-1059.

Bowles, T., Hattie, J., Dinham, S., Scull, J., & Clinton, J. (2014). Proposing a comprehensive model for identifying teaching candidates. *The Australian Educational Researcher*, *41*(4), 365-380

Boyd, D. J., Grossman, P. L., Lankford, H., Loeb, S., & Wyckoff, J. (2008). *Who leaves? Teacher attrition and student achievement*. Washington, DC: National Bureau of Economic Research.

Brunello, G., & Schlotter, M. (2011). *Non-cognitive skills and personality traits: Labour market relevance and their development in education & training systems*. IZA Discussion Paper No. 5743.

Bryman, A. (2015). *Social research methods*: Oxford university press.

Caena, F. (2014). Teacher Competence Frameworks in Europe: policy-as-discourse and policy-as-practice. *European Journal of Education, 49*(3), 311-331.

Callwood, A., Jeevaratnam, K., Kotronoulas, G., Schneider, A., Lewis, L., & Nadarajah, V. D. (2018). Personal domains assessed in multiple mini interviews (MMIs) for healthcare student selection: A narrative synthesis systematic review. *Nurse education today, 64*, 56-64.

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

Campion, M. C., Ployhart, R. E., & MacKenzie, W. I. (2014). The State of Research on Situational Judgment Tests: A Content Analysis and Directions for Future Research. *Human Performance, 27*(4), 283-310.

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), 1-24.

Casey, C., & Childs, R. (2011). Teacher education admission criteria as measure of preparedness for teaching. *Canadian Journal of Education, 34*(2), 3-20.

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.

Casner-Lotto, J., & Barrington, L. (2006). *Are They Really Ready to Work? Employers' Perspectives on the Basic Knowledge and Applied Skills of New Entrants to the 21st Century US Workforce*. Partnership for 21st Century Skills. Massachusetts Avenue NW Suite 700, Washington, DC Retrived from https://files.eric.ed.gov/fulltext/ED519465.pdf

Catano, V. M., Brochu, A., & Lamerson, C. D. (2012). Assessing the reliability of situational judgment tests used in high-stakes situations. *International Journal of Selection and Assessment*, *20*(3), 333-346

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.

Chan, D., & Schmitt, N. (2004). An agenda for future research on applicant reactions to selection procedures: A construct-oriented approach. *International Journal of Selection and Assessment*, *12*(1-2), 9-23

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.

Chesnut, S. R., & Burley, H. (2015). Self-efficacy as a predictor of commitment to the teaching profession: A meta-analysis. *Educational Research Review, 15*, 1-16.

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

Coe, R., Aloisi, C., Higgins, S., & Major, L. E. (2014). *What makes great teaching? Review of the underpinning research.* Project Report. Sutton Trust, London. Retrieved April 12, 2017, from http://dro.dur.ac.uk/13747/1/13747.pdf.

Coe, R.J. (2012) 'The nature of educational research' in Arthur, J., Michael W., Coe R. & Hedges, L. V. (eds). *Research methods and methodologies in education*: Sage publications

Cohen, L., Manion, L., & Morrison, K. (2013). *Research methods in education*. Routledge.

Common, R. K. (2011). Barriers to developing 'leadership'in the Sultanate of Oman. *International Journal of Leadership Studies*, *6*(2), 215-228.

Conway, P. F., Murphy, R., Rath, A., & Hall, K. (2009). *Learning to teach and its implications for the continuum of teacher education: A nine-country cross-national study. Report commissioned by the Teaching Council*. (University College Cork and Teaching Council of Ireland).

Corno, L. Y. N. (2008). On teaching adaptively. *Educational Psychologist*, *43*(3), 161-173.

Corstjens, J., Lievens, F., & Krumm, S. (in press). Contextualized and generic Situational Judgment Tests for selection. In H. Goldstein, E. Pulakos, J. Passmore, & C. Semedo (Eds.) *Handbook of the psychology of recruitment, selection, and retention*. Wiley-Blackwell.

Costa Jr, P. T., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: robust and surprising findings. *Journal of personality and social psychology*, *81*(2), 322

Creemers, B. P., & Kyriakides, L. (2006). Critical analysis of the current approaches to modelling educational effectiveness: The importance of establishing a dynamic model. *School Effectiveness and School Improvement*, *17*(3), 347-366

Creswell, J. W., & Poth, C. N. (2018). *Qualitative Inquiry and Research design: choosing among five approaches*. (4th ed). SAGE

Cronbach, L. J. (1990) *Essentials of psychological testing* (5th ed). New York, happer and Row publishers.

Crook, A. E., Beier, M. E., Cox, C. B., Kell, H. J., Hanks, A. R., & Motowidlo, S. J. (2011). Measuring relationships between personality, knowledge, and performance using single-response situational judgment tests. *International Journal of Selection and Assessment*, *19*(4), 363-373

Cullen, M. J., Sackett, P. R., & Lievens, F. (2006). Threats to the operational use of situational judgment tests in the college admission process. *International Journal of Selection and Assessment*, *14*(2), 142-155

Darling-Hammond, L. (2000). How teacher education matters. *Journal of teacher education*, *51*(3), 166-173.

Darmody, M., & Smyth, E. (2016). *Entry to programmes of initial teacher education.* Dublin: Economic and Social Research Institute and the Teaching Council.

Day, C., Sammons, P., & Stobart, G. (2007). *Teachers matter: Connecting work, lives and effectiveness*. McGraw-Hill Education (UK).

