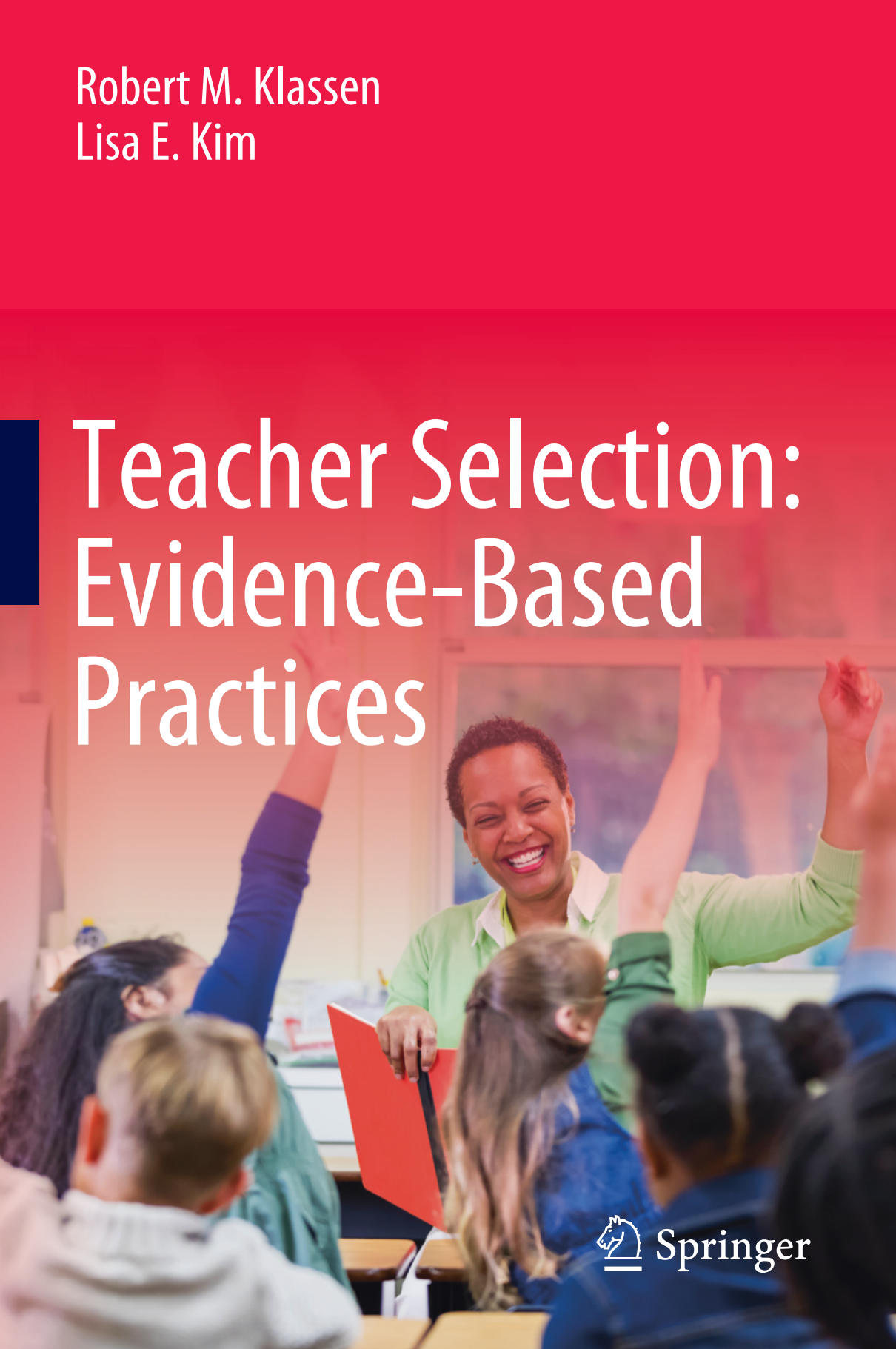
Robert M. Klassen
Lisa E. Kim

Teacher Selection: Evidence-Based Practices



Springer

Teacher Selection: Evidence-Based Practices

Robert M. Klassen · Lisa E. Kim

Teacher Selection: Evidence-Based Practices

Springer

Robert M. Klassen
Department of Education
University of York
Heslington, York, UK

Lisa E. Kim
Department of Education
University of York
Heslington, York, UK

ISBN 978-3-030-76186-8 ISBN 978-3-030-76188-2 (eBook)
https://doi.org/10.1007/978-3-030-76188-2

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Teachers have a profound influence on important educational outcomes: not just short-term academic outcomes, but outcomes related to student well-being, self-concept, and life-long attitudes towards learning. Most of us can recall teachers that had a deep and formative influence on our own love of learning, and on our feelings of safety and belonging in a certain classroom or school. Without doubt, the quality of our teachers has a fundamental influence not only on the personal well-being of individual students, but on 'educational health' and future prosperity at the country level. And yet, there is more care, rigour, and research attention given to how we choose employees into large corporations, financial services, the civil service, or health care than to identifying the best possible prospective teachers. Even in the face of fluctuating demands for teacher training places and teaching positions—whether systems suffer a shortage or a surfeit of applicants for training (and the demand for training spiked in many settings during the Covid-19 crisis)—improving the ways we identify, select, and develop prospective teachers is worth pursuing for a nation's educational health.

This book is based on three fundamental positions. First, there are individual differences in the effectiveness of teachers. Most research shows that teachers grow in their impact on student learning and well-being with time: candidates selected for training are not the finished product. Nevertheless, there is strong evidence that teachers vary in their effectiveness—however defined and measured—that can be traced back to the 'starting point' of selection into training. Second, individual differences at the point of selection play an important role in shaping the future teacher: the personal characteristics (cognitive and non-cognitive attributes) that have developed through a wide range of life circumstances shape the behaviors, attitudes, and beliefs of prospective teachers. Some of these characteristics are mutable; some less so. Third, current teacher selection practices tend to be uninformed by research, with little evidence supporting their efficacy. Research and practices in teacher selection have not kept up with selection research and practices in other fields, at least partially due to a reluctance of some in the education field to embrace the notion of differential teacher effectiveness, and due to a historical mistrust of systematic (and psychometric) approaches to recruitment and selection.

Kahneman, in *Thinking, Fast and Slow* famously described how most of us are prone to overconfidence in trusting our intuition when making judgments, even when we are aware of our biases. Many of us trust our judgment when it comes to making hiring and selection decisions, but research suggests that our judgment is 'riddled with biases', and we tend to implicitly prefer candidates who are similar to us, even when objective data might point us in a different direction. Furthermore, we tend to retrospectively affirm the results of our selection decisions in the absence of supporting evidence. We confidently make decisions based on hidden biases and incomplete data all the time: we are overconfident about our decision-making abilities and trust our intuition to guide us to make accurate decisions. When selecting candidates for teacher training or for teaching jobs, we are influenced by our conscious and unconscious biases, intuition, and by the undue confidence we have in the correctness of our past decision-making.

In this book we take the stance that identifying the best possible candidates for teaching is worth serious scientific consideration because the stakes are so high, not just for individual students, classrooms, and schools, but for a nation's well-being. Selecting the best possible prospective teachers will never be an exact science, and as with all predictions of human behavior there will be hits and misses, but we can use theories and methods from other disciplines, especially organizational psychology and medical education to improve our 'hit rate' of identifying the best possible new teachers. At the heart of selection is the question, *Does this candidate have the potential to be an effective teacher?* and if there are more candidates than places, *Is Candidate X more likely to succeed than Candidate Y in our program (or in practice)?* But understanding, or even defining, teacher effectiveness is not so straightforward, and in this book, after making a case for the importance of teacher selection, we consider in Part I what 'teacher effectiveness' might mean, and how personal characteristics might be associated with successful practice. These personal characteristics, or individual attributes, are not equally predictive of success in all settings, and we consider how these characteristics might vary among individuals and across cultural contexts.

In Part II we delve into selection theory, research, and practice, and examine the problems that crop up when developing a selection program or strategy, and how other fields have wrestled with the challenges of developing and testing selection strategies and methods in their respective contexts. In Part III we turn our attention to teacher selection, and in these four chapters we explore the history of teacher selection, current practices, some evidence-based practices that have been trialled in the last few years, and a consideration of how selection programs might be implemented in a range of real-life settings. We conclude the book in Part IV with a look beyond current practices, and consider how we might apply the lessons of teacher selection to the recruitment and development of prospective teachers.

UNESCO has recently projected a need for nearly 70 million new teachers to be selected and trained in the next decade, and we know that the quality of candidates is influenced by the quality of recruitment, selection, and development processes. Implementing state-of-the-art teacher selection methods can save thousands of hours in the recruitment process, and can act as a 'quick win' in improving the teacher

workforce. Our goal in this book is to highlight that one approach to building a stronger teacher workforce is through research-based teacher selection. We see a future where teachers are selected with the same rigour and care as candidates in other professions, and we hope that this book begins a conversation among researchers, practitioners, and policy-makers about the importance of using the best possible methods to select the best possible teachers.

Heslington, York, UK

Robert M. Klassen
Lisa E. Kim

Acknowledgements

This book is the result of multiple conversations with academic colleagues, not just about teacher recruitment and selection, but about much more interesting topics, like the intersection of life purpose, family, and work. After a decade of research on achievement motivation, I was bothered by the lack of potential application of my research, and I wondered how educational psychology could make a bigger impact on real-life problems. A big thank you to my academic colleagues around the world, but especially Tracy Durksen, Ellen Usher, Flaviu Hodis, Richard Remedios, Paul Tiffin, Sündüs Yerdelen, Lia Daniels, Rauno Parrila, Andrew Joyce-Gibbons, and Colin Anderson for their willingness to engage in discussions about topics that were important for me, if not necessarily for them.

I would also like to highlight the support given by my fantastic colleagues at the Teacher Selection Project at the University of York who have contributed much inspiration for the last six years, and who have made working on this topic so enjoyable and fulfilling. Many people have contributed to our work, but I want to especially thank the *TSP Superheroes!* Group (Liz Maxwell, Jade Rushby, Lisa Bardach, Helen Granger, and Liz Hooper) for making our working environment so enjoyable. I also thank my co-author Lisa Kim for her willingness to challenge ideas and for contributing to this book and the research on teacher selection. It has been a pleasure working with so many of you on our ERC research project and on planning the next stage of our work.

I'm grateful to our supportive education partners from around the world who have joined with us in teacher selection work at their institutions: Tracy and Lynn from UNSW and Wollongong; Michelle from Liverpool Hope; Paul and Jane from Cambridge; Catherine and Linda from Strathclyde, Tallulah, Gemma and colleagues from Teach First; Matthew, Richard and Peter from York St John; Ken and Audrey from Stranmillis; Claire from York; Riitta-Leena and Anna Maija from Jyväskylä; Ina from Teach for Bulgaria; János in Hungary; Anna and Susana in Peru; Ken and the TTC selection team from Malawi; our colleagues at Alpha Plus; and most recently, Lucas, Louise, and M. Elazhari and colleagues in Morocco, and Deborah and colleagues in Kenya. Thanks also to the Social Sciences and Humanities Research

Council of Canada and the European Research Council who have provided generous funding for the research that is represented in this book.

Finally, I'm eternally grateful for the love and encouragement from my wife, Lenore, who has endured my early wake-ups, and has encouraged me to pursue my academic dreams. Thanks also to my kids—Andrea, Danielle, and Joel—who were dragged more-or-less willingly to the UK from Canada, and who are, without doubt, my three proudest accomplishments. And of course, thanks Mom, for your continuous love and support.

Robert M. Klassen

I'm grateful to Professor Rob Klassen and the European Research Council for providing me with opportunities to explore the area of teacher selection when I was a Postdoctoral Research Fellow for the Teacher Selection Project. I would also like to thank the people and organizations I worked with at that time, both in the UK and worldwide, for stretching my thinking in this area. Turning to a more personal arena, I'm very thankful for the unwavering support and love my family has always shown me. Lastly, I would like to thank teachers all over the world for the amazing work that they do.

Lisa E. Kim

Contents

About the Authors

Robert M. Klassen is Professor and Chair of the Psychology in Education Research Centre at the University of York, UK, and an Adjunct Professor in the Department of Educational Psychology at the University of Alberta in Canada. Klassen's research focuses on achievement motivation and the factors that lead to teacher effectiveness. His work on the *Teacher Selection Project* (TSP) was funded by the European Research Council (2015-2021) and conducted in eight countries. He has published over 100 peer-reviewed articles and book chapters on the topics of motivation, and teacher recruitment, selection, and development. He is a Chartered Psychologist in the UK and a Fellow of the American Psychological Association.

Lisa E. Kim is a Lecturer/Assistant Professor in Psychology in Education at the University of York, UK. She is also a Chartered Psychologist (CPsychol) in the UK. She received her Ph.D. in Psychology from the University of Sydney, Australia and worked as a Postdoctoral Research Fellow at the University of York for the Teacher Selection Project, funded by the European Research Council Consolidator Grant. Her research interests lie in understanding what makes a great teacher. Specifically, she investigates the roles, assessment and development of teacher non-cognitive characteristics and teacher effectiveness. She achieves this by using interdisciplinary frameworks, and both quantitative and qualitative methods. For more information, please visit lisaekim.com.

Chapter 1
The Importance of Selecting the Most Effective Teachers

Abstract One way to improve the teacher workforce is to recruit and select the best possible new teachers. Countries with the highest performing education systems tend to pay the most attention to the way that new teachers are recruited and selected into training and professional practice. In the opening chapter we briefly examine the history of teacher selection before making the case that improving teacher selection using research- and theory-based approaches is often neglected by education organizations but can be a 'quick win' to improve education systems. We present five key points to consider when evaluating existing selection programs and when planning new teacher selection programs from scratch. We end this introductory chapter by presenting a roadmap of the book's three sections: Part I Identifying the characteristics of effective teachers; Part II Selection methods and practices: Issues and uses; and Part III Teacher selection: Past, present, and future.

It is empirically noted that one teacher has an effect on pupils that is qualitatively termed inspiring, awakening, and that the personality of another teacher is relatively deadening, dulling. Now here is a problem set for inquiry, whether the sciences which have to be drawn upon are sufficiently advanced to provide material for its solution or not.

John Dewey, Philosopher, psychologist and education reformer, 1929.

The best means of improving a school system is to improve its teachers. One of the most effective means of improving the teacher corps is by wise selection.

Ervin Eugene Lewis, Superintendent of Schools, Michigan, 1925 (Jacob, 2016).

Attempting to recruit and select the best possible people to enter the teaching profession is not a new challenge. The 1920s saw prominent education commentators recognizing the need to improve the identification of candidates who were most likely to maximize student learning. Education reformer John Dewey pondered whether educational and psychological research was able to contribute to the important challenge of teacher selection. Educational psychologist F. B. Knight (1922) examined the methods used to select teachers and concluded, "The kind of information usually asked of a candidate does not correlate… with successful performance" (p. 216,

R. M. Klassen and L. E. Kim, *Teacher Selection: Evidence-Based Practices*,
https://doi.org/10.1007/978-3-030-76188-2_1

1922). In 1928, Eston V. Tubbs, Director of Curriculum for Chicago Public Schools, emphasized the importance of selection: "The fact needs to be emphasized and re-emphasized that too much care and thoughtful consideration cannot be given to the importance attaching to the function of the selection of teachers for our schools" (Tubbs, 1928, p. 332). As universal education became increasingly prevalent in the 1920s, the selection of teachers grew in importance, a fact recognized by the leading educational psychologists, philosophers, and policymakers of the time.

Now, almost a century later, not much has changed. Our quest to identify the best possible teachers in the 2020s chimes with the challenges faced in the 1920s: we continue to search for better ways to recruit, select, and develop new teachers. Teacher selection research has advanced in fits and starts in the last 100 years, but in the last decade the field has been strengthened by cross-pollination from disciplines such as organizational psychology and medical education, where much more research attention has been paid to improving selection practices. The purpose of this book is to examine the most important research relevant to teacher selection, and to propose new and better ways to ensure that teachers with the highest potential for success are selected into the teaching profession.

1.1 Strengthening the Teacher Workforce

The teaching workforce can be strengthened through interventions at four different points across the career span (see Fig. 1.1): (a) attraction/recruitment, (b) selection, (c) initial teacher education (ITE), and (d) professional development. Most efforts

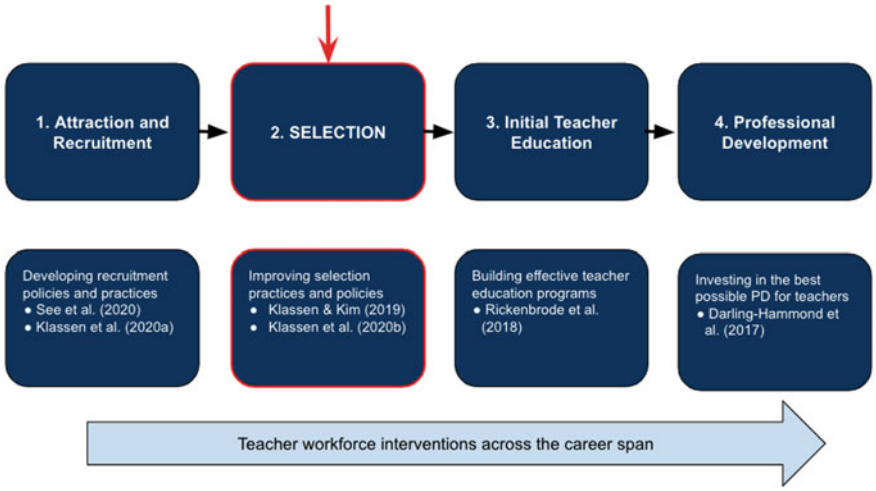


Fig. 1.1 Multiple approaches to improve the quality of the teaching workforce

to improve the teaching workforce have been directed at improving ITE and professional development (e.g., Darling-Hammond, 2017; Rickenbrode, et al., 2018), but the effectiveness of these interventions is mixed (e.g., Jayaram et al., 2012), and costs can be very high. Much less attention has been paid to interventions targeting the first two time points (attraction/recruitment and selection), even though system-wide improvements at these early stages can pay off in an improved workforce with modest investment. Research on attracting and recruiting potential teachers is relatively sparse, with recent reviews (Klassen, Bardach et al., 2021; See et al., 2020) exploring the efficacy of recruitment initiatives, and a recent intervention study highlighting the importance of person-vocation fit in teacher recruitment (Klassen, Granger et al., 2021). Systematic research on teacher selection—in spite of historical calls for greater attention—has been sporadic, but as we shall see in this book, recent advances have rejuvenated the field and provided new pathways to improve educational outcomes.