De Wolff, C. J. (1989). The changing role of psychologists in selection in Herriot, P. (Ed). *Assessment and selection in organisations: methods and practices for recruitment and appraisal*. John Wiley & SONS ltd. England

Dennis M. McInerney, Fraide A. Ganotice, Ronnel B. King, Alexandre J.S. Morin & Herbert W. Marsh (2015) Teachers' Commitment and psychological well-being: implications of self-beliefs for teaching in Hong Kong, *Educational Psychology*, 35:8, 926-945.

Department of Education (the University of York) (2018), Students Research Ethics, Available online at https://www.york.ac.uk/education/intranet/student-research-ethics/

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.

Donnellan, M. B., & Lucas, R. E. (2008). Age differences in the Big Five across the life span: evidence from two national samples. *Psychology and aging*, *23*(3), 558

Douglas, E. F., McDaniel, M. A., & Snell, A. F. (1996, August). The validity of non-cognitive measures decays when applicants fake. *Academy of Management Proceedings*, *(1)*, pp. 127-131.

Doyle, W. (1977). 4: Paradigms for research on teacher effectiveness. *Review of research in education*, *5*(1), 163-198

Duckworth, A. L. (2009). (Over and) beyond high-stakes testing. *American Psychologist*, *64*, 279–280

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.

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: perseverance and passion for long-term goals. *Journal of personality and social psychology*, *92*(6), 1087.

Durksen, T. L., & Klassen, R. M. (2017). The development of a situational judgement test of personal attributes for quality teaching in rural and remote Australia. *The Australian Educational Researcher*, 1-22.

Eignor, D. R. (2001). Standards for the development and use of tests: The Standards for Educational and Psychological Testing. *European Journal of Psychological Assessment*, *17*(3), 157

El-Baz, H. S., & El-Sayegh, S. M. (2015). Competency Domain Model and the Perception of Engineering Managers in the United Arab Emirates. *Engineering Management Journal, 22*(1), 3-12.

Eraut, M. (2000). Design of initial teacher education. *International Journal of Educational Research, 33*(5), 453-456.

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.

Feingold, A. (1994). Gender differences in personality: a meta-analysis. *Psychological bulletin, 116*(3), 429.

Flores, M. A., & Niklasson, L. (2014). Why do student teachers enrol for a teaching degree? A study of teacher recruitment in Portugal and Sweden. *Journal of Education for Teaching, 40*(4), 328-343.

Franchi, L. (2016). Initial teacher education in the university 'My little ship, how ill-laden you are'. *European Journal of Teacher Education, 39*(2), 147-158.

Franzoci, R. (2004). Content analysis. In Hardy, M., & Bryman, A. *Handbook of data analysis* (pp. 548-565). SAGE Publications Ltd.

Gao, M., & Liu, Q. (2013). Personality traits of effective teachers represented in the narratives of American and Chinese preservice teachers: A cross-cultural comparison. *International Journal of Humanities and Social Science, 3(2),* 84-95

Garcia, E. (2014). The Need to Address Non-Cognitive Skills in the Education Policy Agenda1. *Economic policy institute. Briefing paper #386,* Retrieved at https://files.eric.ed.gov/fulltext/ED558126.pdf

Gibbons, A. M., & Rupp, D. E. (2009). Dimension consistency as an individual difference: A new (old) perspective on the assessment center construct validity debate. *Journal of Management, 35*(5), 1154-1180.

Gneema, M. 1996. *Policies and Programs for Arabic Teacher Training*. Aldar Almasryaa Allibnaneea, Cairo, Egypt (Arabic)

Goldhaber, D., Grout, C., & Huntington-Klein, N. (2014). Screen Twice, Cut Once: Assessing the Predictive Validity of Teacher Selection Tools. CEDR Working Paper 2014-9. University of Washington, Seattle, WA.

Gorard, S., & Taylor, C. (2004). *Combining methods in educational and social research*: McGraw-Hill Education (UK).

Gore, J., Barron, R. J., Holmes, K., & Smith, M. (2016). Who says we are not attracting the best and brightest? Teacher selection and the aspirations of Australian school students. *The Australian Educational Researcher, 43*(5), 527-549

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in personality, 37*(6), 504-528.

Grant, L. W., Stronge, J. H., & Xu, X. (2013). A cross-cultural comparative study of teacher effectiveness: Analyses of award-winning teachers in the United States and China. *Educational Assessment, Evaluation and Accountability, 25*(3), 251-276

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

Guenole, N., Chernyshenko, O. S., & Weekly, J. (2017). On Designing Construct Driven Situational Judgment Tests: Some Preliminary Recommendations. *International Journal of Testing*, 1-19

Guo, H., Zu, J., Kyllonen, P., & Schmitt, N. (2016). Evaluation of Different Scoring Rules for a Noncognitive Test in Development. *ETS Research Report Series*, *2016*(1), 1-13.

Gutman, L. M., & Schoon, I. (2013). The impact of non-cognitive skills on outcomes for young people. *Education Endowment Foundation*. Available at: h ttp://educationendowmentfoundation. org. uk/uploads/pdf/Non-cognitive_skills_ literature_review. pdf.

Hamre, B. K., Pianta, R. C., Downer, J. T., DeCoster, J., Mashburn, A. J., Jones, S. M., . . . Brackett, M. A. (2013). Teaching through interactions: Testing a developmental framework of teacher effectiveness in over 4,000 classrooms. *Elementary School Journal, 113*, 461–487.

Harford, J. (2010). Teacher education policy in Ireland and the challenges of the twenty-first century. *European Journal of Teacher Education, 33*(4), 349-360.

Harris, D. N., & Sass, T. R. (2014). Skills, productivity and the evaluation of teacher performance. *Economics of Education Review, 40*, 183-204.