1.2 Teacher Selection is a 'Quick Win'

Selecting the right people for teaching is an important, but often neglected, pathway to improving an education system. In most systems, there are two stages where teachers and prospective teachers are selected: at the point of entry into ITE and at the point of hiring for teaching jobs. The quality—i.e., predictive accuracy—of selection decisions matters at both stages. An evidence-based selection program is beneficial in any training or employment setting, even when the supply of applicants is limited. Even though high quality ITE and professional development opportunities can improve the effectiveness of teachers, the long-term success of an education system is fundamentally dependent on the 'raw materials' that enter the system because there are significant individual differences in the developmental trajectory of new teachers (Atteberry et al., 2015). Improving the quality of prospective teachers at the point of selection is an efficient and economical decision—i.e., a *quick win*—in improving educational outcomes.

Selection methods are important even when recruitment, not selection, is the goal. In some settings, selecting prospective teachers involves sifting through large numbers of applicants. In Finland and Singapore, competition is fierce for each teacher training place (e.g., Sclafani, 2015). In some locations such as Australia and Canada, an oversupply of qualified teachers has traditionally made the job market tight, with many applicants for each teaching position (Galt, 2017). However, teacher supply and demand fluctuate by region and over time, and teacher shortages are routinely predicted with economic cycles. For example, Baker (2020) reports projections of teacher shortages in Australia partly due to COVID-related economic downturns. In the U.S., there are jurisdictions where teachers are in short supply, but many districts have an oversupply of certified teachers for certain subjects and teaching levels (Jacob, 2016). In all of these cases, even in the case of short supply, selection

methods have the potential to provide useful information about the characteristics of prospective teachers.

Applicant pools are closely linked to overall economic climate (more applicants tend to apply when the economy is in crisis, as in the COVID-19 crisis of 2020–2021), and to additional factors related to the reputation of the profession. When the applicant pool is not very deep, the selection process is not as much about filtering candidates as about evaluating key characteristics of candidates for future development. Selection methods are important when the number of applicants exceeds the number of places but developing the best possible selection methods is also important (a) when there is benefit in identifying 'unsuitable' applicants (i.e., *selecting out*), (b) when profiles of candidate strengths and weaknesses can be valuable for future professional development, and (c) when we want to better understand the underlying reasons for teacher attrition. There are compelling reasons to gather the best possible information from selection methods to maximize the chances that the people entering the teacher pipeline will succeed in their work as teachers.

Identifying the most promising teaching candidates is easy in some ways but challenging in other ways. It is relatively easy to identify the brightest, most capable, and academically competent candidates: we can look at academic records, university entrance test scores, and, if needed, there are numerous reliable and valid tests of academic capability that can be administered. The hard part is identifying the *non-cognitive attributes* (sometimes called 'non-academic attributes' or 'soft skills') of prospective teachers. Teachers' non-cognitive attributes, such as emotion regulation, empathy, and resilience have a significant impact on student achievement (e.g., Hattie, 2009), but these attributes are difficult to evaluate in a reliable, valid, and fair way during selection. Identifying and evaluating key non-cognitive attributes of prospective teachers is the 'holy grail' of teacher recruitment and selection because these attributes are essential for effective teaching, but notoriously difficult to measure.

At the foundation of teacher selection research is the belief that individuals vary in personal attributes and experiences, and that these individual differences are related to future behaviors in training and professional contexts. Some of these individual differences are fleeting (e.g., mood), and some are more enduring (e.g., intelligence and personality). In some settings, selectors might ask candidates to self-evaluate these attributes through direct assessment (e.g., *How good are you at working in groups?*), while in other settings, selectors attempt to infer these attributes by observation or through indirect assessments (e.g., through evaluation of personal statements or letters of reference). Research evaluating the effectiveness of teacher selection methods is rare (Klassen & Kim, 2019), but it has the potential to improve the quality of education systems by improving the quality of teachers entering the system.

Facets of teacher effectiveness. In the case of teacher selection, the key 'future behavior' being predicted is the broad, difficult-to-measure, and hotly debated construct of *teacher effectiveness.* Whether the selection program is aimed at selecting candidates for an initial teacher education program or for a teaching position, selectors must consider background factors (e.g., previous teaching-relevant

experience), cognitive factors (e.g., numeracy and literacy skills, subject area knowledge, and general academic aptitude), and non-cognitive factors (e.g., interpersonal skills, personality, and motivation). The selection criteria for ITE mirror the criteria used for selection into teaching jobs, with selection methods that (a) test a wide range of cognitive and non-cognitive attributes, (b) highlight the necessity for strong interpersonal and communication skills, and (c) assess relevant background factors that have been shown to be related to teaching effectiveness. The factors are not too different for selection into training programs and for teaching positions: the important cognitive, non-cognitive, and background factors must all be weighed to identify the strongest candidates for a position in a training program or for employment.

Selection for teacher education and selection for employment. In this book we examine selection research and practice at two points: at entry into teacher education and at entry into employment. The weighing of specific criteria at these two points of selection may change (e.g., secondary school achievement record may hold less weight for selection of practicing teachers than for candidates for initial teacher education), but the overall goal—and the challenges—of selection at the two levels is the same: to identify the attributes associated with teaching success and to assess these attributes in candidates.

The question of *when* to make selection decisions is partly influenced by the question about the stability of individual differences that are associated with teacher effectiveness. In short, the question revolves around the question of how cognitive and non-cognitive attributes are likely to change over time during teacher training and practice. For example, are ITE candidates who score lower in conscientiousness compared to their peers likely to continue scoring lower in conscientiousness over time? Some educational economists have proposed that routine filtering of teachers after hiring based on performance in the classroom provides one solution to the selection problem. For example, Staiger and Rockoff (2010) proposed a model where teachers are filtered after their first year of teaching based on their students' performance. Similarly, Hanushek (2011) suggested continuous filtering of teachers based on classroom performance. What we do know is that individual differences in teacher effectiveness exist and show some stability (e.g., Atteberry et al., 2015), and that evaluating current selection procedures and improving future procedures represents one way to improve educational quality.

1.3 Why Teacher Selection Matters

Improving the effectiveness of teachers is an international concern because it provides one of the most direct approaches to improving student outcomes (OECD, 2005). Countries with higher-performing education systems place importance on developing rigorous approaches to selection (Barber & Mourshed, 2007). In Finland, for example, selection into ITE includes evaluation of academic performance and assessment of personality and interpersonal skills using a range of interviews and tests (Sahlberg, 2021). In Singapore, selection for ITE includes an evaluation of academic

attributes such as grades and national exams, but also non-academic attributes including motivation, passion, values, and commitment to teaching (Sclafani, 2015). Even in education systems where admission into ITE programs is less competitive, there have been calls to improve selection methods, (Heinz, 2013; UK House of Commons, 2012).

Selecting the 'right' teachers can make a difference to student outcomes. The best evidence on teacher effectiveness shows that even relatively modest increases in teacher effectiveness can make a significant difference in student outcomes—and national economic growth—over time (Hanushek, 2011). Although most teachers display dramatic improvement during the first five years of teaching, the relative effectiveness ranks within cohorts remain stable over time (Atteberry et al., 2015). In other words, the most effective new teachers tend to remain in the top cohort over time, and the least effective new teachers tend to stay in the lowest cohort over time, underscoring the need for reliable and valid selection practices. Although most new teachers and most teacher candidates in ITE programs experience success, improving the teaching pipeline by improving the overall quality of teachers makes an impact on key educational and societal outcomes.

Unfortunately, not very much research has been conducted on teacher selection. The selection process represents a predictive hypothesis whereby selectors attempt to predict the future teaching effectiveness of applicants. In order to make the prediction, selectors gather evidence that they believe can help them make valid decisions. Choosing teachers based only on cognitive and background factors, such as GPA or college major has been shown to result in poor outcomes (e.g., Wayne & Youngs, 2003). Teaching is a remarkably complex undertaking that requires a combination of skills, knowledge, and personal attributes. The goal of the selection process is to choose candidates who display strong cognitive attributes such as subject knowledge, literacy and numeracy skills, knowledge of teaching practices and reasoning abilities, coupled with desirable non-cognitive attributes: the psychological characteristics such as interpersonal competence, motivation, and personality traits.

Devising methods that predict which teachers are most likely to be effective is a daunting challenge for two reasons: first, because teaching is a complex, multi-faceted job that requires a host of skills and attributes that may be influenced by a web of contextual factors (e.g., Rimm-Kaufman & Hamre, 2010); and second, because selection is a high-stakes endeavor where candidates may not be able to or wish to accurately report their non-cognitive attributes through direct measures. Current teacher selection practices suffer from an uncertain theoretical and empirical foundation, but recent research provides guidance for new evidence-based selection practices.

Research from educational psychology has examined how psychological characteristics of teachers—motivation, personality, and beliefs about teaching—are related to effective teaching practices. In organizational psychology, research shows how selection practices have changed over time, and points to current practices that predict work performance in a range of professions. Combining theory and research from educational and organizational psychology provides guidance for improving the

validity of teacher selection practices, with a resultant positive influence on students' educational outcomes.

1.4 The Costs and Benefits of Teacher Selection

Improvements to teacher selection represent a 'quick win' when considering methods for improving the workforce, with long-lasting benefits accrued from initial modest investments. There are economic costs and benefits associated with teacher selection. The direct costs of administering selection methods can be considerable. For an ITE program or school district, this might include personnel costs for selection activities (e.g., multiple personnel involved with conducting interviews, observing group activities, or scoring personal statements), fees for training those who administer commercial selection tools, and further costs associated with scoring and reports for selection instruments. Klassen and Kim (2019) recently estimated the costs of selection methods across 32 separate studies, mostly in the United States. The costs were calculated per candidate and included training costs necessary for the administration of specific tools, time costs for the staff needed to carry out selection activities, and implementation costs. Results showed that the costs associated with the selection of applicants averaged $104 (USD) per candidate, ranging from $0—when existing records such as university transcripts were used as the sole method for selection—to nearly $300 per candidate, using an assessment center method requiring multiple assessors. There was no significant relationship between the cost of selection methods and predictive validity ($r = -0.12, p = \text{ns}$); in fact, the data suggested a trend towards an *inverse relationship* between cost and effectiveness. It seems that paying more for selection methods is no guarantee that the methods are more effective.

Economic implications of teacher selection. We know that teacher effectiveness has an influence on student outcomes (e.g., Hattie, 2009), but it also has an important economic impact on students and on society. Hiring a new teacher represents a career-long investment of at least 2 million US dollars (Goldhaber et al., 2014), meaning that selecting a teacher who is less effective may represent a costly mistake. There is also a cost to individual students and to education systems with lower-performing teachers in the classroom. Research by economist Eric Hanushek (2014) shows that replacing the lowest-performing teachers with average teachers would raise U.S. educational achievement to that of higher-performing Canada. In financial terms, replacing a less effective teacher with an average teacher increases students' lifetime income by approximately US $250,000 per classroom (Chetty et al., 2014). Students taught by more effective teachers are more likely to complete high school, attend college (and attend higher-ranked colleges), and enjoy higher future salaries (Chetty et al., 2014). Another selection-related cost is teacher attrition. Teachers who leave the profession prematurely cost school systems up to US $20,000 per teacher (Carver-Thomas & Darling-Hammond, 2017). Research has shown that teachers'

1. Count the cost of selection — time costs and actual expenses
2. Selection methods matter even when recruitment, not selection, is the goal
3. Look for methods with a strong evidence base
4. Explore research on selection methods from other fields
5. New teacher selection methods can be a 'quick win' to improve the teaching workforce

Fig. 1.2 Five key points to consider for teacher selection

non-academic attributes are linked to professional commitment and quitting intention (Klassen & Chiu, 2011). Making poor selection decisions represents a cost for students, school systems, and society.

One way to improve education systems is to improve the quality of entrants into the teaching pipeline. Countries that perform at the highest level in international comparisons (e.g., Singapore, Finland, South Korea), pay the most attention to entrance into the profession, especially to issues of recruitment and selection (Ingvarson & Rowley, 2017). Improving teacher selection practices is a relatively low-cost approach to improving education systems. The selection of teachers and teacher candidates is important not just for students' academic outcomes and well-being, but for a nation's social and economic well-being. Using the best possible selection methods for entry into ITE and employment is a matter of financial accountability for education systems.

Throughout this book we will be examining the social, educational, and economic impacts of improving selection methods of teachers, and proposing new evidence-supported methods for implementation. In Fig. 1.2 we propose five key points to consider in evaluating current selection methods, and to think about when examining possible new methods. Selection programs in ITE programs and in large organizations (school districts and national education systems) have often neglected to consider the cost and effectiveness of selection; we propose that improving teacher selection methods is an important way to improve social and educational outcomes.

1.5 Purpose and Overview of the Book

The premise underlying the book is that a vital—but often overlooked—approach to improving teacher quality is to transform the selection processes that bring people into the profession. However, before improvements to teacher selection practices can be made, there must be a consideration of the key personal characteristics of effective teachers, and of the rigor of available selection methods. By exploring research from education and educational psychology, and from disciplines outside of K-12 education such as organizational psychology and medical education, we offer a theory- and research-based framework for developing and testing selection methods for teacher training and employment.

The first purpose of this book is to consider the importance of selecting effective teachers, and to look at research predicting teacher effectiveness. The second purpose of the book is to consider research on selection practices outside of education, and to see if we in education can learn from how selection practices are developed and tested outside of our field. In particular, we want to consider the contributions of research and theory from psychology and current practices in medicine and other professional fields where selection research has been more systematic than in education. The third purpose of the book is to propose a teacher selection research agenda that will build an understanding of teacher effectiveness and improve educational outcomes. Finally, the fourth purpose is to propose practical ways that selection research can be implemented to improve teacher selection practice.

Sections of the book. The book is laid out in three sections. After the present introductory chapter, we present Part I *Identifying the characteristics of effective teachers* (authored by Robert Klassen) with a focus on research and theory investigating teacher effectiveness (Chap. 2) and on research that examines the role of key attributes in teacher selection (Chap. 3). Part II *Selection methods and practices: Issues and uses* (authored by Lisa E. Kim) begins with an overview of issues and challenges in selection methods and practices (Chap. 4), and then presents an outline of the selection methods used by medical schools, law schools, and large organizations (Chap. 5). Part III *Teacher selection: Past, present, and future* (Klassen) explores historical and current selection practices in education (Chap. 6). In Chap. 7 we consider how prospective teachers' non-cognitive attributes can be evaluated using situational judgment tests (SJTs), a method that has not been widely used for screening large numbers of applicants in education settings, even though it is widely used outside of education. In Chap. 8 we present how multiple mini-interviews (MMIs), originally developed and tested for interviewing for entrance into medical education, show promise for intensive interviewing in education settings. In Chap. 9, we present an overview of the challenges of implementing teacher selection methods, and in Chap. 10, we ponder how the lessons learned in teacher selection can be applied to recruiting and developing new teachers. Chapter 11 provides a final summary of the book with some suggestions about the future direction of teacher selection work. Together these 11 chapters provide a robust foundation of research and theory supporting teacher selection practices, along with practical guidance for implementing some of the newest and most promising selection methods.

1.6 Chapter Summary

Strengthening the teacher workforce begins with improving selection methods into the profession. A close look at approaches used in other disciplines, such as medical education, and at new research and strategies recently implemented in teacher education can improve how we choose prospective teachers. Improvements in teacher selection have the potential to improve the teaching workforce and by extension, the potential to influence entire education systems and thereby countries' economic

prosperity. This book is among the first of its kind to propose evidence-supported approaches to improve selection in education. It is our sincere hope that readers will come away with a new understanding and will consider new approaches to maximize the chance of selecting the very best candidates to enter the profession.

References

Atteberry, A., Loeb, S., & Wyckoff, J. (2015). Do first impressions matter? Predicting early career teacher effectiveness. *AERA Open, 1*, 1–23.

Baker, J. (2020, May 28). Stragegy needed to attract top students as teacher shortage looms. *The Sydney Morning Herald*. Retrieved from https://www.smh.com.au/national/nsw/strategy-needed-to-attract-top-students-as-teacher-shortage-looms-20200528-p54x9m.html

Barber, M., & Mourshed, M. (2007). *How the world's best-performing schools systems come out on top*. McKinsey & Company.

Carver-Thomas, D., & Darling-Hammond, L. (2017). *Teacher turnover: Why it matters and what we can do about it.* Palo-Alto, CA: Learning Policy Institute.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review, 104*, 2593–2632.

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective teacher professional development.* Palo Alto, CA: Learning Policy Institute.

Dewey, J. (1929). *The sources of a science of education. The Kappa Delta Pi lecture series*. Horace Liveright.

Galt, V. (2017, June 16). How aspiring teachers navigate a tight labour market. *The Globe and Mail*. Retrieved from: https://www.theglobeandmail.com/report-on-business/careers/management/how-aspiring-teachers-navigate-a-tight-labour-market/article35307960/

Goldhaber, D., Grout, C., & Huntington-Klein, N. (2014). Screen twice, cut once: Assessing the predictive validity of teacher selection tools. (CEDR Working Paper No. 9). Seattle, WA: Center for Education Data and Research. Retrieved from http://www.cedr.us/papers/working/CEDR%20WP%202014-9.1.pdf

Hanushek, E. A. (2011, Summer). *Valuing teachers: How much is a good teacher worth?* Education Next, *11*(3), 41–45.

Hanushek, E. A. (2014). Boosting teacher effectiveness. *What lies ahead for America's children and their schools*, 23–35.

Hanushek, E. A., & Rivkin, S. G. (2012). The distribution of teacher quality and implications for policy. *Annual Review of Economics, 4*, 131–157.

Hattie, J. (2009). *Visible learning: A synthesis of 800+ meta-analyses on achievement*. Routledge.

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*, 93–114.

Ingvarson, L., & Rowley, G. (2017). Quality assurance in teacher education and outcomes: A study of 17 countries. *Educational Researcher, 46*, 177–193.