Hashim, H., Damio, S. M., & Hussin, N. (2013). Malaysian Educators Selection Inventory (MEdSI) Certified Interviewers' Perspectives on its Implementation: A Preliminary Study. *Procedia-Social and Behavioral Sciences, 90*, 171-177.

Hassan, A. S., & Alzubaidi, A. S. (2009) Grade Distribution at the Faculty of Education at Sultan Qaboos. *University Journal of Educational & Psychological Science, 10 (1),* 12-33

Hattie, J. (2009). *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. London: Routledge

Heckman, J. J. (2000). Policies to foster human capital. *Research in economics, 54*(1), 3-56.

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

Heinz, M. (2013). Tomorrow's teachers—selecting the best: An exploration of the quality rationale behind academic and experiential selection criteria for initial teacher education programmes. *Educational Assessment, Evaluation and Accountability*, *25*(2), 93-114

Hemmi, K., & Ryve, A. (2015). Effective mathematics teaching in Finnish and Swedish teacher education discourses. *Journal of Mathematics Teacher Education, 18*(6), 501-521

Henry, G. T., Bastian, K. C., Fortner, C. K., Kershaw, D. C., Purtell, K. M., Thompson, C. L., & Zulli, R. A. (2014). Teacher preparation policies and their effects on student achievement. *Education*, *9*(3), 264-303.

Henry, G. T., Purtell, K. M., Bastian, K. C., Fortner, C. K., Thompson, C. L., Campbell, S. L., & Patterson, K. M. (2014). The effects of teacher entry portals on student achievement. *Journal of Teacher Education, 65*(1), 7-23.

Herriot, P. (Ed.1989). *Assessment and selection in organisations: methods and practices for recruitment and appraisal*. John Wiley & SONS ltd. England

Higher Education Admission Center (HEAC) (Oman). 2017. In http://www.heac.gov.om/

Hobson, A. J., Ashby, P., McIntyre, J., & Malderez, A. (2010). International approaches to teacher selection and recruitment. OECD Education Working Papers, No.47, OECD Publishing

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.

Hobson, A. J., Malderez, A., Tracey, L., Giannakaki, M., Pell, G., & Tomlinson, P. D. (2008). Student teachers' experiences of initial teacher preparation in England: Core themes and variation. *Research Papers in Education*, *23*(4), 407-433.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. Sage publications.

Hong, J. Y. (2012). Why do some beginning teachers leave the school, and others stay? Understanding teacher resilience through psychological lenses. *Teachers and Teaching*, *18*(4), 417-440.

House of Commons Education Committee. (2012). Great Teachers: Attracting, Training and Retaining the Best. Government's Response to the Committee's Ninth Report of Session 2010-12. London: House of Commons

Howard, A., Choi, M., & Krig, N. (2011). Similarities and differences in managerial judgement around the world: Developing an international SJT. *Assessment and Development Matters*, *3*(2), 2.:

Hughes, D. J. (2017). Psychometric validity: Establishing the accuracy and appropriateness of psychometric measures. In P., Irwing, T., Booth, DJ Hughes (Eds.), *Wiley handbook of psychometric testing: A multidisciplinary approach to survey, scale, and test development.* Hoboken, NJ: Wiley.

Husbands, A., Rodgerson, M. J., Dowell, J., & Patterson, F. (2015). Evaluating the validity of an integrity-based situational judgement test for medical school admissions. *BMC medical education*, *15*(1), 144.

Iles, P. (1999). *Managing staff selection and assessment*. Open University Press. Buckingham (UK).

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.

Islam, M. M., & Al-Ghassani, A. (2015). Predicting College Math Success: Do High School Performance and Gender Matter? Evidence from Sultan Qaboos University in Oman. *International Journal of Higher Education*, *4*(2), 67-80.

Issan, S. A., (2011) Akhlaqyaat Almihnah Altarpoyah Altalymyah lada Almullem fee Almadrassa (Professional ethics in schools), *Resalat Altarbeeah (32*). Muscat (Arabic)

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.

Jackson, D. J., LoPilato, A. C., Hughes, D., Guenole, N., & Shalfrooshan, A. (2016). The internal structure of situational judgement tests reflects candidate main effects: Not dimensions or situations. *Journal of Occupational and Organizational Psychology, 90*(1), 1-27

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

Jiang, H. (2016). Revisiting individualism and collectivism: a multinational examination of pre-service teachers' perceptions on student academic performances. *Intercultural Education, 27*(1), 101-110

Joharry, O., Lihanna, B., & Nabilah, A. (2008). The construction of the Malaysian educators selection inventory (MEdSI): a large scale assessment initiative. *Asian Journal of University Education, 4*(1), 113-126.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher, 33*(7), 14-26.

Jolijn Hendriks, A. A., Perugini, M., Angleitner, A., Ostendorf, F., Johnson, J. A., De Fruyt, F., ... & Conner, M. (2003). The five-factor personality inventory: cross-cultural generalizability across 13 countries. *European journal of personality*, *17*(5), 347-373

Joshi, H. (2014). *'Non-cognitive'skills: What are they and how can they be measured in the British cohort studies?. A Literature Review.* London: Institute of Education.

Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment. Research Paper. MET Project. Bill & Melinda Gates Foundation.

Kaplan, R., & Saccuzzo, D. (2009). *Psychological testing: Principles, applications, and issues*: Cengage Learning.

Kasten, N., & Freund, P. A. (2015). A meta-analytical multilevel reliability generalization of Situational Judgment Tests (SJTs). *European Journal of Psychological Assessment*

Kautz, T., Heckman, J. J., Diris, R., Ter Weel, B., & Borghans, L. (2014). *Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success*. (No. w20749). National Bureau of Economic Research.

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

Kell, H. J. (2018). Noncognitive proponents' conflation of "cognitive skills" and "cognition" and its implications. *Personality and Individual Differences, 134*, 25-32.

Kell, H. J., Rittmayer, A. D., Crook, A. E., & Motowidlo, S. J. (2010). Situational content moderates the association between the Big Five personality traits and behavioral effectiveness. *Human Performance, 23*(3), 213-228

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.

Kennedy, A. (2015). What do professional learning policies say about purposes of teacher education?. *Asia-Pacific journal of teacher education*, *43*(3), 183-194.

Kim, L. E., Dar-Nimrod, I., & MacCann, C. (2017). Teacher personality and teacher effectiveness in secondary school: Personality predicts teacher support and student self-efficacy but not academic achievement. *Journal of Educational Psychology*. 10(3). 309-323

Kim, Y., & Park, H. (2013). An Investigation of the Competencies Required of Airline Cabin Crew Members: The Case of a Korean Airline. *Journal of Human Resources in Hospitality & Tourism, 13*(1), 34-62.

Kington, A., Reed, N., & Sammons, P. (2014). Teachers' constructs of effective classroom practice: variations across career phases. *Research Papers in Education*, *29*(5), 534-556.

Klassen, R. M., & Kim, L. E. (2017a). Assessing critical attributes of prospective teachers: Implications for selection into initial teacher education programs, In: BJEP. pp. 5-22. http://eprints.whiterose.ac.uk/124823/

Klassen, R. M., & Tze, V. M. C. (2014a). Teachers' self-efficacy, personality, and teaching effectiveness: A meta-analysis. *Educational Research Review, 12*, 59-76.

Klassen, R. M., Al-Dhafri, S., Hannok, W., & Betts, S. M. (2011). Investigating pre-service teacher motivation across cultures using the Teachers' Ten Statements Test. *Teaching and Teacher Education, 27*(3), 579-588.

Klassen, R. M., Durksen, T. L., Al Hashmi, W., Kim, L. E., Longden, K., Metsäpelto, R. L., ... & Györi, J. G. (2018). National context and teacher characteristics: Exploring the critical non-cognitive attributes of novice teachers in four countries. *Teaching and Teacher Education*, *72*, 64-74

Klassen, R., Durksen, T., Kim, L. E., Patterson, F., Rowett, E., Warwick, J., Warwick, P. & Wolpert, M. A. (2017b). Developing a Proof-of-Concept Selection Test for Entry into Primary Teacher Education Programs. *International Journal of Assessment Tools in Education (IJATE)*, *4*(2), *96-114*.

Klassen, R., Durksen, T., Rowett, E., & Patterson, F. (2014b). Applicant Reactions to a Situational Judgment Test used for Selection into Initial Teacher Training. *International Journal of Educational Psychology, 3*(2), 104-124.

Klassen, R.M. & Kim, L. E (2017c, september). Selecting Teachers and Prospective Teachers: A Meta-Analysis. Paper presented at the 17th Biennial EARLI conference, Tampere, Finland

Klassen. R. (2016). Can you spot a good teacher from their characteristics? Guardian Online: http://www.theguardian.com/teacher-network/2016/mar/02/can-you-spot-a-good-teacher-from-their-characteristics

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

Kolman, L., Noorderhaven, N. G., Hofstede, G., & Dienes, E. (2003). Cross-cultural differences in Central Europe. *Journal of Managerial Psychology*, *18*(1), 76-88

Konrath, S. H., & Grynberg, D. (2013). The Neurobiology and Psychology of Empathy

Krumm, S., Lievens, F., Hüffmeier, J., Lipnevich, A. A., Bendels, H., & Hertel, G. (2015). How "situational" is judgment in situational judgment tests?. *Journal of Applied Psychology*, *100*(2), 399

Kubr, M. (1993). *How to select and use consultants: A client's guide*. OIT,.

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.

Kyriacou, C. (2007). *Essential teaching skills* (3rd ed.): Nelson Thornes.

Lakhal, S., Sévigny, S., & Frenette, É. (2015). Personality and student performance on evaluation methods used in business administration courses. *Educational Assessment, Evaluation and Accountability, 27*(2), 171-199

Lee, L.-T., & Lee, T.-T. (2011). Investigating soft skills for success in the workforce: Perceptions of elementary school teachers. *International Review of Social Sciences and Humanities, 1*(2), 140-149.

Lemay, J. F., Lockyer, J. M., Collin, V. T., & Brownell, A. K. W. (2007). Assessment of non-cognitive traits through the admissions multiple mini-interview. *Medical education, 41*(6), 573-579.

Lievens, F. (2006). International situational judgment tests. In J. A. Weekley & R. E. Ployhart (Eds.), *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. (2017). Assessing Personality–Situation Interplay in Personnel Selection: Toward More Integration into Personality Research. *European Journal of Personality, 31*(5), 424-440.

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., & Patterson, F. (2011). The validity and incremental validity of knowledge tests, low-fidelity simulations, and high-fidelity simulations for predicting job performance in advanced-level high-stakes selection. *Journal of Applied Psychology, 96*(5), 927

Lievens, F., Buyse, T., & Sackett, P. R. (2005). The operational validity of a video-based situational judgment test for medical college admissions: illustrating the

importance of matching predictor and criterion construct domains. *Journal of Applied Psychology*, *90*(3), 442

Lievens, F., Corstjens, J., Sorrel, M. Á., Abad, F. J., Olea, J., & Ponsoda, V. (2015). The Cross-cultural Transportability of Situational Judgment Tests: How does a US-based integrity situational judgment test fare in Spain?. *International Journal of Selection and Assessment*, *23*(4), 361-372.