Jacob (2016). The power of teacher selection to improve education. *Evidence Speaks Reports, 1*, 12.

Jayaram, K., Moffit, A., & Scott, D. (2012). Breaking the habit of ineffective professional development for teachers. *McKinsey on society*, 1–12.

Klassen, R., Bardach, L., Rushby, J., & Durksen, T. L. (2021). Examining teacher recruitment strategies in England. *Journal of Education for Teaching*. https://doi.org/10.1080/02607476.2021.1876501

Klassen, R. M., & Chiu, M. M. (2011). The occupational commitment and intention to quit of practicing and pre-service teachers: Influence of self-efficacy, job stress, and teaching context. *Contemporary Educational Psychology, 36*, 114–129.

Klassen, R. M., Granger, H., & Bardach, L. (2021). Attracting prospective STEM teachers using realistic job previews: A mixed methods study. *European Journal of Teacher Education*. https://doi.org/10.1080/02619768.2021.1931110

Klassen, R. M., & Kim, L. E. (2019). Selecting teachers and prospective teachers: A meta-analysis. *Educational Research Review, 26*, 32–51.

Klassen, R. M., Kim, L. E., Rushby, J., & Bardach, L. (2020). Can we improve how we screen applicants for initial teacher education? *Teaching and Teacher Education, 87,* 102949.

Knight, F. B. (1922). Qualities related to success in elementary school teaching. *The Journal of Educational Research, 5*, 207–216.

OECD: Organisation for Economic Cooperation and Development (2005). *Teachers matter: Attracting, developing and retaining effective teachers.* Paris, France: OECD Publications. Retrieved from: http://www.oecd.org/education/school/attractingdevelopingandretainingeffectiveteachers-finalreportteachersmatter.htm

Rickenbrode, R., Drake, G., Pomerance, L., & Walsh, K. (2018). *2018 Teacher Prep Review: A review of the nation's teacher preparation programs.* National Council on Teacher Quality.

Rimm-Kaufman, S. E., & Hamre, B. K. (2010). The role of psychological and developmental science in efforts to improve teacher quality. *Teachers College Record, 112*, 2988–3023. http://www.tcrecord.org

Sahlberg, P. (2021). *Finnish lessons 3.0* (3rd ed.). Teachers College Press.

Sclafani, S. K. (2015). Singapore chooses teachers carefully. *Phi Delta Kappan, 97*(3), 8–13.

See, B. H., Morris, R., Gorard, S., & El Soufi, N. (2020). What works in attracting and retaining teachers in challenging schools and areas? *Oxford Review of Education, 46*, 678–697.

Staiger, D. O., & Rockoff, J. E. (2010). Searching for effective teachers with imperfect information. *Journal of Economic Perspectives, 24*, 97–118.

Tubbs, E. V. (1928). The selection of teachers. *Peabody Journal of Education, 5–6*, 323–332.

UK House of Commons Education Committee (2012). *Great teachers: attracting, training and retaining the best*. Government Response to the Committee's Ninth Report of Session 2010–2012. London: House of Commons.

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research, 73*, 89–122.

Part I
Identifying the Characteristics of Effective Teachers

Chapter 2
What Does 'Teacher Effectiveness' Look like?

Abstract Making selection decisions is, at its heart, a prediction about future effectiveness. For teacher training, the key predictive question is, Will this applicant succeed in our program? For teaching jobs, the central predictive question is, Will this applicant have a positive influence on student achievement and wellbeing? In this chapter, we provide a working definition of teacher effectiveness and explore theories and models of teacher effectiveness. Next, we consider the challenges inherent in measuring teacher effectiveness, with a look at value-added approaches, classroom observations, and student ratings. Understanding how teacher effectiveness changes over time raises important implications for teacher selection, and building our knowledge about the academic, psychological, and even financial outcomes of selecting the most effective teachers is crucial to building the teacher workforce.

Making selection decisions is, at its heart, a prediction about future effectiveness. For teacher training, the key predictive question is, *Will this applicant succeed in our program?* For teaching jobs, the central predictive question is, *Will this applicant have a positive influence on student achievement and wellbeing?* There is no shortage of research showing that teachers make an important contribution to academic outcomes (e.g., Fauth et al., 2019; Hanushek, 2014; Kane et al., 2008; Rivkin et al., 2005), and we know there is considerable variation in how much individual teachers contribute to these outcomes (e.g., Atteberry et al., 2015; Xu et al., 2015). However, defining teacher effectiveness can be challenging because teaching is complex, and teacher behaviors do not necessarily influence student outcomes in a direct or linear fashion (Skourdoumbis & Gale, 2013). Furthermore, not very much is known about the trajectory of teacher effectiveness: how it can be identified in applicants, how it develops through training, and how it changes over time. In this chapter, we explain what we mean by *teacher effectiveness*, and explore its conceptualization, its development over time, how it varies from person to person, and how it might be measured in ways that are reliable and valid.

R. M. Klassen and L. E. Kim, *Teacher Selection: Evidence-Based Practices*,
https://doi.org/10.1007/978-3-030-76188-2_2

2.1 Teacher Effectiveness

We define 'teacher effectiveness' as the extent to which teachers carry out the socially-agreed objectives associated with the job, primarily, but not exclusively, pertaining to student learning (Campbell et al., 2003). The definition recognizes that teachers are evaluated first and foremost on the impact they make on students' learning, but also that other outcomes—social, professional/collegial, community-related—are secondarily associated with effectiveness. The current focus on evaluating teacher effectiveness through measuring student achievement gains or through observing specific lessons using systematic observation protocols captures important aspects of teacher effectiveness, i.e., student learning, but also misses other, more nuanced aspects of teaching.

Teacher effectiveness includes an interaction of personal characteristics and behavior; that is, who the teacher *is* (individual attributes, background factors, teaching-related experiences) and what a teacher *does* (i.e., behaviors in the classroom that include teaching and assessment strategies, ways of relating to students, peers, and the community). Our primary interest in this book is in exploring the individual attributes or personal characteristics that influence t*eacher effectiveness*, rather than examining teaching behaviors that comprise the broader category of *effective teaching.* There is no shortage of research on effective teaching practices (e.g., the RAND report on teacher effectiveness; Stecher et al., 2018), but researchers in education and psychology have paid less attention to studying the personal characteristics of effective teachers. We recognize the critical importance of studying the behaviors that underpin effective teaching—the strategies used, the assessment approaches adopted, and the preparation and planning supporting effective teaching—but our key interest is in understanding the general and particular characteristics that lead to these teaching behaviors, especially when we consider how important teacher effectiveness is to student outcomes.

Hattie's meta-analysis on student achievement outcomes. How important is teacher effectiveness to student outcomes? Hattie's synthesis of over 800 meta-analyses related to students' academic achievement (Hattie, 2009; also see Hattie & Zierer, 2019) summarizes the relative contributions from the student, from the home, the school, and from the curriculum. His 'barometer of influences' rates the relative influence of the major contributors to learning, with four categories: reverse effects, describing interventions in which the students actually *lose* progress ($d = -0.20$ to -0.10), developmental effects, where the factor does not make much more of an impact than the expected maturational improvement ($d = 0.0$ to 0.15), teacher effects ($d = 0.15$ to $d = 0.40$) where the effect is similar to the education gains typically accomplished with a teacher in a school year, and zone of desired effects ($d > 0.40$), for the influences that have the greatest impact on student learning.

Table 2.1 presents the ranking of average effects from each of the major contributors to learning, with 'teacher' factors ranked first, with a mean effect size of $d = 0.49$, followed by effects from the curricula, from teaching (i.e., teaching practices), and lesser effects from student, home, and school factors. Many of the key

Table 2.1 Average effects for contributions to student achievement (adapted from Hattie, 2009)

Contribution	# of meta-analyses	*d*	*SE*
Student	139	0.40	0.044
Home	36	0.31	0.058
School	101	0.23	0.072
Teacher	**31**	**0.49**	**0.049**
Curricula	144	0.45	0.076
Teaching	365	0.42	0.071
Average	136	0.40	0.062

Note For educational outcomes, effect sizes can be classified as small ($d = 0.20$), medium ($d = 0.40$), and large ($d \geq 0.60$)

within-teacher variables included in Hattie's analyses include individual attributes such as the expectations teachers hold about their students' academic potential, with other variables focusing on teaching behaviors in the classroom. Student achievement results from a complex interaction of environmental factors, within-person factors, and behavioral factors.

The '*What makes great teaching?*' report. A 2014 report focused on teaching effectiveness—*What makes great teaching?*—provides one way to understand factors related to successful teaching. Coe et al. (2014) defined teaching effectiveness as teaching which leads to "improved student achievement using outcomes that matter to their future successes" (p. 2). The authors of the review focused primarily on factors related to teachers and teaching, with an emphasis on classroom factors associated with measurable student achievement. Six evidence-supported components of effectiveness were included in their general framework for teaching quality: pedagogical content knowledge, quality of instruction, classroom climate, classroom management, teacher beliefs, and professional behaviors. The authors found *strong evidence* of impact on student outcomes for (a) pedagogical content knowledge and (b) quality of instruction, with *moderate evidence* of effectiveness for (c) classroom climate and (d) classroom management, and *some evidence* of effectiveness for (e) teacher beliefs, and (f) professional behaviors. Although it is difficult to separate teachers' individual attributes from their teaching practices, the Sutton Trust review suggests that a wide range of factors are related to effective teaching.

Kunter's COACTIV model. Teacher effectiveness is dynamic, because it changes over time as teachers gain experience and learn new approaches to engaging with students. But teachers vary widely in their effectiveness (Atteberry et al., 2015), and these inter-individual differences are influenced by the interaction between individual attributes and external factors. The COACTIV model of teacher effectiveness is built on a *dynamic interactionist view* in which individual attributes interact with contextual and background factors to influence student outcomes (Kunter et al., 2013). This view of teacher effectiveness recognizes that inter-individual differences in teacher effectiveness may be related to individual attributes or to background or contextual factors. Teacher effectiveness is formed through the interaction

between individual attributes, relevant experiences, and learning opportunities. In Kunter et al.'s model, learning opportunities include informal (learning by doing) opportunities as well as formal activities, such as those presented in initial teacher training and professional development. Their work is built on the notion that teachers' competence exists as a continuum (e.g., Krauss et al., 2020) and develops over time, influenced by individual attributes that are evident at entry into training and practice (Kunter et al., 2013).

Figure 2.1 presents an adapted version of the COACTIV model. In the model, the broad educational and social environment (contextual factors) have an overarching influence on all aspects of teaching and learning through its relationship with learning opportunities, teacher effectiveness (comprised of professional competence and practice), and student and teacher outcomes. Teachers' individual attributes include those that are malleable and likely to change over time (e.g., pedagogical knowledge) and those that are more trait-like and resistant to change (e.g., personality). These individual attributes provide a foundation that does not just influence professional competence and professional practice, but also influences how teachers engage in available learning opportunities. The adaptation of Kunter et al.'s model provides a theoretical explanation of variations in teacher effectiveness, by noting how individual attributes influence teacher effectiveness, resulting in differential effects on student and teacher outcomes.

Individual differences in teacher effectiveness. The COACTIV model acknowledges the ways in which individual attributes contribute to variation in teacher effectiveness. The evidence for variation in teacher effectiveness is strong, yet school systems are often reluctant to publicly acknowledge variation in teachers' effectiveness (Paufler & Sloat, 2020; Weisberg et al., 2009). The pattern of argument

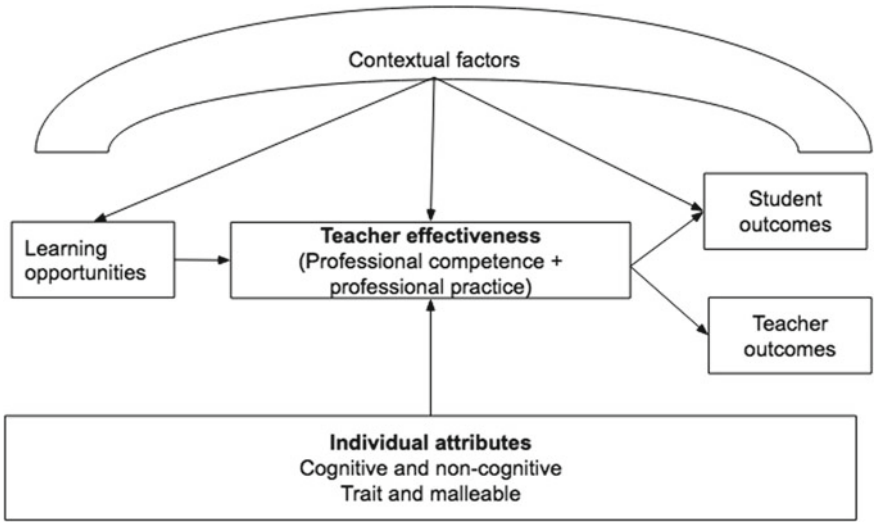


Fig. 2.1 Model of teacher effectiveness (adapted from Kunter et al., 2013)

between those protecting teachers' interests and those operating school systems is well known, with one side arguing that arbitrary and unreliable measurements of effectiveness threaten teachers with potentially arbitrary and discriminatory employment practices, and the other side arguing that accountability is needed to identify cases of good and exceptionally poor practice. The argument that teachers are all equally effective was termed by Weisberg et al. (2009) the *Widget Effect*, defined as the belief that teachers in a system function as identical, interchangeable parts, with no difference in instructional effectiveness. In their study of 12 American school districts in four states, representing approximately 15,000 teachers, they found that almost all teachers were rated as either *good* or *great*, excellence was unrecognized, and poor performance was mostly unaddressed. In almost all school districts, no meaningful information on teachers' strengths and weaknesses was collected, and the effectiveness data that were collected were almost never used for selection (or retention) purposes.

Consideration of variation in teacher effectiveness is especially important when evaluating new teachers at the point of selection into training and employment. Hiring a new teacher represents a career-long investment of at least two million dollars (Goldhaber et al., 2014), meaning that selecting a less effective teacher represents a costly mistake. Typically, beginning teachers become more effective in the first few years of their careers, thus a selection process is more about predicting an effectiveness trajectory than predicting effectiveness in the first year alone. Staiger and Rockoff (2010) estimated that students in the classroom of a first-year teacher gain 0.06 to 0.08 standard deviations of achievement in mathematics and language arts less than similar students assigned to experienced teachers. A recent systematic review of the research on the relations between teacher experience and effectiveness (Podolsky et al., 2019) found that almost all studies (28/30) showed a positive and significant association between experience and effectiveness, with effectiveness rising sharply in the first few years of a career, with a continuing upward trajectory into the second (and often) third decade of teaching. However, the validity of measures of teacher effectiveness is often disputed, largely due to the complexity of the job.

2.2 Measuring Teacher Effectiveness

Measuring teacher effectiveness presents one of the greatest challenges for researchers and policymakers because teacher effectiveness depends on a complex interaction between teachers x students x subject x school. It can be difficult to reliably measure one factor in the interaction. In 1917, Pittenger, writing in the *Journal of Educational Psychology,* spoke of *teacher measurement*, referring to the development of a consistent manner of measuring the "qualities of teaching merit" (p. 103). Pittinger recognized that "there are those who believe that the movement toward teacher measurement is a monstrous innovation, which threatens the holiest traditions of the education profession" (p. 103). But measuring teacher effectiveness can play an important role in sustaining healthy school systems: understanding

how and when teachers are most effective can lead to targeted professional development offerings, improved instructional practices, and a less biased understanding of teacher effectiveness.

Assessment of teacher effectiveness can take on a range of forms, with some methods (e.g., classroom observations, value-added models, student ratings) showing stronger validity evidence than other methods (e.g., principal judgment, teacher self-reports, and analysis of teaching portfolios (Coe et al., 2014). The use of value-added models of teacher effectiveness is gaining increased exposure, but is contested, with particular concern that value-added models fail to adequately account for differences in student backgrounds (e.g., Darling-Hammond, 2015). Other approaches to measuring teaching effectiveness involve the use of systematic classroom observation tools, such as the CLASS framework (e.g., Pianta & Hamre, 2009), which assesses teaching behaviors including emotional support, classroom organization, and instructional support. Assessment of teaching behaviors using instruments such as the CLASS has been shown to be robust and of benefit for designing interventions aimed at enhancing teaching practices.

In the next section, we describe three different approaches to measuring teacher effectiveness, and although far from exhaustive, the list covers some of the most-researched and discussed approaches.

Value-added approaches. The idea of using value-added measures of teacher effectiveness reduces a complex calculus of interactions to a simple-to-understand equation: teacher effectiveness = gains in student achievement. A value-added approach evaluates the impact of teachers on students' standardized test scores. Increasingly popular—and controversial—especially in the United States, value-added approaches control for relevant student factors such as prior test scores and demographics, and purport to provide an unbiased measure of the causal impact of teacher effectiveness. The approach is conceptually appealing: if teacher effectiveness can be reliably and validly separated from other environmental influences on student learning, then it is possible to identify the teachers that are having the most (and least) impact in the classroom. The work of influential American educational economists such as Raj Chetty, Jonah Rockoff, Thomas Kane, and Douglas Staiger have been influential in establishing the prominence of value-added approaches for teacher evaluation. Value-added scores are used to rate and rank teachers, and to make personnel and funding decisions.

Not surprisingly, value-added methods for teacher evaluation have met with strong opposition from those closely aligned with the profession. Although it is widely acknowledged that value-added methods are preferred over static measures of student learning outcomes because they capture change over time, the use of value-added methods to evaluate relative teacher effectiveness has been a cause for concern (American Educational Research Association, 2015). The approach is based on a set of assumptions that are frequently violated: (a) that student achievement is measured in a reliable way by standardized tests, (b) that individual teachers are the key contributors to students' learning over the time period measured, and (c) that students are randomly assigned to teachers in and across schools (Darling-Hammond, 2015). Critics raise questions about the stability of value-added scores:

teachers show a high level of annual fluctuation in value-added performance, with half of teachers in the bottom 20% of rankings in one year scoring in the top half in the following year (Darling-Hammond, 2015). Although value-added approaches are intuitively appealing, other approaches using observational approaches have gained in popularity as a way to measure teacher effectiveness.