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., & Buyse, T. (2009). The effects of response instructions on situational judgment test performance and validity in a high-stakes context. *Journal of Applied Psychology*, *94*(4), 1095

Lind, G. (2005, August). The cross-cultural validity of the Moral Judgment Test: Findings from 29 cross-cultural studies. In *Presentation at the conference of the American Psychological Association* (pp. 18-21).

Liu, S., & Meng, L. (2009). Perceptions of teachers, students and parents of the characteristics of good teachers: A cross-cultural comparison of China and the United States. *Educational Assessment, Evaluation and Accountability*, *21*(4), 313-328

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

Madkoor, A. (2005). *The Future Teacher: for best performance*. Dar Alfekr Alarabee, Cairo, Egypt (Arabic)

Makransky, G., Havmose, P., Vang, M. L., Andersen, T. E., & Nielsen, T. (2017). The predictive validity of using admissions testing and multiple mini-interviews in undergraduate university admissions. *Higher Education Research & Development*, *36*(5), 1003-1016.

Maria Assunção Flores (2001) Person and Context in Becoming a New Teacher, *Journal of Education for Teaching*, 27:2, 135-148

Marshall, C., & Rossman, G. B. (2014). *Designing qualitative research*. Sage publications.

Mat Kasim, M., Ramli, R., Ibrahim, H., Ghazali, M. I., Mohmad Kamal, F., & Vikneswari, S. (2012). Prioritizing of criteria in teacher-candidate selection

process by analytic hierarchy process. *International Journal of Management Studies (IJMS)*, *20*(2), 15-28.

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

McAllister, G., & Irvine, J. J. (2002). The role of empathy in teaching culturally diverse students: A qualitative study of teachers' beliefs. *Journal of Teacher Education*, *53*(5), 433-443

McCarthy, J. M., Bauer, T. N., Truxillo, D. M., Anderson, N. R., Costa, A. C., & Ahmed, S. M. (2017). Applicant Perspectives During Selection: A Review Addressing "So What?,""What's New?," and "Where to Next?". *Journal of Management*, *43*(6), 1693-1725

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

McFarland, L. A., & Ryan, A. M. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology*, *85*(5), 812.

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.

Menter, I., Hulme, M., Elliot, D., Lewin, J., Baumfield, V., Britton, A., ... & Patrick, F. (2010). *Literature review on teacher education in the 21st century*. Scottish Government Social Research

Meriac, J. P. (2015). Examining relationships among work ethic, academic motivation and performance. *Educational psychology*, *35*(5), 523-540.

Mertens, D. M. (2014). *Research and Evaluation in Education and Psychology: Integrating Diversity With Quantitative, Qualitative, and Mixed Methods:*Sage Publications.

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

Metsäpelto, R. L., & Poikkeus, A. M (2017) Developing Student Selection for Teacher Training: Piloting Situational Judgment Test in Finland. Paper presented at the 17th Biennial EARLI conference, Tampere, Finland

Mohamed, Z., Valcke, M., & De Wever, B. (2017). Are they ready to teach? Student teachers' readiness for the job with reference to teacher competence frameworks. *Journal of Education for Teaching*, *43*(2), 151-170.

Motowidlo, S. J., Crook, A. E., Kell, H. J., & Naemi, B. (2009). Measuring procedural knowledge more simply with a single-response situational judgment test. *Journal of Business and Psychology*, *24*(3), 281-288.)

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., Ghosh, K., Mendoza, A. M., Buchanan, A. E., & Lerma, M. N. (2016). A context-independent situational judgment test to measure prosocial implicit trait policy. *Human Performance*, *29*(4), 331-346

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)

Muijs, D., & Reynolds, D. (2010). *Effective teaching: Evidence and practice*: Sage.

Muijs, D., Kyriakides, L., van der Werf, G., Creemers, B., Timperley, H., & Earl, L. (2014). State of the art–teacher effectiveness and professional learning. *School effectiveness and school improvement*, *25*(2), 231-256

National Centre for Statistic and Information (NCSI) (Oman). (2018). Retrieved in July 12th, 2018 from www.ncsi.gov.om

Nguyen, N. T., Biderman, M. D., & McDaniel, M. A. (2005). Effects of response instructions on faking a situational judgment test. *International Journal of Selection and Assessment*, *13*(4), 250-260.

Nielsen, T., & Friderichsen, I. S. (2017). Non-cognitive skills as admission criteria for university: A systematic review. In *17th Conference of the European Association for Research in Learning and Instruction*.

O'Connell, M. S., Hartman, N. S., McDaniel, M. A., Grubb, 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

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

Oliver, P. (2010). *The Student's Guide to Research Ethics*. Maidenhead: Open University Press.

Oman Education Council (OEC). (2014). *The final report for the National seminar: Education in Oman and the road to the future*, Muscat (in Arabic)

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

Oman Education Council (OEC). (2016b). *Education in Oman: national indicators & international comparisons*, Muscat (in Arabic)

Oman Education Council (OEC). (2017a). *Education and development trends and current and future employment opportunities in the labour market*, A symposium held in Muscat - May 2017 (in Arabic)

Oman Education Council (OEC). (2017b) The philosophy of Education in the Sultanate of Oman. Retrieved from https://www.educouncil.gov.om/downloads/AZ27826wxDsr.pdf

Oman Ministry of Education (OMoE). (2014a). *National Education for all Report in Oman.* Muscat

Oman Ministry of Education (OMoE). (2014b). The Specialised Centre for Professional Training of Teachers. Retrivied from http://havasapps.com/test/moe/wp-content/uploads/2016/07/TT_Information_E.pdf