Classroom observations. One of the most widely implemented approaches to measuring teacher effectiveness is observation of teachers in the classroom (Coe et al., 2014). Unstructured observations are regularly used by school principals to monitor the quality of instruction being delivered in a school. More formal classroom observation systems are used to make judgments about teacher effectiveness. One of the most widely used classroom observation tools is the CLASS (Classroom Assessment Scoring System) developed by Pianta and Hamre (2009). The CLASS, administered by trained observers, measures three factors of classroom teaching: emotional support (classroom climate, teacher sensitivity, and empathy for student perspectives), classroom organization (behavior management, productivity, and instructional organization), and instructional support (quality of classroom feedback, concept development, and communication). The CLASS instrument, unlike other observation protocols, differentiates between primary, middle school, and secondary school contexts, with different versions of the instrument for each context, with stronger evidence for predictive validity in the early years (e.g., Sandilos et al., 2019).

Another well-validated observation system is Danielson's *Framework for Teaching* (FfT; Danielson, 2007). The framework evaluates four aspects of effective teaching: planning and preparation, classroom environment, instruction, and professional responsibilities. The FfT does not provide explicit observation protocols in the same way as the CLASS measure, but rather offers a categorization of teaching practices that are deemed to be supportive of effective teaching. Unlike the CLASS instrument, the FfT does not differentiate between different levels of teaching (e.g., primary and secondary). Studies examining the validity of the FfT show mixed results, with some studies showing positive correlations with student achievement gains (e.g. Gallagher, 2004) but other studies showing more equivocal results (e.g., Kimball et al., 2004; Sandilos et al., 2019).

Recent work in the Netherlands has focused on developing a teacher observation method—the International Comparative Analysis of Learning and Teaching (ICALT)—that includes six domains of teaching behaviors: safe learning climate, classroom management, clear instruction, activating teaching methods, learning strategies, and differentiation (van der Lans et al., 2017). The ICALT offers a unique perspective among teacher effectiveness measures; it is situated in Fuller's (1969) three-stage theory of teacher concerns which proposes that teachers proceed through developmental stages as they progress through their career. In the first stage, according to Fuller, teachers are primarily concerned with the self; secondly, teachers are concerned with the 'tasks' of teaching; and finally, teachers are concerned with the impact on student learning. Under the umbrella of this developmental perspective, the van der Lans et al. study used Rasch modeling to show that the ICALT content was ordered in a way that was congruent with Fuller's stage model of teacher development.

Classroom observations are subject to several methodological problems, even when systematic approaches are implemented. Pre-existing beliefs about the teacher or teaching methods can bias an observer's perspective, with *halo effects* (the tendency for overall impressions of a person to influence observation ratings) and other biases potentially influencing observation scores. Furthermore, the reliability and generalizability of observations can be suspect if based on a modest number of observations (Muijs, 2006).

Student ratings. Most teacher observation systems use 'expert' observers to carry out ratings of teacher behaviors, but another approach is to use a different kind of 'experts'; that is, students in the classroom. Researchers in the Measures of Effective Teaching (MET) project, funded by the Bill and Melinda Gates Foundation, used the Tripod survey instrument (developed by Ferguson), which assesses students' perceptions of the classroom environment (e.g., Ferguson, 2009). The Tripod surveys consist of 36 items divided into seven categories (7Cs): Care, Control, Clarify, Challenge, Captivate, Confer, and Consolidate. Table 2.2 provides example items from the elementary (i.e., primary) version of the survey.

Reliability within each of the categories was strong (in the range of 0.80), and validity, measured as the relationship with teacher value-added measures, was significant. The relationship with teacher effectiveness measures varied according to category. In the validation study, the Tripod categories that were most strongly correlated with student achievement gains in English and mathematics were 'control' and 'challenge', with raw correlations of 0.22, and disattenuated correlations (i.e., corrected for measurement error) around 0.40 (Kane & Cantrell, 2010). A recent factor analysis of the Tripod (Wallace et al., 2016) found little support for the stability of the proposed seven factors but did find that student ratings of teacher behavior and the classroom environment were associated with teachers' value-added scores. Measures other than the Tripod have been developed and used in other contexts, with, for example, Kyriakides' (2005) student rating protocol showing significant correlations with student achievement gains in Cyprus.

Measuring teacher effectiveness can be conducted in a range of ways, but classroom observation, value-added models, and student ratings have a stronger evidence base than other approaches, such as principal judgments, teacher self-reports, or

Table 2.2 Examples of items from the Tripod measure of student perceptions of teacher effectiveness (elementary version)

Category	Example Item
Care	*I like the way my teacher treats me when I need help*
Control	*Our class stays busy and does not waste time*
Clarify	*My teacher explains difficult things clearly*
Challenge	*My teacher pushes everybody to work hard*
Captivate	*School work is interesting*
Confer	*My teacher wants us to share our thoughts*
Consolidate	*My teacher takes the time to summarize what we learn each day*

analysis of teacher portfolios (Coe et al., 2014). Using a triangulation approach with multiple evidence-supported measures provides the best chance of accurately capturing teacher effectiveness.

2.3 Trajectories of Teacher Effectiveness

Researchers have posited that teacher effectiveness tends to improve with experience, but only to a point (Rockoff et al., 2011), with new teachers becoming more effective as they gain experience, but with the 'experience effect' declining after the first few years of teaching (Hanushek, 2014). Recent reviews of the research challenge this truncated teacher growth theory, with Podolsky et al. (2019) proposing that effectiveness continues to increase into the second and third decades of teaching experience. Jackson et al. (2014) considered teaching effectiveness as the ability to increase students' "stock of human capital" (p. 802) through teaching behaviors such as communication with students, classroom management, or encouragement of greater efforts.

Although many new teachers gain effectiveness over the first few years of their careers, others succumb to the 'reality shock' phenomenon experienced during the first one or two years of teaching, and leave the profession (Ingersoll, 2001). Although there are many causes of new teachers' reality shock—socialization into the profession, unexpectedly heavy workload, difficulties with teacher-student interactions—being unprepared to manage classroom disturbances is a major cause of the phenomenon (Dicke et al., 2015). Teachers may overcome the initial shock of facing classroom realities through a combination of targeted interventions (e.g., Dicke et al.), or through increases in expertise that come with classroom experience.

Most teachers increase in effectiveness over time, but research that follows the trajectories of beginning teachers shows that *relative* effectiveness may be stable; that is, new teachers' effectiveness can vary substantially. Predicting heterogeneity in teacher effectiveness is at the heart of the selection process because it represents an attempt to predict which teachers will show the highest, and most stable levels of improvement in effectiveness, especially at the beginning of a teaching career. Uncovering the within-teacher factors that lead to teacher effectiveness is at the heart of the teacher selection process.

Atteberry's work on effectiveness within large cohorts of new teachers shows that *relative* effectiveness is stable (Atteberry et al., 2015); that is, new teachers' initial effectiveness is predictive of future effectiveness, especially for those who initially display the highest and lowest levels of effectiveness. The researchers collected value-added student data in mathematics and English language arts from the classrooms of over 3000 teachers in New York during the first five years of their careers. After dividing the sample into quintiles of initial performance, the researchers compared the performance of teachers at each quintile over the next five years.

The key finding from the study was that, on average, initial job performance measured after the first year of teaching predicted teacher effectiveness in years 2–5,

and that the effect was more predictive than education or SAT (university entrance) scores. On average, the most effective teachers in the first year retained their effectiveness relative to their peers over time; the least effective teachers tended to stay in that group over time. The effectiveness trajectory of low- and high-performing teachers over time was not perfect—some of the lower performing teachers became higher performing, and some higher performing teachers became lower performing—but the pattern of consistency of effectiveness was stable for the group overall. Atteberry et al. (2015) concluded that accurately identifying the effectiveness of early career teachers had the potential to dramatically improve educational outcomes for students.

The finding of stable patterns of teacher effectiveness—with lower effectiveness and higher effectiveness teachers tending to show stable rank ordering over time—is not unique to Atteberry's study. Xu and colleagues (Xu et al., 2015) measured teacher performance trajectories in high- and low-poverty school settings in mathematics. Teacher effectiveness levels improved the fastest at the beginning of teachers' careers (i.e., 0–5 years), plateaued at 6–10 years, and resumed growing at 10–15 years of experience. The authors found that the fastest growing teachers improved significantly faster annually than slower growing teachers; that is, the students of novice teachers with initial low effectiveness showed annual lower achievement growth than students in the classes of higher effectiveness teachers. Teachers who were initially in the top effectiveness quartile tended to show a faster rate of improvement than teachers in the lower performing quartile; in fact, as much as 80% faster growth rate than their slower improving peers. About half of the total variation in teachers' performance was found within teachers, with about one-quarter to one-third explained by increasing experience, and the remainder by classroom and school level characteristics.

Trajectories of motivation profiles. It is not only teachers' effects on student achievement that shows stability, but teachers' motivation patterns also show stability over time. Watt and Richardson measured the motivation of pre-service teachers during their teacher training programs in Australia (Watt & Richardson, 2008). Using cluster analysis, the researchers found a sizable proportion of participants with low motivation, so-called 'lower engaged desisters,' who showed little change in motivation profiles over the course of the teacher training program. This low motivation group of pre-service teachers were disaffected with teaching as a career from the beginning of their training. A follow up study in the United States (Watt et al., 2014) that traced the motivation profiles over the course of teacher training resulted in similar findings. The authors concluded that the findings of a stable profile of low motivation pre-service teachers indicate a need to closely examine the process for the recruitment of teacher training candidates. The implications for selection are clear: *who* you select into teacher education matters, and selection decisions have long-term effects on teaching outcomes.

2.4 Teacher Effectiveness and Related Outcomes

Teachers influence students' academic achievement and social development, but they also influence factors related to learning outcomes, such as motivation and emotions. Much of the research on longer-term outcomes of teacher effectiveness comes from studies of large-scale databases by educational economists, rather than educational psychologists, who have suggested that teachers can affect more distal outcomes, like salary in adulthood.

Achievement outcomes. The contribution of teacher effectiveness to students' academic achievement is well documented. Teacher effectiveness is multi-faceted, and some effectiveness factors are stronger predictors of achievement outcomes than others. For example, Rockoff and colleagues (Rockoff et al., 2011) found that measuring a broad range of teacher characteristics, including cognitive and non-cognitive variables, noticeably increased the accuracy of the prediction of student achievement outcomes. Teachers have a systematic and measurable effect on students' achievement outcomes, at least when the outcomes are measured with standardized tests (Jackson et al., 2014). Test scores are the most frequently used measure of student outcomes. However, other student outcomes may also be important to understand the consequences of teacher effectiveness, including career aspirations, motivation profiles, and long-term financial outlooks.

Student motivation and emotions outcomes. We know that teacher effectiveness is associated with improved student learning, but the process through which teachers influence student outcomes is worth exploring. Although teachers' classroom practices (e.g., instructional strategies) represent one pathway influencing student learning, another pathway is through transmission of motivation and emotions. In this way, students' motivation and emotions are influenced by teachers' motivation and emotions. Teachers influence student motivation by encouraging students' persistence, effort, and resilience when obstacles are encountered or when success is elusive (Anderman & Midgley, 1997). In Zee & Koomen's, 2016 heuristic model, teacher motivation (especially self-efficacy) is linked to the quality of classroom processes such as instructional support, classroom organization, and emotional support. These classroom processes, in turn, influence not only students' academic achievement but also their motivation, which in turn reciprocally influences teachers' engagement and motivation (Zee & Koomen, 2016).

Emotions, too, serve an important role in learning. Positive emotions, such as enthusiasm experienced during learning can spur on continued effort and lead to a satisfying learning experience. Negative emotions such as anger or anxiety can hinder progress and may result in lowered effort and achievement. Teachers' emotions are transmitted to students: a two-phase study by Frenzel and her colleagues conducted in Germany showed that teachers' enjoyment of mathematics was transmitted to students, and that the effect was mediated by the level of teacher enthusiasm (Frenzel et al., 2009). A three-wave longitudinal model confirmed the positive reciprocal links between teachers' and students' enjoyment, mediated by perceptions of each other's classroom behaviors (Frenzel et al., 2018). The impact of effective teachers on student

Table 2.3 Estimates of financial impact of variation in teacher effectiveness

Study	Financial impact
Kane and Staiger (2002)	A one SD increase in teacher effectiveness represents a lifetime earnings gain of around $330,000 to $760,000 for a class of 20 students
Hanushek and Rivkin (2012)	A teacher in the top 15% with a class of 20 yields at least $240,000 in class-level economic gain compared with an average teacher
Chetty et al. (2014)	Teacher impacts for the bottom 5% of teachers are greater than $250,000 lifetime earnings per class
Hanushek and Woessmann (2011)	Replacing the least effective 5%–8% of teachers with average teachers would bring student achievement up by 0.4 SDs, resulting in $70 trillion added GDP to US economy

motivation and emotion outcomes can be considered just as important as learning outcomes, because enhanced motivation and positive emotions can have a lasting effect on student learning.

Financial outcomes. Effective teachers play an important role in influencing financial outcomes for students (see Table 2.3 for summary of financial impact of variation in teacher effectiveness). Research by economists Hanushek and Rivkin (2012) shows that replacing low performing teachers with average teachers would raise U.S. educational achievement to that of Canada and Finland. In financial terms, replacing a less effective teacher with an average teacher increases students' lifetime income by approximately $250,000 per classroom (Chetty et al., 2014). Students taught by more effective teachers are more likely to complete high school, attend college (and attend higher-ranked colleges), and enjoy higher future salaries (Chetty et al., 2014).

Another cost associated with teacher effectiveness is attrition. Teachers who leave the profession prematurely cost school systems up to $20,000 (Carver-Thomas & Darling-Hammond, 2017). Research has shown that teachers' individual attributes are linked to professional commitment and quitting intention (Klassen & Chiu, 2011). Making bad selection decisions is costly for students, school systems, and society as a whole. Improving teacher effectiveness at the systems level is a relatively low-cost approach to improving education systems and boosting economic outcomes for students and for society as a whole.

2.5 Are Effective Teachers Born or Made?

Teacher effectiveness influences multiple outcomes, but is it something that is innate in prospective teachers? In education, the debate about individual differences in teacher effectiveness has been hotly contested. On the one side of the debate, some researchers endorse the 'qualification hypothesis,' whereby teacher education and

professional development represent the most (or only) important source of influence on teaching effectiveness (see Kunter et al., 2013). In this view, the individual attributes of applicants are not too important, because high quality teacher education can 'fill in the gaps' in prospective teachers. On the other side of the argument, the 'good teacher,' 'born teacher,' or 'individual aptitude' hypothesis puts forward the notion that variations in success in teaching are due to specific and stable within-person attributes that teachers and prospective teachers bring into the classroom (Kennedy et al., 2008). These personal attributes vary among individuals, thus the identification of individuals with a particular set of characteristics is important when selecting for training and employment.

There has been strong resistance to the 'born teacher' hypothesis in the popular media and by some teacher educators. For example, a Seattle Times op-ed in 2012 opened with the claim "Some people think that good teachers are born; educators know that good teachers are made. They are made over time, through education, perseverance, practice, and guidance" (Knapp, 2012). Influential educational researcher Darling-Hammond (2006) labelled the born teacher hypothesis a 'damaging myth' and a 'superstition' that resulted in policies that relied on 'some kind of prenatal alchemy' (p. ix) to identify and prepare effective teachers. Opponents to the born teacher position hold that linking stable individual attributes with teacher effectiveness weakens the importance of the role played by training and development, and suggests that teacher educators, prospective teachers, and practicing teachers can do little to improve their effectiveness beyond the constraints provided by their personal make-up.

Theory and research on the 'born teacher' debate. A number of key theories have provided a framework for the born-or-made debate. Dispositional explanations of teacher effectiveness align with an entity perspective in Dweck's (2000) entity vs. incremental model of human abilities (see Fig. 2.2). In this model, important individual attributes that influence behavior are viewed as either (a) innate and unchangeable (*entity* view), or (b) malleable and influenced by training and experience (*incremental* view). Rather than providing an explanatory model of how human behavior is either innate or learned, Dweck's model addresses the consequences for learning of adopting one of the two stances, primarily for student learning: students who believe that their own abilities are malleable, rather than fixed, tend to display higher levels of perseverance and effort. Although primarily focused on students, the incremental-entity heuristic can also be adapted to understand opposing views of the development of teacher effectiveness.

Relevant to the born-or-made debate, some individual attributes, including personality traits and attitudes, seem to be relatively stable over time, and are stable and robust predictors of occupational outcomes (e.g., Spengler et al., 2015). The long-term predictiveness of personality and other individual attributes has been explained by life course models (e.g., Shanahan et al., 2014), which show how factors such as conscientiousness are long-term and stable predictors of outcomes through the life course. The relevance to teacher selection is clear: some measurable attributes seem to be stable and significantly related to important occupational outcomes. However, the powerful impact of effective teacher training and professional development cannot

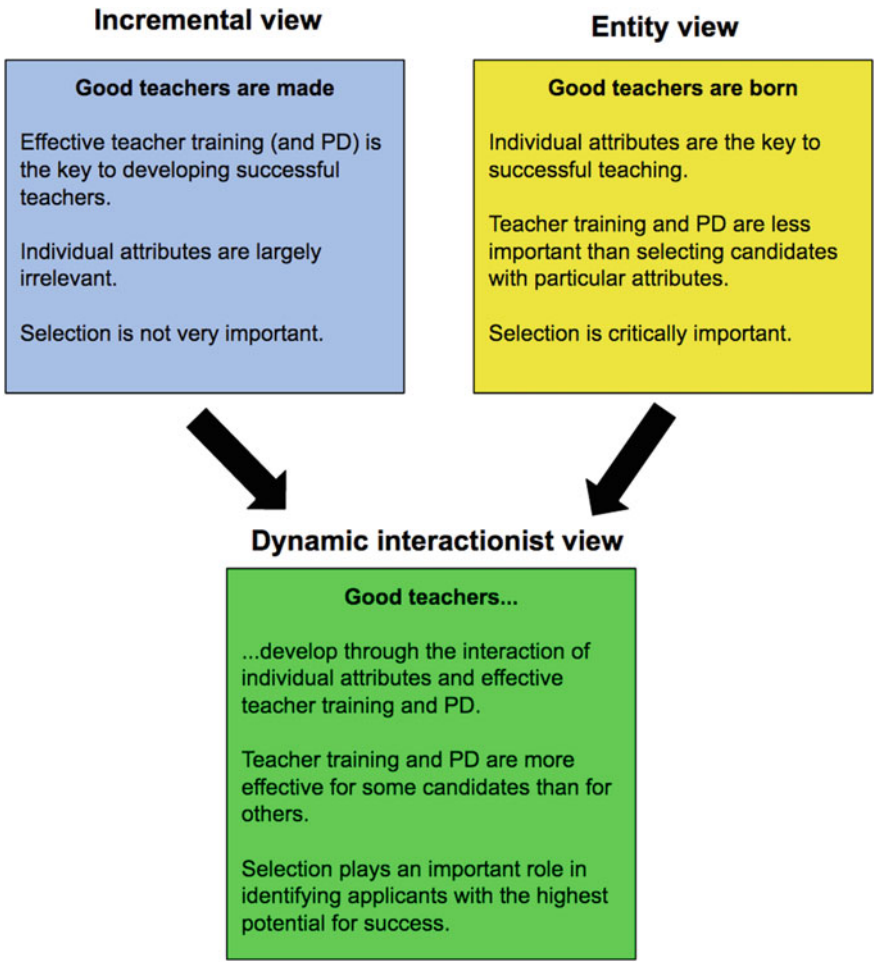


Fig. 2.2 Teacher selection and the 'Are teachers born or made?' debate (adapted from Klassen & Kim, 2017)

be denied: we know that teachers develop and improve over time. Our view of the born-or-made debate reflects a dynamic interactionist view in which good teachers develop through the interaction of individual attributes and high-quality professional training and development opportunities.