Oman Ministry of Education (OMoE). (2011a). *The National Report for PIRLS 2011*. Muscat, Oman: MOE(in Arabic)

Oman Ministry of Education (OMoE). (2011b). *The National Report for TIMSS 2011*. Muscat, Oman: MOE(in 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 Ministry of Education (OMoE). (2015). The guidence for the schools' job discribtion. Rretrieved from www.moe.gov.om (in Arabic)

Oman Ministry of Education (OMoE). (2018a). Recruitment procedures. Retrieved from http://home.moe.gov.om/arabic/module.php?module=topics-showtopic&CatID=1&ID=5660 (in Arabic)

Oman Ministry of Education (OMoE). (2018b). The Annual Educationla Statistics Book 2017/2018. Retrieved from http://home.moe.gov.om/arabic/images/library/file/MOEAnnrep%202017-2018.pdf

Oman Ministry of Higher Education (OMoHE). (2014). *The Student's Guide to enroll in Higher Education Institutes 2014/2015*. Muscat, Oman: MoHE(in Arabic)

Oman Ministry of Higher Education (OMoHE). (2017). *The Annual Statistical Report of Higher Education in Oman 2016/2017*. Muscat, Oman: MoHE(in Arabic)

Oman Ministry of Higher Education (OMoHE). (2018). Graduates survey. Retrieved from http://ogss.gov.om/index.php/ar/information-center-2/reports

Oman Ministry of Information (OMoI). (2016). Oman 2016. Muscat

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

Oostrom, J. K., Born, M. P., Serlie, A. W., & van der Molen, H. T. (2012). Implicit trait policies in multimedia situational judgment tests for leadership skills: Can they predict leadership behavior?. *Human Performance*, *25*(4), 335-353

Organisation for Economic Co-operation and Development, Santiago, P., & Source OECD. (2005). *Teachers matter: Attracting, developing and retaining effective teachers*. Organisation for Economic Co-operation and Development.

Othman, J., Borhan, L., Abdullah, N., Jaafar, W., Marzuki, W., Noah, S. M., ... & Ibrahim, S. (2008). The construction of the Malaysian Educators Selection Inventory (MEdSI): a large scale assessment initiative. *Asian Journal of University Education*, *4*(1), 113-126

Pallant, J. (2010, 4th Edition). *SPSS survival manual: A step by step guide to data analysis using SPSS for Windows (versions 10 and 11): SPSS student version 11.0 for Windows*. Milton Keynes: Open University Press.

Patterson, F., Ashworth, V., Mehra, S., & Falcon, H. (2012a). Could situational judgement tests be used for selection into dental foundation training?. *British dental journal, 213*(1), 23-26

Patterson, F., Ashworth, V., Murray, H., Empey, L., & Aitkenhead, A. (2014). Analysis of the Situational Judgement Test for selection to the Foundation Programme 2014

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

Patterson, F., Baron, H., Carr, V., Plint, S., & Lane, P. (2009). Evaluation of three short-listing methodologies for selection into postgraduate training in general practice. *Medical education, 43*(1), 50-57

Patterson, F., Ferguson, E., & Thomas, S. (2008). Using job analysis to identify core and specific competencies: implications for selection and recruitment. *Med Educ, 42*(12), 1195-1204.

Patterson, F., Ferguson, E., Lane, P., Farrell, K., Martlew, J., & Wells, A. (2000). A competency model for general practice: implications for selection, training, and development. *British Journal of General Practice, 50*(452), 188-193.

Patterson, F., Kerrin, M., Born, M. P., Oostrom, J. K., & Prescott-Clements, L. (2015a). What are values and how can we assess them? Implications for values-based recruitment in healthcare. *Researching Medical Education*, 167-180

Patterson, F., Knight, A., Dowell, J., Nicholson, S., Cousans, F., & Cleland, J. (2016a). How effective are selection methods in medical education? A systematic review. *Medical Education, 50*(1), 36-60

Patterson, F., Lievens, F., Kerrin, M., Munro, N., & Irish, B. (2013a). The predictive validity of selection for entry into postgraduate training in general practice: evidence from three longitudinal studies. *Br J Gen Pract, 63*(616), e734-e741.

Patterson, F., Lievens, F., Kerrin, M., Zibarras, L., & Carette, B. (2012c). 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., Prescott-Clements, L., Zibarras, L., Edwards, H., Kerrin, M., & Cousans, F. (2016b). Recruiting for values in healthcare: a preliminary review of the evidence. *Advances in Health Sciences Education, 21*(4), 859-881

Patterson, F., Rowett, E., Hale, R., Grant, M., Roberts, C., Cousans, F., & Martin, S. (2016c). 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

Patterson, F., Tavabie, A., Denney, M., Kerrin, M., Ashworth, V., Koczwara, A., & MacLeod, S. (2013b). A new competency model for general practice: implications for selection, training, and careers. *British Journal of General Practice, 63*(610), e331-e338.

Patterson, F., Zibarras, L., & Ashworth, V. (2015b). Situational judgement tests in medical education and training: Research, theory and practice: AMEE Guide No. 100. *Med Teach,* 1-15. doi:10.3109/0142159X.2015.1072619

Patterson, F., Zibarras, L., Carr, V., Irish, B., & Gregory, S. (2011). Evaluating candidate reactions to selection practices using organisational justice theory. *Medical education, 45*(3), 289-297

Peeters, H., & Lievens, F. (2005). Situational judgment tests and their predictiveness of college students' success: The influence of faking. *Educational and Psychological Measurement, 65*(1), 70-89.:

Peus, C., Braun, S., & Frey, D. (2013). Situation-based measurement of the full range of leadership model—Development and validation of a situational judgment test. *The Leadership Quarterly, 24*(5), 777-795.

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.

Ployhart, R. E., & Ehrhart, M. G. (2003). Be careful what you ask for: Effects of response instructions on the construct validity and reliability of situational judgment tests. *International Journal of Selection and Assessment, 11*(1), 1-16.)

Ployhart, R. E., & Holtz, B. C. (2008). The diversity–validity dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology, 61*(1), 153-172

Ployhart, R. E., & MacKenzie Jr, W. I. (2011). Situational judgment tests: A critical review and agenda for the future. *APA handbook of industrial and organizational psychology, Vol. 2. Selecting and developing members for the organization* (pp. 237-252). Washington, DC, US: American Psychological Association. http://dx.doi.org/10.1037/12170-008

Ployhart, R. E., Weekley, J. A., Holtz, B. C., & Kemp, C. (2003). Web-based and paper-and-pencil testing of applicants in a proctored setting: are personality, biodata, and situational judgment tests comparable?. *Personnel Psychology*, *56*(3), 733-752.

Praetorius, A. K., Lauermann, F., Klassen, R. M., Dickhäuser, O., Janke, S., & Dresel, M. (2017). Longitudinal relations between teaching-related motivations and student-reported teaching quality. *Teaching and teacher education*, *65*, 241-254

Prasad, J. J., Showler, M. B., Schmitt, N., Ryan, A. M., & Nye, C. D. (2016). Using Biodata and Situational Judgment Inventories across Cultural Groups. *International Journal of Testing*, *17*(3), 210-233

Punch, K. F. (2013). *Introduction to social research: Quantitative and qualitative approaches*: Sage.

Ramli, R., Ghazali, M. I., Ibrahim, H., Kasim, M. M., Kamal, F. M., & Vikneswari, S. (2013). A Hybridized Competency-Based Teacher Candidate Selection System. *World Academy of Science, Engineering and Technology, International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering*, *7*(5), 1117-1121

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

Rawling, K. (1985). *The Seven-Point Plan by Alec Rodger: New perspectives fifty years on.* NFER-Nelson.

Reed, R. R., & Tsaur, J. R. (2008). *The importance of non-cognitive skills: Observations from Asian educational systems*. Retrieved from econ.sinica.edu.tw

Relf, S., & Hobbs, D. (1999). The recruitment and selection of new teachers of mathematics: the needs of secondary schools versus the teacher training agenda. *Research Papers in Education*, 14(2), 165-180.

Robert A. Roe. Designing a selection system in Herriot, P. (Ed.1989). *Assessment and selection in organisations: methods and practices for recruitment and appraisal*. John Wiley & SONS ltd. England

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: a meta-analysis of longitudinal studies. *Psychological bulletin*, *132*(1), 1.

Robertson-Kraft, C., & Duckworth, A. L. (2014). True grit: Trait-level perseverance and passion for long-term goals predicts effectiveness and retention among novice teachers. *Teachers College record (1970)*, *116*(3).

Rockstuhl, T., Ang, S., Ng, K. Y., Lievens, F., & Van Dyne, L. (2015). Putting judging situations into situational judgment tests: Evidence from intercultural multimedia SJTs. *Journal of Applied Psychology*, *100*(2), 464.:

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.

Sahlberg, P. (2015). *Teach for Finland? Why it won't happen*.". Washington Post, 12.

Salgado, J. F., & Moscoso, S. (2002). Comprehensive meta-analysis of the construct validity of the employment interview. *European Journal of Work and Organizational Psychology*, *11*(3), 299-324.

Salleh, A. M. L. A., Al-Kalbani, M. S. A., & Mastor, K. A. (2010). Testing the five factor personality model in Oman. In Proc. the wseas international conference on sociology, psychology, philosophy. At Penang: WSEAS (pp. 11-17

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.

Sammons, P., Lindorff, A. M., Ortega, L., & Kington, A. (2016). Inspiring teaching: learning from exemplary practitioners. *Journal of Professional Capital and Community*, *1*(2), 124-144.

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), 54-71.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological bulletin*, *124*(2), 262.

Schmitt, D. P., Long, A. E., McPhearson, A., O'Brien, K., Remmert, B., & Shah, S. H. (2016). Personality and gender differences in global perspective. *International Journal of Psychology*. *52*, 45-56.

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

Schumacher, G., Grigsby, B., & Vesey, W. (2015). Determining effective teaching behaviors through the hiring process. *International Journal of Educational Management, 29(1)*, 139-155.

Schwager, I. T., Hülsheger, U. R., Bridgeman, B., & Lang, J. W. (2015). Graduate Student Selection: Graduate record examination, socioeconomic status, and undergraduate grade point average as predictors of study success in a western European University. *International Journal of Selection and Assessment*, *23*(1), 71-79

Schwartz, S. H. (1999). A theory of cultural values and some implications for work. *Applied psychology*, *48*(1), 23-47

Shane S. Bush, Jerry J. Sweet, Kevin J. Bianchini, Doug JohnsonGreene, Pamela M. Dean & Mike R. Schoenberg (2018): Deciding to adopt revised and new psychological and neuropsychological tests: an inter-organizational position paper, The Clinical Neuropsychologist, DOI: 10.1080/13854046.2017.1422277

Sharma, S., Gangopadhyay, M., Austin, E., & Mandal, M. K. (2013). Development and validation of a situational judgment test of emotional intelligence. *International Journal of Selection and Assessment*, *21*(1), 57-73.

Shemitt & Ott-Holand (2012) Theoritical and Practical Issues: Research Needs in Shemitt, N. (Ed). (2012) *The Oxford Handbook of Personnel Assessment and Selection* . OUP USA, 2012.