The born-or-made debate and teacher selection. The born-or-made debate has clear implications for teacher selection. Figure 2.2 highlights the relevance of selection from three viewpoints: the incremental view, the entity view, and a dynamic interactionist view. For those with an incremental view, selection is not very important, since key attributes and skills can be developed through effective teacher training and professional development. Many people involved in teacher education hold the view

that appropriate learning opportunities are the key to the development of effective teachers, i.e., an incremental view. However, for those with an entity view, selection is *everything*, because individual attributes are resistant to change, even with effective training. In the entity view, teacher training and PD are less important than choosing teachers with 'the right stuff.' Our interactionist view of teacher effectiveness is influenced by the arguments from the incremental view of human abilities whereby training and professional development improves teacher effectiveness, but also by an entity view where individual attributes—sometimes resistant to change—play an important role in influencing positive outcomes.

2.6 Chapter Summary

The purpose of this chapter was to consider how understanding teacher effectiveness is core to understanding teacher selection. We defined teacher effectiveness and examined how individual attributes relate to effective teaching. Teacher effectiveness makes an important difference for student outcomes, including academic, motivation, and even future financial outcomes. We also considered the 'born-or-made' debate, with consideration of incremental and entity views of teacher effectiveness, before settling on a dynamic interactionist view. In the next chapter we delve into research that explores how individual characteristics are related to teacher effectiveness, setting the stage for the consideration of what ITE program directors, school principals, and education authorities might look for in their quest for selecting the best possible prospective teachers.

References

American Educational Research Association. (2015). AERA statement on use of value-added models (VAM) for the evaluation of educators and educator preparation programs. *Educational Researcher, 44*, 448–452.

Anderman, E. M., & Midgley, C. (1997). Changes in achievement goal orientations, perceived competence, and grades across the transition to middle-level schools. *Contemporary Educational Psychology, 22*, 269–298.

Atteberry, A., Loeb, S., & Wyckoff, J. (2015). Do First Impressions Matter? Predicting Early Career Teacher Effectiveness. *AERA Open, 1*, 1–23.

Campbell, R. J., Kyriakides, L., Muijs, R. D., & Robinson, W. (2003). Differential teacher effectiveness: Towards a model for research and teacher appraisal. *Oxford Review of Education, 29*, 347–362.

Carver-Thomas, D. & Darling-Hammond, L. (2017). *Teacher turnover: Why it matters and what we can do about it.* Palo Alto, CA: Learning Policy Institute.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review, 104*, 2593–2632.

Coe, R., Aloisi, C., Higgins, S., & Major, L. E. (2014). *What makes great teaching? A review of the underpinning research.* Sutton Trust.

Danielson, C. (2007). Enhancing professional practice: A framework for teaching. ASCD.

Darling-Hammond, L. (2006). *Powerful teacher education*. Jossey-Bass.
Darling-Hammond, L. (2015). Can valued added add value to teacher evaluation? *Educational Researcher, 44*, 132–137.
Dicke, T., Elling, J., Schmeck, A., & Leutner, D. (2015). Reducing reality shock: The effects of classroom management skills training on beginning teachers. *Teaching and Teacher Education, 48*, 1–12.
Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality and development*. Psychology Press.
Fauth, B., Decristan, J., Decker, A. T., Büttner, G., Hardy, I., Klieme, E., Kunter, M. (2019). The effects of teacher competence on student outcomes in elementary science education: The mediating role of teaching quality. *Teaching and Teacher Education, 86,* 102882.
Ferguson, R. (2009). *Tripod student survey, MET project upper elementary and MET project secondary versions.* Distributed by Cambridge Education, Westwood, MA.
Frenzel, A., & C., Becker-Kurz, B., Pekrun, R., Goetz, T., & Lüdtke, O. (2018). Emotion transmission in the classroom revisited: A reciprocal effects model of teacher and student enjoyment. *Journal of Educational Psychology, 110*, 628–639.
Frenzel, A. C., Goetz, T., Lüdtke, O., Pekrun, R., & Sutton, R. E. (2009). Emotional transmission in the classroom: Exploring the relationship between teacher and student enjoyment. *Journal of Educational Psychology, 101*, 705–716.
Fuller, F. (1969). Concerns of teachers: A developmental conceptualization. *American Educational Research Journal, 6*, 207–226.
Gallagher, H. A. (2004). Vaughn Elementary's innovative teacher evaluation system: Are teacher evaluation scores related to growth in student achievement? *Peabody Journal of Education, 79*(4), 79–107.
Goldhaber, D., Grout, C., and Huntington-Klein, N. (2014). *It's selective but is it effective? Exploring the predictive validity of teacher selection tools.* CEDR Policy Brief 2014–9. University of Washington, Seattle, WA.
Hanushek, E. A. (2014). Boosting teacher effectiveness. *What lies ahead for America's children and their schools*, 23–35.
Hanushek, E. A., & Rivkin, S. G. (2012). The distribution of teacher quality and implications for policy. *Annual Review of Economics, 4*, 131–157.
Hanushek, E. A., & Woessmann, L. (2011). The economics of international differences in educational achievement. In *Handbook of the Economics of Education* (Vol. 3, pp. 89–200). Elsevier.
Hattie, J. A. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.
Hattie, J., & Zierer, K. (2019). *Visible Learning Insights*. Routledge. https://doi.org/10.4324/9781351002226
Ingersoll, R. M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal, 38*, 499–534.
Jackson, C. K., Rockoff, J. E., & Staiger, D. O. (2014). Teacher effects and teacher-related policies. *Annual Review of Economics, 6*, 801–825.
James, W. (1899/2015) *Talks to teachers.* Retrieved from: http://www.gutenberg.org/ebooks/16287?msg=welcome_stranger#I__PSYCHOLOGY_AND_THE_TEACHING_ART
Kane, T., & Cantrell, S. (2010). Learning about teaching: Initial findings from the measures of effective teaching project. *MET Project Research Paper, Bill & Melinda Gates Foundation*, *9*.
Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review, 27*, 615–631.
Kane, T. J., & Staiger, D. O. (2002). The promise and pitfalls of using imprecise school accountability measures. *The Journal of Economic Perspectives, 16*, 91–114.
Kennedy, M. M., Ahn, S., & Choi, J. (2008). The value added by teacher education. In M. Cochran-Smith, S. Feiman-Nemser, D. J. McIntyre, & K. E. Demers (Eds.), *Handbook of research on teacher education* (3rd ed., pp. 1249–1273). Routledge.

Kimball, S. M., White, B., Milanowski, A. T., & Borman, G. (2004). Examining the relationship between teacher evaluation and student assessment results in Washoe County. *Peabody Journal of Education, 79*(4), 54–78.

Klassen, R. M., & Chiu, M. M. (2011). The occupational commitment and intention to quit of practicing and pre-service teachers: Influence of self-efficacy, job stress, and teaching context. *Contemporary Educational Psychology, 36*, 114–129.

Klassen, R. M., & Kim, L. E. (2017). Assessing critical attributes of prospective teachers: Implications for selection into initial teacher education programs. In D. W. Putwain & K. Smart (Eds.), *British Journal of Educational Psychology Monograph Series II: Psychological Aspects of Education* (pp. 5–22). Wiley.

Knapp, J. (2012, September 11). Good teachers are made, not born. *Seattle Times* http://www.seattletimes.com/opinion/op-ed-good-teachers-are-made-not-born/

Krauss, S., Bruckmaier, G., Lindl, A., Hilbert, S., Binder, K., Steib, N., & Blum, W. (2020). Competence as a continuum in the COACTIV study: The "cascade model." *ZDM, 52*, 311–327.

Kunter, M., Kleickmann, T., Klusmann, U., & Richter, D. (2013). 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). Springer, US.

Kyriakides, L. (2005). Drawing from teacher effectiveness research and research into teacher interpersonal behaviour to establish a teacher evaluation system: A study on the use of student ratings to evaluate teacher behaviour. *The Journal of Classroom Interaction, 40*, 44–66.

Muijs, D. (2006). Measuring teacher effectiveness: Some methodological reflections. *Educational Research and Evaluation, 12*, 53–74.

Paufler, N. A., & Sloat, E. F. (2020). Using standards to evaluate accountability policy in context: School administrator and teacher perceptions of a teacher evaluation system. *Studies in Educational Evaluation, 64,* 100806.

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

Pittenger, B. F. (1917). Problems of teacher measurement. *Journal of Educational Psychology, 8*, 103–110.

Podolsky, A., Kini, T., & Darling-Hammond, L. (2019). Does teaching experience increase teacher effectiveness? A review of US research. *Journal of Professional Capital and Community, 4*, 286–308.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*, 417–458.

Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2011). Can you recognize an effective teacher when you recruit one? *Education, 6*, 43–74.

Sandilos, L. E., Sims, W. A., Norwalk, K. E., & Reddy, L. A. (2019). Converging on quality: Examining multiple measures of teaching effectiveness. *Journal of School Psychology, 74*, 10–28.

Shanahan, M. J., Hill, P. L., Roberts, B. W., Eccles, J., & Friedman, H. S. (2014). Conscientiousness, health, and aging: The life course of personality model. *Developmental Psychology, 50*(5), 1407–1425.

Skourdoumbis, A., & Gale, T. (2013). Classroom teacher effectiveness research: A conceptual critique. *British Educational Research Journal, 39*, 892–906.

Spengler, M., Brunner, M., Damian, R. I., Lüdtke, O., Martin, R., & Roberts, B. W. (2015). Student characteristics and behaviors at age 12 predict occupational success 40 years later over and above childhood IQ and parental socioeconomic status. *Developmental Psychology, 51*, 1329–1340.

Staiger, D. O., & Rockoff, J. E. (2010). Searching for effective teachers with imperfect information. *Journal of Economic Perspectives, 24*, 97–118. https://doi.org/10.1257/jep.24.3.97

Stecher, B. M., Holtzman, D. J., Garet, M. S., Hamilton, L. S., Engberg, J., Steiner, E. D., … Chambers, J. (2018). *Improving teaching effectiveness: Final Report.* Santa Monica, CA: RAND Education.

van der Lans, R., van de Grift, W. J., & van Veen, K. (2017). Developing an instrument for teacher feedback: Using the Rasch model to explore teachers' development of effective teaching strategies and behaviors. *The Journal of Experimental Education. Online First:* https://doi.org/10.1080/00220973.2016.1268086

Wallace, T. L., Kelcey, B., & Ruzek, E. (2016). What can student perception surveys tell us about teaching? Empirically testing the underlying structure of the Tripod Student Perception survey. *American Educational Research Journal, 53*, 1834–1868.

Watt, H. M. G., & Richardson, P. W. (2008). Motivations, perceptions, and aspirations concerning teaching as a career for different types of beginning teachers. *Learning and Instruction, 18*, 408–428.

Watt, H. M. G., Richardson, P. W., & Wilkins, K. (2014). Profiles of professional engagement and career development aspirations among USA preservice teachers. *International Journal of Educational Research, 65*, 23–40.

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness. *New Teacher Project.*

Xu, Z., Ozek, U., & Hansen, M. (2015). Teacher performance trajectories in high- and lower-poverty schools. *Educational Evaluation and Policy Analysis, 37*, 458–477.

Zee, M., & Koomen, H. M. Y. (2016). Teacher self-efficacy and its effects on classroom processes, student academic adjustment, and teacher well-being: A synthesis of 40 years of research. *Review of Educational Research, 86*, 981–1015.

Chapter 3
The Role of Individual Attributes in Teacher Selection

Abstract In this chapter, we explore research on the relations between teachers' individual attributes and their effectiveness, and build an understanding about how this relationship is integral to building teacher selection procedures. We begin by exploring what we mean by 'individual attributes', and particularly focus on the distinction between cognitive and non-cognitive attributes, and how these relate to teacher effectiveness. We conclude the chapter by taking an international and cross-cultural perspective through a look at what educators in a range of diverse contexts deem to be the most important non-cognitive attributes for the success of novice teachers.

Before making decisions about *who* to select, most selectors think carefully about *what* they want to evaluate when they make selection decisions. Subject knowledge? Intelligence? Interpersonal skills? Commitment to the profession? People throughout history have pondered the most important individual characteristics of effective teachers. Confucius, portrayed by Chinese historians as the *model teacher for ten thousand ages*, was renowned for his kindness, humility, and caring for others. Plato praised his teacher, Socrates, for his wisdom, courage, and moral character. In his 1899 book *Talks to Teachers,* American philosopher and psychologist William James proposed that effective teachers display "an additional endowment altogether, a happy tact and ingenuity to tell us what definite things to say and do when the pupil is before us" (1899/2015). In 1922, F.B. Knight pondered the attributes of successful graduates of teacher training institutions, noting "Of a hundred graduates… quite probably some will make excellent teachers, a larger number will do well, and a few will fail…. What qualities possessed by a candidate and ascertainable by a prospective employer are correlated highly enough with teaching success to be worth considering in a sound selective technique?" (p. 207).

These historical perspectives all share one underlying sentiment: the most effective teachers possess high levels of key attributes that set them apart from less effective teachers. The central point of teacher selection is to first identify what these attributes are, and second, to develop a reliable, valid, and fair method of evaluating these characteristics. Recent reviews exploring the links between teachers' psychological

R. M. Klassen and L. E. Kim, *Teacher Selection: Evidence-Based Practices*,
https://doi.org/10.1007/978-3-030-76188-2_3

characteristics and critical outcomes (e.g., Bardach et al., 2020) provide new insight into the most important of these key attributes for teacher selection.

In this chapter, we explore research on the relations between teachers' individual attributes and their effectiveness and build an understanding about how this relationship is integral to building teacher selection procedures. We begin by exploring what we mean by 'individual attributes', and particularly focus on the distinction between cognitive and non-cognitive attributes, and how these relate to teacher effectiveness. We conclude the chapter by taking an international and cross-cultural perspective through a look at what educators in a range of diverse contexts deem to be the most important non-cognitive attributes for the success of novice teachers.

3.1 Individual Attributes

Regardless of the methods used for selection, predicting the effectiveness of future teachers requires building an algorithm that includes applicants' background experiences, academic profile, and individual attributes. We define individual attributes as personal characteristics that include *cognitive* domains (e.g., subject area knowledge, pedagogical knowledge, reasoning ability) and *non-cognitive* domains (e.g., motivation, interpersonal skills, personality, beliefs, attitudes, and dispositions). Some individual attributes show considerable change over the course of teacher training and practice (e.g., pedagogical knowledge) while other attributes, such as personality, remain relatively stable. For example, Klassen and Durksen (2014) showed that preservice teachers' self-efficacy showed a robust increase during their major teaching practicum as they progressed through the practicum, while their levels of teaching stress decreased. Teachers' self-efficacy is theorized as a dynamic motivation construct that is influenced by past experiences, verbal persuasion, social modelling, and interpretation of physiological states. Other individual attributes are more stable, with personality traits, such as conscientiousness and agreeableness, less susceptible to change over time; it is these more fixed attributes that selectors may want to focus on at the point of selection.

What attributes do selectors look for? Selection panels for teacher training and principals hiring for teaching jobs aim to choose candidates who possess particular *background factors*, which may include a degree in a relevant subject, a teaching qualification, and relevant experiences; strong *cognitive* attributes (i.e., cognitive abilities, subject knowledge and expertise, literacy and numeracy skills, knowledge of teaching practices, communication skills); and a profile showing desired *non-cognitive* attributes, which may include a cluster of motivation, personality, resilience, commitment, and engagement factors. In Chap. 2 we showed how a dynamic interactionist view combines elements from an incremental view (*good teachers develop with appropriate teacher training and professional development*) and from an entity view (*good teachers seem to be born with certain personal characteristics that influence teaching success*). Deciding about the weighting of

background, cognitive, and non-cognitive factors is an important part of setting up selection processes.

When selectors have to make decisions about who to choose for initial teacher education (ITE) programs or for teaching jobs, they make choices about the weighting of the background factors, cognitive attributes, and non-cognitive attributes, balanced against teacher supply and demand. In some cases, principals or school district human resources personnel may make selection decisions based on subject shortages. In the United States, over half of all school districts and up to 90% of the most-deprived school districts report difficulties recruiting and retaining science teachers (Goldhaber et al., 2015). In these cases, and in other cases where supply is low, the background factor of a candidate's degree in a relevant science subject might trump cognitive and non-cognitive attributes. In situations where the supply is high, as has been the case in some parts of Canada, where half of graduating teachers face five or more years of job searching before landing a first permanent job (Brock & Ryan, 2016), employers will pay more attention to cognitive and non-cognitive attributes when the pool of prospective teachers with similar background factors is deep. In Fig. 3.1, we outline a model of how cognitive attributes, background factors, and non-cognitive attributes contribute to teaching behaviors. In our model, no weightings are given for each contributing factor; in specific contexts, the relative weightings are dependent on localized factors such as supply and demand of teachers.