Shepard, L. A. (1993). Chapter 9: Evaluating Test Validity. *Review of research in education*, *19*(1), 405-450.

Sherman, R. A., Nave, C. S., & Funder, D. C. (2013). Situational construal is related to personality and gender. *Journal of Research in Personality*, *47*(1), 1-14

Shultz, M. M., & Zedeck, S. (2011). Predicting lawyer effectiveness: Broadening the basis for law school admission decisions. *Law & Social Inquiry, 36*(3), 620-661.

Simon, E., Walsh, K., Paterson-Brown, F., & Cahill, D. (2015). Does a high ranking mean success in the Situational Judgement Test?. *The clinical teacher*, *12*(1), 42-45.

Sorrel, M. A., Olea, J., Abad, F. J., de la Torre, J., Aguado, D., & Lievens, F. (2016). Validity and reliability of situational judgement test scores: A new approach based on cognitive diagnosis models. *Organizational Research Methods*, *19*(3), 506-532.

Stankov, L. (2013). Non-cognitive predictors of intelligence and academic achievement: An important role of confidence. *Personality and Individual Differences, 55*, 727–732

Stankov, L., & Lee, J. (2014). Quest for the best non-cognitive predictor of academic achievement, *Educational Psychology, Vol. 34, No.* 1, 1–8

Stebbins. R. A. (Ed.). (2001*). Exploratory Research in the Social Sciences*. Thousand Oaks, CA: SAGE Publications.

Stemig, M. S., Sackett, P. R., & Lievens, F. (2015). Effects of organizationally endorsed coaching on performance and validity of situational judgment tests. *International Journal of Selection and Assessment*, *23*(2), 174-181

Stemler, S. E., Aggarwal, V., & Nithyanand, S. (2016). Knowing What NOT To Do Is a Critical Job Skill: Evidence from 10 different scoring methods. *International Journal of Selection and Assessment*, *24*(3), 229-245.)

Stephens, P., Tønnessen, F. E, & Kyriacou, C. (2004). Teacher training and teacher education in England and Norway: a comparative study of policy goals. *Comparative Education, 40*(1), 109-130.

Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health measurement scales: a practical guide to their development and use*. Oxford University Press, USA

Strikwerda-Brown, J., Oliver, R., Hodgson, D., Palmer, M., & Watts, L. (2008). Good Teachers / Bad Teachers: How Rural Adolescent Students' Views of Teachers Impact on Their School Experiences. *Australian Journal of Teacher Education, 33*(6). doi:10.14221/ajte.2008v33n6.3

Stronge, J. H. (2007). *Qualities of effective teachers*: ASCD.

St-Sauveur, C., Girouard, S., & Goyette, V. (2014). Use of Situational Judgment Tests in Personnel Selection: Are the different methods for scoring the response options equivalent?. *International Journal of Selection and Assessment*, *22*(3), 225-239.

Sultan Qaboos University (SQU). 2017. In http://www.squ.edu.om/

Tatto, M. T. (2008). Teacher policy: A framework for comparative analysis. *Prospects*, *38*(4), 487-508.

Tettegah, S., & Anderson, C. J. (2007). Pre-service teachers' empathy and cognitions: Statistical analysis of text data by graphical models. *Contemporary Educational Psychology*, *32*(1), 48-82

Thorsteinson, T. J. (2018). A meta-analysis of interview length on reliability and validity. *Journal of Occupational and Organizational Psychology*, *91*(1), 1-32.

Timoštšuk, I., & Ugaste, A. (2012). The role of emotions in student teachers' professional identity. *European Journal of Teacher Education*, *35*(4), 421-433.

TIMSS (2015). In http://timssandpirls.bc.edu/timss2015/international-results/download-center/

Tseng, W. S. (Ed.). (2001). *Handbook of cultural psychiatry*. Academic Press.

Turner, S., & Turner, T. (1997). Entering Initial Teacher Education: science graduates' views of the interview and selection process. *European journal of teacher education, 20(2),* 125-136

UNESCO. (2013). *Education for All Global Monitoring Report 2013/4 Teaching and Learning: Achieving Quality for All*: Paris: UNESCO.

Vinchur, A. J., & Koppes Bryan, L. L. A history of personnel selection and assessment. In Schmitt, N. (Ed.). (2012). *The Oxford handbook of personnel assessment and selection*. OUP USA. 9-30.

Wang, A. H., Coleman, A. B., Coley, R. J., & Phelps, R. P. (2003). *Preparing Teachers around the World.* Policy Information Report.

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.

Watt, H. M., & Richardson, P. W. (2007). Motivational factors influencing teaching as a career choice: Development and validation of the FIT-Choice Scale. *The Journal of experimental education*, *75*(3), 167-202.

Waxman, H. C. (1983). Effect of teachers' empathy on students' motivation. *Psychological reports*, *53*(2), 489-490.

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.

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.

Winterton, J. (2009). Competence across Europe: highest common factor or lowest common denominator?. *Journal of European industrial training*, *33*(8/9), 681-700.

Wyatt, M. (2013). Motivating teachers in the developing world: Insights from research with English language teachers in Oman. *International Review of Education*, *59*(2), 217-242.

Zayed, K., Abu Hilal, M., & Diabat, F. (2011). Attitudes of Omani PE Teachers and PE College Students Towards Teaching PE. *Scentific Journal of Physical Education & Sports, Helwan University*, *2*(63), 269-279.

Zhou, K. (2016) *Non-cognitive skills: Definitions, measurement and malleability*, Paper commissioned for the Global Education Monitoring Report 2016, Education for people and planet: Creating sustainable futures for all.