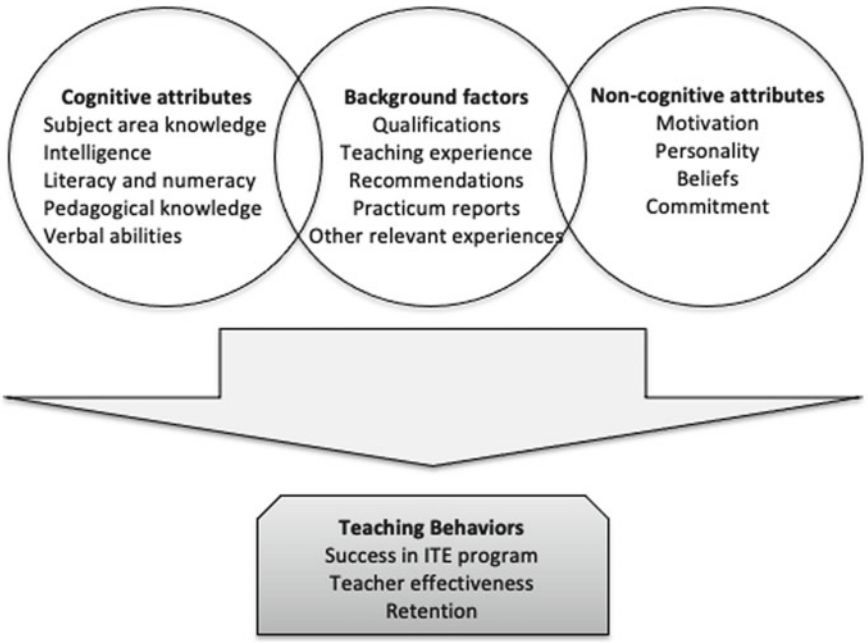


Fig. 3.1 Examples of cognitive attributes, background factors, and non-cognitive attributes that contribute to teaching behaviors

3.2 The Distinction Between Cognitive and Non-Cognitive Attributes

When examining the profile of prospective teachers, a selector will almost always scrutinize background factors (relevant qualifications) and will usually seek evidence of cognitive attributes, perhaps through proxies such as grades in academic courses, or in some cases, scores on university entrance exams (such as SAT [Scholastic Aptitude Test] scores in the United States). In many settings, there is often marked interest in the assessment of *non-cognitive attributes* (sometimes called 'non-academic' attributes) that are deemed important for teaching. The term 'non-cognitive attributes' refers to within-person variables variously described as beliefs, motives, personality traits, and dispositions (e.g., Patterson et al., 2015). In selection research, cognitive attributes (sometimes called 'academic' attributes) typically refer to variables that reflect reasoning skills (in the US measured by the Scholastic Aptitude Test, SAT) or academic achievement (e.g., grade point average [GPA]). For selection in education settings, cognitive attributes may also refer to subject area knowledge and pedagogical knowledge. Most teacher selection processes attempt to assess candidates' non-cognitive and cognitive attributes with the belief that both contribute to future success in teaching, as portrayed in the model in Fig. 3.1.

3.3 Individual Attributes and Teacher Effectiveness

When people look back to their time in school, they tend to remember their teachers' individual attributes more than their teaching practices or behaviors (Urdan & Pajares, 2008). Teachers' individual attributes shape their students' learning engagement, although these attributes interact with teaching methods in complex ways. When admissions teams select applicants for teacher training or when principals hire teachers for vacant positions, the goal is to identify applicants who display the personal attributes *and* the knowledge about teaching methods that are believed to lead to successful outcomes. But there are many questions about finding teachers with the attributes that make up 'the right stuff': are these attributes liable to change? Does it matter if applicants display these attributes at the point of selection, or can they be developed? To understand teacher selection, it is important to consider the attributes that research shows to be related to effective teaching and, furthermore, to understand which attributes are likely to be subject to development over time.

The stability of the elements of teacher effectiveness varies according to context: current interactionist approaches suggest that the expression of traits and attributes depends on the interaction between the person and the particular situation. For example, in latent state-trait theory (Steyer et al., 1999), momentary expression of trait tendencies—states—are underpinned by an underlying latent trait that may be expressed differently according to setting. The individual attributes underpinning teacher effectiveness may develop over time, but patterns of the expression of

these attributes are associated within individuals. The importance of the individual attributes remains high throughout a teacher's career but identifying and evaluating these attributes is especially important at entry into the profession, because the potential benefits of identifying individual variation in effectiveness is highest (Atteberry et al., 2015).

3.4 Research on Attributes Related to Teacher Effectiveness

The salience of the trio of background factors, cognitive attributes, and non-cognitive attributes is not determined only by local demands (for example, by the need for teachers in a particular subject area), but also by the weight of predictive evidence for each of the three factors. The prediction utility of the three factors vary according to purpose and context, and in this next section we examine evidence for some of the predictors typically included in teacher selection.

Cognitive abilities. It seems reasonable that the cognitive abilities of teachers play an important role in their performance in the classroom, and selectors typically include some kind of measure or proxy of cognitive abilities in selection decisions. Sautelle et al. (2015) surveyed teachers and non-teachers about their views of the attributes believed to lead to effective teaching and found that cognitive ability was rated as the most important attribute for teacher selection, followed by the personality traits of conscientiousness and agreeableness. The 'bright person hypothesis' (e.g., Kunter et al., 2013) proposes that because teaching is a highly demanding and complex job, effective teachers possess a high degree of cognitive flexibility and intellectual capacity. Kunter and colleagues tested this hypothesis alongside two other hypotheses: 'the knowledgeable teacher hypothesis' (effective teachers acquire profession-specific knowledge) and the 'professional competence hypothesis' (effective teachers display an interplay of knowledge, skills, attitudes, and motivation). The authors found that a combination of pedagogical content knowledge, teacher beliefs, and teacher enthusiasm had an effect on students' mathematics achievement gains and mathematics enjoyment. In contrast, teachers' general cognitive ability (as measured by GPA) was unrelated to student achievement and enjoyment.

Bardach and Klassen (2020) examined the role of teachers' cognitive abilities in relation to teacher effectiveness through a systematic review of 27 studies that explored teachers' cognitive abilities (including measures of intelligence and proxies of cognitive abilities such as college entrance scores and academic skills assessments) and teacher effectiveness, measured as student achievement growth and external observer ratings. Although intelligence test scores have been shown to be a valid predictor of job performance in multiple studies (e.g., Ones et al., 2012), Bardach and Klassen found that half of the studies they reviewed did not show any statistically significant effect, whereas the other half showed negative effects. Studies that linked proxies of teachers' cognitive abilities and effectiveness were more differentiated, with no relation between college entrance test scores and effectiveness, small or null effects of basic skills tests and effectiveness, and some (but tenuous) effects of

teachers' mathematics abilities on students' mathematics achievement. In terms of predictors to be examined in the teacher selection process, the authors concluded that non-cognitive attributes may be a more promising avenue of development in teacher selection research.

Verbal abilities. Communication skills are at the top of many people's list of attributes when considering teacher effectiveness. When asked to name the key predictors of teacher effectiveness, principals consistently report that communication skills are of high importance (e.g., Harris et al., 2010; Ralph et al., 1998; Tamir, 2019). However, finding evidence for the predictive relationship between communication skills and teacher effectiveness proves challenging. Aloe and Becker (2009) conducted a meta-analysis of 19 studies that explored the relations between teachers' verbal ability and effectiveness and concluded that the relation was "very weak at best" (p. 620), with median r value of 0.03 (*ns*). The one facet of verbal ability that seemed more salient for effectiveness was verbal fluency ($r = 0.17$, $p < 0.05$), although most of the positive evidence came from the disputed (and dated) *Equality of Educational Opportunity* dataset from the 1960s. The results of this meta-analysis are contrary to expectations, and there is currently a dearth of follow-up research that explains the nature of the relation between teachers' communication skills and their effectiveness.

Non-cognitive attributes. The challenge during teacher selection is to assess applicants' non-cognitive attributes in a way that is reliable, valid, and fair. However, identifying which attributes to target, and determining *how* to measure these attributes is a challenge. There is a further divide in relation to the question of teacher selection: which attributes are readily measured in a *high-stakes* situation; i.e., when applicants feel pressure to respond in a socially or professionally desirable manner? For example, considerable research has examined teachers' self-efficacy in relation to effectiveness, and the findings are generally robust (e.g., Klassen & Tze, 2014), but can self-efficacy be readily assessed when applicants are aware that their responses are linked to decision-making processes i.e., in a high-stakes situation? An individual completing a teaching self-efficacy measure anonymously for research purposes may respond differently when faced with the same measure during a selection process. Consider a question on the frequently used *Teachers' Sense of Efficacy Scale* (Tschannen-Moran & Woolfolk Hoy, 2001): *How much can you do to help students value learning?* (1 = *Nothing* to 9 = *A great deal*). During a selection process, an applicant to an ITE program or for a teaching job will likely respond on the high end of the scale, regardless of their actual confidence levels, thus providing information that may not be useful for the selection team.

A number of non-cognitive attributes have received research attention in the last few years, including research on teachers' motivation, e.g., self-efficacy (e.g., Klassen et al., 2011), personality (e.g., Kim et al., 2019) and emotions (e.g., Chang & Taxer, 2020). Other non-cognitive attributes have been investigated in single studies (e.g., *dispositions* (Fonseca-Chacana, 2019), and see Bardach et al., 2020), but in the next section we focus on three broad categories of attributes—motivation (especially self-efficacy), personality, and emotions—and look at their relevance and promise for teacher selection.

Self-efficacy. Teachers' self-efficacy—the confidence teachers hold about their individual and collective capabilities to influence student achievement—is one of the key motivating factors that influence teacher behaviors and student outcomes (Bandura, 1997). Measures of teachers' self-efficacy are usually formulated along the lines of '*How confident are you that you can….* e.g., *manage student behavior*? Self-efficacy refers to domain-specific confidence, not general confidence, with the level of specificity depending on the purpose of the measurement. The continuum of teacher self-efficacy specificity ranges from the very general (*How confident are you that you can teach effectively?*) to the very specific (*How confident are you that you can effectively manage Gregory's behavior in Class 9b today?*). In spite of the considerable volume of research on teachers' self-efficacy over the last few decades, research on the links between teachers' self-efficacy and their effectiveness is less common, with Klassen et al. (2011) finding that fewer than 3% of studies directly linked the two variables. In a follow-up meta-analysis that assessed the associations between two psychological characteristics—teachers' self-efficacy and teacher personality—Klassen and Tze (2014) found a (medium-sized) correlation of $r = 0.28$ between teachers' self-efficacy and observed teaching performance, but the correlation with student achievement was much smaller ($r = 0.08$). One problem with incorporating measures of teachers' self-efficacy into high-stakes selection processes, pointed out by Bardach et al., 2020, is that due to social desirability effects, self-efficacy may be a less useful attribute for teacher selection. Although self-efficacy can be a valid predictor of actual behavior, the self-report scales typically to measure self-efficacy are transparent, i.e., the 'best' answers are obvious. Using explicit measures (i.e., direct questions) can affect all high-stakes assessment, but some individual attributes, such as self-efficacy, may be especially prone to the effects.

Personality. The link between teacher personality and effectiveness is intuitively appealing. Personality, defined as enduring, relatively stable clusters of behaviors, cognitions, and emotional patterns, is one of the preeminent conceptualizations of how individuals vary in their interactions with the environment. Research on the links between teachers' personality and effectiveness has produced mixed results. Corcoran and O'Flaherty (2018) examined the relations between Big Five personality traits and pre-service teachers' teaching performance at Year 2 and Year 4 in their teacher training program. At Year 2, significant correlations were found between three personality traits—conscientiousness, agreeableness, and extraversion—and teaching performance, but these correlations disappeared in Year 4. Klassen and Tze (2014) conducted a meta-analysis of research examining the effects of teachers' self-efficacy and personality on effectiveness and found a modest association ($r = 0.08$) between composite personality and effectiveness. Kim et al. (2019) conducted a meta-analysis covering 25 studies on Big Five personality traits, teacher effectiveness and burnout, and found that four of five traits were significantly related to effectiveness, with extraversion ($r = 0.17$, $p < 0.05$) the largest, and agreeableness ($r = 0.03$, $p > 0.05$), the smallest. The authors proposed that teacher personality research could potentially inform teacher selection practices, although Bardach et al.'s review of the teacher personality literature pointed to inconsistent and "overall not very

promising findings" (p. 18) when aggregating results across existing syntheses. Additional work on applying personality measures to teacher selection would be needed before inclusion in selection systems.

Emotion regulation. Teaching is an emotion-laden activity, and links have been shown between teachers' emotions and a wide range of outcomes including classroom effectiveness, teacher well-being and health, and student emotions and motivation (Taxer & Gross, 2018). The ability that teachers possess to regulate their emotional state, and to identify and manage the emotional state of their students, has been examined in a wide range of recent research. Chang and Taxer (2020) conducted a pair of studies exploring how teachers regulated their emotions in response to student misbehavior. In the first study, they found that students' non-cooperation and defiance spurred teachers' anger and frustration, with teachers most reporting that they most frequently used the regulation strategy of *suppression* (deliberately inhibiting behavioral expression) to manage their emotional responses. In Study 2, the authors found that teachers who reported high levels of the emotional regulation strategies of reappraisal (modifying their appraisal of a situation) and suppression were the ones who also reported higher levels of emotional exhaustion on a daily basis. Taxer and Gross (2018) found that teachers regulated their emotions to achieve multiple goals, including *hedonic* goals (reducing the experience of anger), and to increase their teaching effectiveness. Research on teachers' emotions may lag somewhat behind research on teachers' motivation and personality, but the work done to date shows the importance of the ability for teachers to regulate their emotions in teaching situations. In Bardach et al.'s integrative review, emotional intelligence (the ability to perceive and interpret emotions in themselves and others) was noted as a potential candidate for teacher selection, with promising (but limited) research showing links with teacher effectiveness.

Review of teachers' personal characteristics. Bardach and colleagues (Bardach et al., 2020) conducted a 'synthesis of syntheses' or *meta-synthesis* whereby they aggregated the findings from quantitative meta-analyses and qualitative reviews that explored the relations between teachers' psychological characteristics (i.e., non-cognitive attributes) and teacher effectiveness, including multiple student outcomes and teacher outcomes such as well-being. Their search uncovered a total of 24 syntheses focused on: teachers' motivation (self-efficacy [6 syntheses] and causal attributions [1]), personality (3), expectations (2), emotion-related factors (5), and mindfulness (7 syntheses). Results confirmed the important role that self-efficacy plays in teachers' performance and well-being, with some indication that self-efficacy is also associated with retention in the profession. Causal attributions (the causes to which individuals ascribe their success or failure) and teacher expectations were found to be promising predictors of effectiveness, whereas personality and mindfulness were deemed to be less-promising predictors. Finally, emotion-related factors (emotional intelligence, emotional labor, and enthusiasm) were, as a group, found to be promising predictors of key teacher outcomes.

3.5 Inductive, Deductive, and Integrated Approaches

Not all of the research on individual attributes has used a deductive approach to understand the key contributors to teacher effectiveness. For the purpose of teacher selection, individual attributes have mostly been identified using a 'top-down' research-based approach, i.e., by targeting particular attributes in a selection process based on a strong research foundation. In contrast, a 'bottom-up' or inductive approach has been used to develop some teacher selection tools, e.g., situational judgment tests (SJTs), whereby key attributes are identified during the content development process (e.g., Campion et al., 2014). With an inductive approach, researchers work alongside expert practitioners to identify 'critical incidents' in professional practice, assigning inductive categories to the content. A third approach to identify key attributes for teacher selection is to adopt a 'construct-informed integrated approach' (see Klassen et al., 2020), which combines deductive and inductive approaches to the identification of key attributes. In Fig. 3.2 we show an example of how the attributes underlying an SJT used for selection were identified using an integrated approach, with three attributes—self-efficacy, emotion regulation, and conscientiousness—derived from extensive research on effectiveness, and three attribute clusters—organization and planning, empathy and communication, and resilience and adaptability—derived from extensive interactions with expert practitioners. The benefits of using an integrated construct-informed approach to develop selection tests such as SJTs is that the heterogeneous test content more clearly reflects the heterogeneity of performance on the job it is meant to be predicting. The disadvantages of an integrated

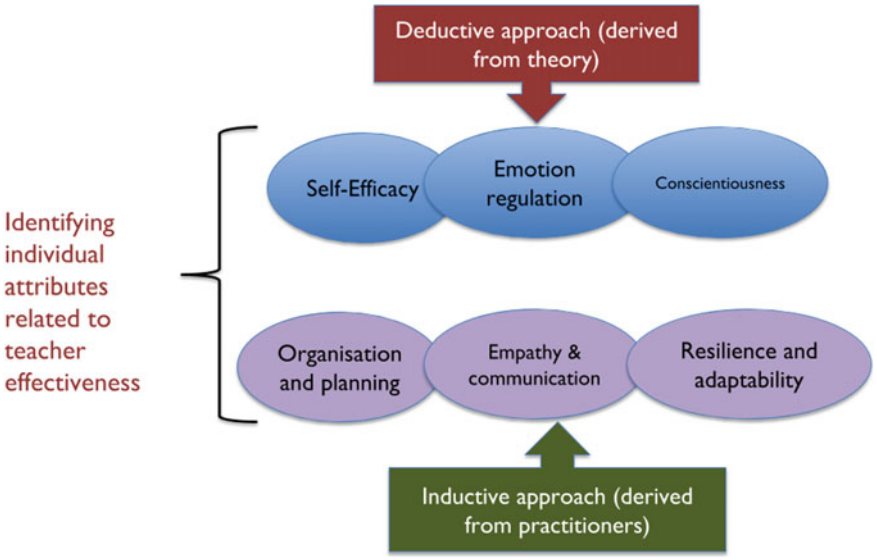


Fig. 3.2 An integrated approach to identifying non-cognitive attributes

approach include difficulties in interpreting scores and low internal consistency. Increasing attention is paid to how teachers' non-cognitive attributes relate to their effectiveness (e.g., Bardach et al., 2020), but most of the work has been conducted in Western settings, and little is known about the universality of the attributes heretofore identified.

3.6 Cross-Cultural Perspectives: Which Attributes Are Universal?

Teachers around the world share many of the same day-to-day activities: structured teaching in a classroom, less structured interacting with students in and outside of classroom settings, collaborating with fellow teachers and principals, and communicating with parents and other community members about educational issues. The make-up of daily teaching practices may vary from country-to-country or region-to-region. For example, the proportion of teacher-directed vs. student-directed activities may vary, or the nature of teacher-student interactions: McIntyre et al. (2017) showed that there are significant differences in the way that teachers interact with students in the UK and Hong Kong. An understanding of the attributes that are possessed by an *effective* teacher may not be shared across cultures, and most of the research on teacher effectiveness has been undertaken in single, usually Western, settings. Comparative research has shown that teachers' values and priorities vary across contexts: Meng & Muñoz (2016) reported that perceptions of teacher effectiveness were influenced by behaviorist beliefs in American teachers and by moralistic (Confucian) beliefs in Chinese teachers. Although teachers' activities may look similar across cultures, the way that cultural context influences the individual attributes associated with effective teaching is not entirely clear. In Fig. 3.3, we propose a model in which culture—the shared beliefs, goals, and values that guide the way we understand relationships, expectations, duties, and activities (Schwartz, 1994)—directly influences the educational environment (e.g., educational standards and national expectations), but also the preferred personal characteristics (such as non-cognitive attributes), and the kinds of behaviors that are expected in the classroom. In this model, culture also plays an indirect role through the mediating effects of the educational environment (e.g., nationally agreed standards), which has a direct effect on the expected personal characteristics and teaching behaviors. The model offers one way to understand how the cultural context might influence understandings about the desired individual attributes of effective teachers.

Klassen et al. (2018) recently conducted a cross-cultural study exploring how experienced teachers and teacher educators in four culturally disparate countries—England, Finland, Malawi, and Oman—conceptualized the key non-cognitive attributes of novice teachers. The study emerged from work on developing SJTs for selecting teachers into ITE programs in multiple international settings. In the SJT development process, non-cognitive attributes are identified and defined, and then

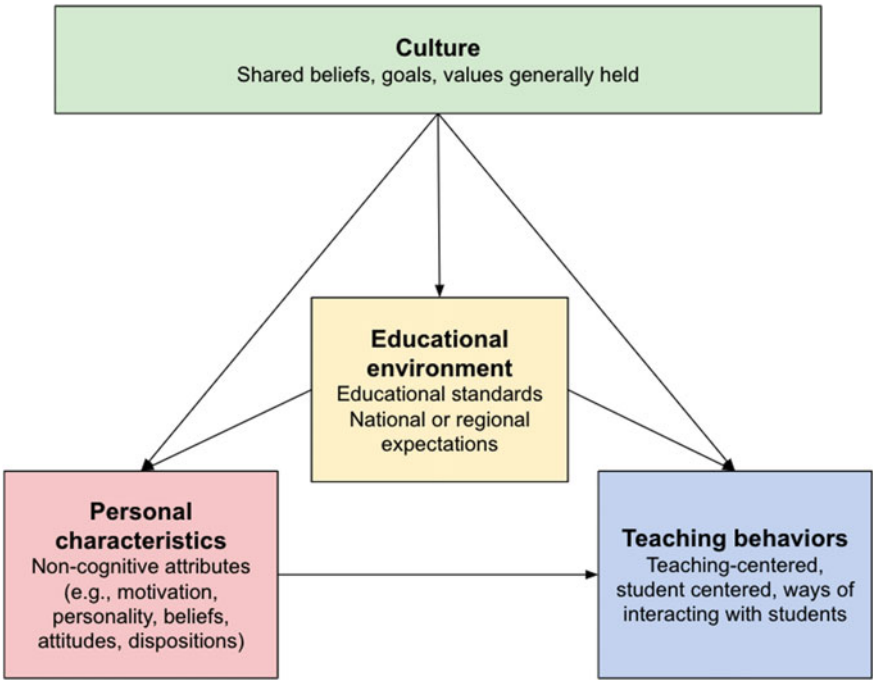


Fig. 3.3 How cultural context influences the educational environment, non-cognitive attributes, and teaching behaviors (adapted from Klassen et al., 2018)

form the basis for item development, usually using a critical incident approach that illustrates a particular attribute, or cluster of attributes (e.g., resilience and adaptability). The study originated in England, where three clusters of inductively derived attributes were initially identified by the iterative work of several panels of experienced educators: empathy and communication, organization and planning, and resilience and adaptability. Next, these clusters were introduced and debated in three comparison countries—Finland, Malawi, and Oman—as part of a teacher selection tool-developing process.

After multiple rounds of discussion of the key attributes, experts in each of the three comparison countries endorsed the initial cluster of attributes, suggesting that some core attributes of effective teachers span cultural contexts. However, in each setting, additional attributes were proposed: in Finland, *cooperation and fostering of community*; in Malawi, *autonomy, integrity and community relations, motivation and commitment,* and *reflection and creativity*; in Oman, *professional ethics, and enthusiasm and motivation.* Findings from this study suggested that certain attribute clusters are universally endorsed across very different cultural contexts, but at the same time, important additional attributes were identified in each setting, with these additional attributes aligned with salient cultural practices and beliefs. Identifying

the key non-cognitive attributes for teacher selection demands attention not only to 'educational universals', but also to critical cultural factors.

3.7 Chapter Summary

In this chapter, we looked at a range of personal characteristics of effective teachers, and how research on cognitive and non-cognitive attributes has shown that certain attributes are reliably associated with teacher effectiveness and may be valuable for use in teacher selection. Cognitive attributes such as academic achievement and subject knowledge have long been associated with teacher selection practices but are insufficient to reliably predict teaching success. Unfortunately, the research picture on measuring non-cognitive attributes in high-stakes settings is far from clear, with some attributes, such as teachers' self-efficacy showing strong links with teacher effectiveness, but with moderate utility for selection. Other attributes, such as personality, have long been used for personnel selection outside of education, but may have limited appeal for teacher selection. Some promise was shown in the broad domain of emotions, with the potential for measures of emotion regulation and emotional intelligence showing some potential for application to teacher selection. Cultural context influences beliefs about key teacher attributes and using an integrated deductive and inductive approach to identifying key attributes shows promise, especially when crossing cultural boundaries. In the next chapter, we begin to look at selection research and theory, and in particular at some of the key issues and challenges facing the field.

References

Aloe, A. M., & Becker, B. J. (2009). Teacher verbal ability and school outcomes. Where is the evidence? *Educational Researcher, 38*, 612–624.

Atteberry, A., Loeb, S., & Wyckoff, J. (2015). Do first impressions matter? Predicting early career teacher effectiveness. *AERA Open, 1*(4), 1–23.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. Freeman.

Bardach, L., & Klassen, R. M. (2020). Smart teachers, successful students? A systematic review of the literature on teachers' cognitive abilities and teacher effectiveness. *Educational Research Review, 30,* 100312.

Bardach, L., Klassen, R. M., & Perry, N. E. (2020). Teachers' psychological characteristics: Do they matter for teacher effectiveness, teachers' well-being, retention, and interpersonal relations? An integrative review. *Educational Psychology Review*. https://doi.org/10.1007/s10648-021-09614-9

Brock, A., & Ryan, T. G. (2016). Exploring the gap between teacher certification and permanent employment in Ontario: An integrative literature review. *Canadian Journal of Educational Administration and Policy, 175*, 1–28.

Campion, M. C., Ployhart, R. E., & MacKenzie, W. I., Jr. (2014). The state of research on situational judgment tests: A content analysis and directions for future research. *Human Performance, 27*, 283–310.

Chang, M.-L., & Taxer, J. (2020). Teacher emotion regulation strategies in response to classroom misbehavior. *Teachers and Teaching*. https://doi.org/10.1080/13540602.2020.1740198

Corcoran, R. P., & O'Flaherty, J. (2018). Factors that predict pre-service teachers' teaching performance. *Journal of Education for Teaching, 44*, 175–193.

Fonseca-Chacana, J. (2019). Making teacher dispositions explicit: A participatory approach. *Teaching and Teacher Education, 77*, 266–276.

Goldhaber, D., Krieg, J., Theobald, R., & Brown, N. (2015). Refueling the STEM and special education teacher pipelines. *Phi Delta Kappan, 97*(4), 56–62.

Harris, D. N., Rutledge, S. A., Ingle, W. K., & Thompson, C. C. (2010). Mix and match: What principals really look for when hiring teachers. *Education, 5*(2), 228–246.

James, W. (1899/2015) *Talks to teachers.* Retrieved from: http://www.gutenberg.org/ebooks/16287?msg=welcome_stranger#I__PSYCHOLOGY_AND_THE_TEACHING_ART

Kim, L. E., Jörg, V., & Klassen, R. M. (2019). A meta-analysis of the effects of teacher personality on effectiveness and burnout. *Educational Psychology Review, 31*, 163–195.

Klassen, R. M., & Durksen, T. L. (2014). Weekly self-efficacy and work stress during the final teaching practicum: A mixed methods study. *Learning and Instruction, 33*, 158–169.

Klassen, R. M., Durksen, T. L., Al Hashmi, W., Kim, L. E., Longden, K., Metsäpelto, R.-L., Poikkeus, A. M., & Györi, J. (2018). Cultural context and teacher characteristics: Exploring the non-cognitive attributes of prospective teachers in four countries. *Teaching and Teacher Education, 72*, 64–74.

Klassen, R. M., Kim, L. E., Rushby, J., & Bardach, L. (2020). Can we improve how we screen applicants for initial teacher education? *Teaching and Teacher Education, 87,* 102949.

Klassen, R. M., & Tze, V. M. C. (2014). Teachers' self-efficacy, personality, and teaching effectiveness: A meta-analysis. *Educational Research Review, 12*, 59–76.

Klassen, R. M., Tze, V. M., Betts, S. M., & Gordon, K. A. (2011). Teacher efficacy research 1998–2009: Signs of progress or unfulfilled promise? *Educational Psychology Review, 23*, 21–43.

Knight, F. B. (1922). Qualities related to success in elementary school teaching. *The Journal of Educational Research, 5*, 207–216.

Kunter, M., Kleickmann, T., Klusmann, U., & Richter, D. (2013). 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). Springer, US.

McIntyre, N., Mainhard, M. T., & Klassen, R. M. (2017). Are you looking to teach? Cultural, temporal, and dynamic insights into expert teacher gaze. *Learning and Instruction, 49*, 41–53.

Meng, L., & Munoz, M. (2016). Teachers' perceptions of effective teaching: A comparative study of elementary school teachers from China and the USA. *Educational Assessment, Evaluation and Accountability, 28*, 179–199.

Ones, D. S., Dilchert, S., & Chockalingam, V. (2012). Cognitive Abilities. In N. Schmitt (Ed.), *The Oxford Handbook of Personnel Assessment and Selection.* New York: Oxford University Press, USA. http://dx.doi.org/https://doi.org/10.1093/oxfordhb/9780199732579.013.0010

Patterson, F., Zibarras, L., & Ashworth, V. (2015). Situational judgement tests in medical education and training: Research, theory and practice: AMEE Guide No. 100. *Medical Teachers, Aug 27,* 1–15 E-pub ahead of print.

Ralph, E. G., Kesten, C., Lang, H., & Smith, D. (1998). Hiring new teachers: What do school districts look for? *Journal of Teacher Education, 49*(1), 1–10.

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*, 54–71.

Schwartz, S. H. (1994). Beyond individualism-collectivism. In U. Kim, H. C. Triandis, C. Kagitcibasi, S. C. Choi, & G. Yoon (Eds.), *Individualism and collectivism: Theory, method and applications* (pp. 85–119). Sage.

Steyer, R., Schmitt, M., & Eid, M. (1999). Latent state-trait theory and research in personality and individual differences. *European Journal of Personality, 13*, 389–408.

Tamir, E. (2019). What principals look for when hiring new teachers. *Leadership and Policy in Schools*. https://doi.org/10.1080/15700763.2019.1638423

Taxer, J. L., & Gross, J. J. (2018). Emotion regulation in teachers: The "why" and "how." *Teaching and Teacher Education, 74*, 180–189.

Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17*, 783–805.

Urdan, T., & Pajares, F. (Eds.). (2008). *The ones we remember: Scholars reflect on teachers who made a difference*. IAP.

Part II
Selection Methods and Practices: Issues and Uses

Chapter 4
Issues and Challenges in Selection

Abstract In the previous chapter we explored the attributes one can assess when selecting teacher candidates. In this chapter, we consider how these attributes can be measured. Choosing selection tools to be included in a selection program can be a complex task that requires consideration of a wide range of factors. Central to this decision-making is examining the tool's criterion-related validity; that is, how much a score on that test is associated with a criterion (Lievens et al., 2021). Additionally, other factors must be considered, including the potential adverse impact of their use, applicants' perceptions of the selection program, and how faking and coaching might influence applicants' performance. Ignoring these issues can lead to a relatively homogenous pool of selected applicants, applicants with negative perceptions of the selection procedure, and suboptimal selection decisions based on scores that may not be 'true' reflections of the applicants. In this chapter, we will examine some issues and challenges that should be considered when choosing and implementing a selection program. We will also suggest some ways of addressing key challenges in teacher selection.

4.1 Test Quality and Theoretical Models Underpinning Selection

Examining the psychometric properties of a selection test (also referred to as tool or method) is important in assessing its quality. Properties such as whether the test measures what it is intended to measure (validity) and whether it measures a characteristic consistently (reliability) are often considered. In selection, one of the most important factors that is considered is the tests' criterion-related validity; that is, how strongly the test scores are associated with outcomes that are considered to be important in the field (e.g., job performance, academic achievement).

Comparing the evidence of criterion-related validity across selection tests is key to deciding which test(s) to use in the selection procedure. Such comparisons are published in meta-analyses, which are consolidated collections of studies with validity estimates that take into account factors such as sampling error, range restriction, and measurement reliability (Hunter & Schmidt, 1990). One of the first most

R. M. Klassen and L. E. Kim, *Teacher Selection: Evidence-Based Practices*,
https://doi.org/10.1007/978-3-030-76188-2_4

comprehensive meta-analyses on personnel selection methods was conducted by Schmidt and Hunter (1998), who assessed the validity of 17 tools in association with overall job performance. Many researchers have conducted studies and meta-analyses in this area since, including Lievens et al. (2021) who updated Schmidt and Hunter's (1998) table. The updated findings are generally consistent with the original findings; the strongest predictor of job performance is cognitive ability test (corrected validity = 0.51), and the weakest predictors are graphology (0.02) and age (0.03). These findings attest to the importance of considering criterion-related validity when choosing tests to be included in a selection program. Looking at Schmidt and Hunter's meta-analysis for evidence of incremental validity, the best predictors of overall job performance were cognitive ability tests combined with integrity tests, cognitive ability tests combined with structured interviews, and cognitive ability tests combined with work samples (composite validity = 0.65). These findings indicate that using multiple selection tools may be more effective than using a single tool.

Moreover, it is important to consider the theoretical models that the tools are based on when choosing selection tools to measure certain attributes. Let's consider personality as an example since personality measures are often included as part of personnel selection procedures. Personality can be theoretically conceptualized in a myriad of ways, which is reflected by the existence of a variety of measures. They include Goldberg's (1992) Big Five Factor Markers, based on the Big Five theory; Hogan Personality Inventory (Hogan & Hogan, 2007), based on the socioanalytic theory; and Myers-Briggs Type Indicator (Myers et al., 1998), based on Jung's theory of psychological types. There are also personality tools, though are loosely based on theory and have little evidence base, that are used in some fields. It is recommended that selection tools are chosen not based on what is well-known and 'makes sense' (i.e., has high face validity) but on the rigor of the theory and evidence on which the tool is built on. Going back to the example of personality, the Big Five is one of the most accepted theoretical models among personality psychologists (e.g., Sackett et al., 2017) and tools based on the Big Five are usually better predictors of outcomes than those that are not (e.g., Salgado, 2003). Hence, using an assessment tool based on a strong theoretical model and empirical base is important when building a selection program.

4.2 Adverse Impact

Adverse impact refers to the phenomenon whereby members of a subgroup (e.g., using categories such as ethnicity, gender, and age) are selected at different rates than members of another subgroup. Formally, adverse impact is said to occur if the selection ratio for the lower scoring group divided by the selection ratio for the group with the highest selection ratio is less than 0.8 (Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, & Department

of Justice, 1978). Understanding the sources of adverse impact can help inform decisions on which assessment tools are used in the selection process, and how these tools are used.

It can be difficult to choose assessment tools that have the highest validity and ensures high diversity among the accepted applicants, otherwise known as the diversity–validity dilemma. That is, some of the strongest predictors of important outcomes are associated with large ethnic and sex subgroup differences (Pyburn et al., 2008). Some researchers have claimed that "it is unreasonable to expect that one can maximize both the performance and ethnic diversity of selected individuals" (Sackett et al., 2001, p. 302). A selection strategy emphasizing diversity would look different to a strategy emphasizing validity (Sackett & Roth, 1996). In practice, there is often a trade-off between expected performance and diversity of the accepted applicants when choosing selection assessment tools.

Certain selection methods lend themselves to higher adverse impact than others. Typically, *d* statistics are used to understand differences between subgroups, which are the mean of one group minus the mean of the other subgroup divided by their pooled standard deviation. Hence, *d* values represent differences in standard deviation units, such that $d = 0.50$ indicates that the two subgroups differ by 0.50 standard deviation units. The greater the *d* value, the greater the difference between the subgroups (with a positive value indicating higher scores from the first subgroup). Using assessment tools showing high *d* values means that it is less likely that different subgroups are selected equally. Ployhart and Holtz (2008) conducted a meta-analysis on the ethnic and sex subgroup differences of 19 predictors, ranging from general cognitive ability, personality domains, situational judgment tests, and assessment centers. General cognitive ability tests were found to be a source of adverse impact (with *d* value of 0.99 between White and Black subgroups), and, as discussed earlier in the chapter, it is one of the strongest predictors of overall job performance. The predictors with the smallest *d* value between subgroups were the personality domains, particularly conscientiousness. Test methodologies also ranged in their *d* values. Situational judgment tests showed small differences between males and females ($d = -0.06$ for video versions; $d = -0.12$ for written versions) and larger differences between ethnic groups (e.g., $d = 0.31$ for video versions and $d = 0.40$ for written versions between White and Black subgroups). Assessment centers, on the other hand, reported White–Black differences of $d = 0.60$ or less depending on the content.

Ployhart and Holtz (2008) proposed 16 different strategies to minimize adverse impact, which was classified into five categories. First, predictors with smaller subgroup differences should be used (e.g., structured interviews and assessment centers over general cognitive ability). Second, combining scores from multiple tests can reduce subgroup differences (e.g., using multiple cognitive and non-cognitive predictors to balance out known subgroup differences). This modular approach not only can reduce potential subgroup differences, but it can also be a way of improving the theoretical underpinnings of selection processes (Lievens & Sackett, 2017). Third, removing construct irrelevant predictor score variances (e.g., language, cultural differences) can reduce subgroup differences. The fourth approach is to allow

applicants to practice or to be retested. The fifth and final category is to increase applicant feedback to maximize diversity in recruitment and performance in the selection procedure. Ployhart and Holtz concluded that the first two categories were the most effective in reducing subgroup differences as they most directly address the problem. Specifically, they outlined two effective strategies included using alternative measurement methods that measure multiple constructs (e.g., interviews, SJTs) instead of tests dedicated to measuring only one construct (e.g., cognitive ability tests); and to assess both cognitive and non-cognitive constructs so that the subgroup differences can be balanced out between assessing both types of constructs.

Various strategies can be applied to the context of teacher selection to maximize the diversity of selected applicants. For example, Ployhart and Holtz (2008)'s two strategies of choosing appropriate measurement methods and assessing a wide range of job-relevant cognitive and non-cognitive constructs may be effective here. These two strategies have been applied in medical education selection, where both cognitive ability tests (with high inter-group differences) and SJTs (with lower subgroup differences; Whetzel et al., 2008) are used to measure both cognitive and non-cognitive constructs in countries like the UK (Patterson et al., 2017) and Belgium (Lievens et al., 2016; see Chap. 5 for more details). As such, there is argument to measuring both cognitive and non-cognitive constructs and using their appropriate measurement methods (e.g., SJTs), when selecting applicants into teacher education programs and the teaching profession to help meet the various goals of the selection process (i.e., high validity and diversity).

4.3 Applicants' Perceptions of Selection Procedures

Practitioners are increasingly viewing selection procedures as an interactive social process (McCarthy et al., 2017). That is, not only is the selection committee choosing the applicants, but the applicants are choosing the organization or program based on the perceptions they have formed during the selection procedure. Thus, as a selection committee aims to secure top applicants into their organization or program, they may be inevitably wary of how applicants may be perceiving their selection procedure. Accordingly, different fields have examined applicants' perceptions of selection procedures, including in personnel selection (Truxillo & Bauer, 2011), medical education selection (Patterson et al., 2011), and teacher education selection (Bardach et al., 2021; Klassen et al., 2014).

Applicants' perceptions of the selection process can influence their performance, intentions, and behavior. A systematic review on applicant perspectives found that the valence of applicants' views (i.e., how positive or negative they felt about the process) correlated with their actual and perceived performance (McCarthy et al., 2017), indicating that positive applicant perceptions are important for their performance and thus the selection committee's use of their scores. Furthermore, the review also reported that applicant perceptions are positively correlated with their view of

the organization, intention to accept the organization's job offer, and likelihood to recommend the employer to others.

Gilliland developed one of the first theoretical models of applicants' perceptions of fairness of selection systems (Gilliland, 1993). Based on organizational justice theory (Greenberg & Cropanzano, 1993; Lind & Tyler, 1988), Gilliland specified two elements associated with applicants' reactions. The first element was procedural justice—the fairness of the selection procedures—which is associated with the type of test administered, human resource policies, and human resource staff's interaction with applicants. These procedural justice rules are associated with structural fairness (factors associated with the selection process, such as timing of the feedback results and the tests) and social fairness (communication and treatment of the applicants; Bauer et al., 2001). The second element was distributive justice—the fairness of outcomes—which is associated with how hiring decisions are made, expectations of applicants' performance, salience of discrimination, and accommodating to applicants' special needs. Although both elements are associated with applicants' perception of the fairness of the selection system and the applicants' perception of the fairness of the outcome, the procedural justice rules are more strongly associated with the former and distributive justice rules are more strongly associated with the latter. Thus, to improve applicant perceptions on the selection procedure and the outcomes, it is helpful to consider the factors associated with these two justice elements and to address how they are reflected in the selection practice.

However, care must be taken when factoring in applicants' perceptions when making decisions associated with the selection process. Robertson and Smith (2001) concluded from their review of studies on applicants' perceptions of fairness that perceptions of fairness are closely intertwined with self-interest. That is, if an outcome or a treatment is favorable for the applicant, the selection method is often perceived as fair. On the other hand, if an outcome or a treatment is unfavorable for the applicant, the selection method is often perceived as unfair. The conflation between perception and outcome for unsuccessful applicants may be due to their self-esteem being damaged by rejection. That is, the cause of how they're feeling may be attributed, through a self-serving bias, to the unfairness of the system (Chan et al., 1998; McCarthy, et al., 2017).

Applicants' perceptions and teacher selection. In the case of teacher selection, applicants' perceptions should be one of the factors considered when developing selection procedures. As outlined above, perceptions are important to consider as they can influence not only applicants' decisions about accepting training places or jobs but also influence their motivation, attitudes, and performance during the selection process. For example, the negative attitude of an applicant may hinder performance in cognitive ability tests and truthful responding in non-cognitive tests. In addition, there may be reputational risks to negative applicant perceptions, not only for the organization, but for the profession as a whole. Poor selection processes, or processes that are perceived to be poor, may negatively impact the program and/or the teaching profession. However, since there can be a conflation between perception and outcome, further research is necessary to consider how applicants' perceptions

can change before and after the selection outcomes are revealed, and to explore ways such that self-interests do not conflate with perceptions of fairness.

4.4 Faking

Cognitive ability assessments are hard to fake because the answers are typically dichotomous and knowledge-based. In contrast, tests of non-cognitive attributes are often easier to fake because it is sometimes more obvious to surmise what the socially desirable responses are and to respond in line with what the 'expected' responses are perceived to be. Also, according to a review on selection interviews, "most applicants fake at least to some degree" (Melchers et al., 2020, p. 123). The concern for faking is that the scores from the procedures may not be a true reflection of the applicants' attributes, which raises questions as to how much one can trust and use the scores for selection purposes.

Before the issue of faking in selection is discussed, it is helpful to define some commonly used terms. *Response bias* is a generic term indicating systematic sets of responses that deviate from accurate or 'true' responses, whether that is done unintentionally (e.g., lack of self-awareness, misinterpretation of items) or intentionally (Furnham, 1986). When done intentionally, it can be classified as *faking*. Faking has two types: *faking good* (systematically presenting a better image of the self) and *faking bad* (systematically presenting a worse image of the self). The type of faking used by applicants in selection contexts is usually *faking good*, also known as *socially desirable responding* (Paulhus, 2002). That is, applicants attempt to present themselves in a way that does not accurately reflect their 'true' self but is in line with what they believe the selection committee is looking for.

Do applicants fake? The first question that may come to mind is, *do applicants really not answer truthfully?* Yes, test-takers routinely engage in faking in self-report assessments (Griffith et al., 2007; McFarland & Ryan, 2000; Melchers et al., 2020) and in interviews (Melchers et al., 2020). One cannot conclude, however, that *all* applicants fake *all the time*. Rather, some do in some circumstances. The extent to which a test-taker engages in faking is a function of various factors, including their ability (Snell et al., 1999), willingness, and opportunity to engage in faking (Levashina & Campion, 2006). When non-cognitive assessments are used for high-stakes decisions, such as for selection into training or employment, a test-taker may be more willing to engage in faking than in a low-stakes situation because it is deemed as more important for the individual and there is a potentially greater pay-off (Ziegler et al., 2011).

The number of applicants who engage in faking is not trivial across fields. For job applications, researchers have estimated that 30–50% of job applicants fake during their job application process (Donovan et al., 2003; Griffith et al., 2007). In the context of medical school selection, a study examined applicants' reported personality profiles and the extent to which they were faked (Griffin & Wilson, 2012). The authors found that both successful and unsuccessful applicants reported similar

personality profiles at the point of selection. However, the successful applicants, when tested on their personality four months later, showed that they had faked good on all five personality factors. More specifically, 62.7% of the applicants had faked good on at least one of the five personality domains, indicating that more than half of the applicants had misrepresented their personality profiles. This raises the possibility that applicants to other training programs may also engage in faking.

Does faking matter? Faking matters for selection programs because it can affect the assessment framework: the choice of constructs that are assessed during selection and the predictive validity of the tests used to assess those constructs. When applicants alter their responses to appear more attractive to the organization or program that they are applying for, faking is often manifested as higher scores on job- or program-relevant traits and lower variance in the scores. In effect, the assessment tool may no longer be measuring their 'true' levels of the construct. In turn, the assessment tool may not be able to predict the desired outcomes as accurately. That is, both the construct validity (Pauls & Crost, 2005) and the test–criterion relationship is adversely affected (Holden, 2007; Holden et al., 2001). Faking is also a fairness issue that can have meaningful consequences for applicants. It is possible that honest applicants may not be accepted into a job or program because other applicants faked their responses.

The issue of fairness and test–criterion relationship was investigated by Mueller-Hanson et al. (2003). Specifically, they aimed to investigate the effect of faking in the quality of the selection decisions (i.e., who was selected in) and the criterion-related validity. The experiment began by asking participants to complete an achievement motivation measure. Participants completed this measure in their randomly assigned conditions: in either a control condition (instructed to provide honest answers) or in an incentive group (instructed that the participants with high achievement orientation scores will be selected to the second stage of the experiment, where they could earn $20). In reality, all participants were invited to the second stage of the experiment and were asked to complete a performance test. The influence on the proportion of selected individuals from the two conditions and the performance between the two groups were examined at different selection ratios (i.e., proportion of people who would be selected based on their achievement motivation score). The study found that when achievement orientation was used as a selection criterion, a greater proportion of participants from the incentive group was selected at smaller selection ratios (i.e., less than 60%) but were similar at larger selection ratios (i.e., higher than 60%). However, the 'selected' participants from the control group performed better in the performance test than the 'selected' participants from the incentive group, and the difference between the scores between the two groups increased as the selection ratios decreased. These results indicate that participants' faking not only could change who is selected but also could decrease the test–criterion relationship. As such, for real-life selection practices, including teacher selection, the extent to which participants can and do fake should be considered when choosing methods and developing the assessment framework.

What can we do about faking in teacher selection? Though the potential issue of faking in teacher selection methods was raised decades ago (Sheldon, 1959),

cumulative research on this topic is relatively sparse. Given evidence that some applicants to jobs and other programs do fake, precautions should be taken in teacher selection practices. So, what can practitioners do about this issue? No selection method will ever be able to completely remove response distortion, but there are various methods that can help reduce the extent to which applicants distort their responses and to statistically correct scores to account for response distortion. Here are some suggestions:

1. Choose assessment methods that are harder to fake. Evidence indicates that test formats such as SJTs (Levashina & Campion, 2006), structured interviews (Van Iddekinge et al., 2005), and forced-choice questions (Jackson et al., 2000) are less prone to faking.
2. For interviews, choose a structure and content that are harder to fake. A meta-analysis by Barrick et al. (2009) indicates that scores from highly-structure interviews are less affected by faking good than low-structured interviews. Moreover, interview questions about past behavior are harder to fake than situational questions (Levashina & Campion, 2006).
3. Explore using statistical models that can take into account potential faking. Faking can be conceptualized as an interaction between a situation and a person, such that the measurement resulting from this includes measurement error (Schmidt et al., 2003). Some statistical methods, such as those that model socially desirable responding (e.g., Ziegler & Buehner, 2009), can partial out faking to capture only the 'true' construct, though research in this area is still developing.
4. Include items that can help detect faking to improve the selection process. Burns and Christiansen (2011) review various ways to measure faking, including using bogus items (examining applicants' endorsement of items which do not exist), items measuring overclaiming (like bogus items but more general; Paulhus et al., 2003), and examining idiosyncratic items responses (examining items which have different distributions under normal vs. faking conditions). Considering whether, and if so which type, of faking detection measure to include in the selection procedure could be helpful.

Thus, when we are selecting individuals into teacher education programs and into teaching positions, strategies to address potential faking should be considered. Such steps will provide programs and organizations greater confidence in applicant scores and in their selection decisions.

4.5 Coaching Effects

If a selection procedure is high-stakes and competitive, commercial coaching programs for these are likely to exist. For example, 51.4% of applicants selected to the interview stage of an Australian medical education program reported to have attended coaching sessions (Griffin et al., 2008). Coaching programs can take different forms,

ranging from familiarizing individuals with the test structure and content, teaching strategies on how to approach the test, providing feedback from practice tests, to delivering exercises that claim to develop the ability and knowledge necessary for the test (Messick & Jungelbut, 1981).

Coaching can affect the scores of both cognitive and non-cognitive tests. Certain types of ability tests are particularly prone to coaching effects and have been the focus of numerous meta-analyses. For example, Becker (1990) found that coaching led to increases in Scholastic Aptitude Test (SAT) Verbal scores by 0.09 *SD*s and SAT Math scores by 0.16 *SDs*. Greater coaching time is associated with greater score gains, though the relationship is not linear but logarithmic — after initial gains are made, an exponential increase in coaching time is required to continue to raise test scores at the same rate (Hausknecht et al., 2007; Messick & Jungelbut, 1981). It is also possible to be coached on non-cognitive tests, whereby individuals can be instructed on how to respond to items and how to alter their responses to match the profiles that may be, or perceived to be, desired by the selection committee. Looking at the evidence in different fields, the effects of coaching differ by the assessment tool. The effects are mixed for interviews (e.g., Maurer et al., 1998, 2001), and the effects are larger for overt integrity tests ($d = 1.54$) than for covert integrity tests ($d = 0.36$; Alliger & Dwight, 2000). The effects are moderate for SJTs ($d = 0.50$; Lievens et al., 2012), and minimal for MMIs (Griffin et al., 2008).

Selection committees must consider the possible effects of coaching. Just like the effects of faking, applicants' possible engagement in coaching makes it harder for selection committees to decipher the extent to which applicants' scores are accurate reflections of applicants' attributes. Moreover, coaching may increase biases related to income discrepancies. That is, candidates who can afford, and thereby receive coaching, may have an advantage in the tests, which presents issues of fairness and equity. Thus, it is worth considering how prevalent coaching is and how to minimize its influence on selection procedures and their outcomes.

One solution to minimizing coaching effects is to understand in which conditions the effects are the smallest. For example, there seems to be variation as to how much SJTs can be coached. Patterson et al. (2013) argued that the more complex the response formats are in SJTs, the harder they are to coach. They argued that the 'pick the best' test instructions (e.g., *Pick the best response out of these four responses*) are easier to coach than ranking (e.g., *Rank all five possible responses in order*) and rating instructions (e.g., *Rate the appropriateness of each response option*). Furthermore, collecting pre-test scores and propensity scores (score differences on other variables) may be helpful to statistically model coaching effects, although this information may be difficult to obtain and more research is needed to determine how best to use this kind of data. Moreover, test items should be regularly refreshed, as it should be assumed that the test items are fully exposed (Moshinsky et al., 2017). In sum, selection committees should be aware of the effects of coaching and provide conditions that can minimize the effects of coaching.

4.6 Chapter Summary

Selecting people for education programs and employment is an important but a difficult endeavor that requires consideration of various issues and challenges that comes with testing and measurement. Those designing, choosing, and implementing selection programs must strive to ensure that optimal decisions can be made based on the information that is collected for selection. In this chapter we explored some of the challenges that come with making important selection decisions with regards to selection tool properties and applicant behavior. We also proposed some ways to address these challenges. In the next chapter we look at some selection practices and methods used in fields outside of teaching and teacher education and how these challenges have been navigated.

References

Alliger, G. M., & Dwight, S. A. (2000). A meta-analytic investigation of the susceptibility of integrity tests to faking and coaching. *Educational and Psychological Measurement, 60*, 59–72.

Bardach, L., Rushby, J. V., Kim, L. E., & Klassen, R. M. (2021). Using video- and text-based situational judgement tests for teacher selection: A quasi-experiment exploring the relations between test format, subgroup differences, and applicant reactions. *European Journal of Work and Organizational Psychology, 30*, 251–264.

Barrick, M. R., Shaffer, J. A., & DeGrassi, S. W. (2009). What you see may not be what you get: Relationships among self-presentation tactics and ratings of interview and job performance. *Journal of Applied Psychology, 94*, 1394–1411.

Bauer, T. N., Truxillo, D. M., & Sanchez, R. J. (2001). Applicant reactions to selection: Development of the selection procedural justice scale (SPJS). *Personnel Psychology, 54*, 387–419.

Becker, B. J. (1990). Coaching for the scholastic aptitude test: Further synthesis and appraisal. *Review of Educational Research, 60*, 373–417.

Burns, G. N., & Christiansen, N. D. (2011). Methods of measuring faking behavior. *Human Performance, 24*, 358–372.

Chan, D., Scmitt, N., Jennings, D., Clause, C. S., & Delbridge, K. (1998). Applicant perceptions of test fairness: Integrating justice and self-serving bias perspectives. *International Journal of Selection and Assessment, 6*, 232–239.

Donovan, J. J., Dwight, S. A., & Hurtz, G. M. (2003). An assessment of the prevalence, severity, and verifiability of entry-level applicant faking using the randomized response technique. *Human Performance, 16*, 81–106.

Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, & Department of Justice. (1978). Uniform guidelines on employee selection procedures. *Federal Register, 43*, 38290–39315.

Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and Individual Differences, 7*(3), 385–400.

Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *Academy of Management Review, 18*, 694–734.

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment, 4*, 26–42.

Greenberg, J., & Cropanzano, R. (1993). The social side of fairness: Interpersonal and informational classes of organizational justice. In R. Cropanzano (Ed.), *Series in applied psychology. Justice in*

the workplace: Approaching fairness in human resource management (pp. 79–103). Lawrence Erlbaum Associates.

Griffin, B., Harding, D. W., Wilson, I. G., & Yeomans, N. D. (2008). Does practice make perfect? The effect of coaching and retesting on selection tests used for admission to an Australian medical school. *The Medical Journal of Australia, 189*, 270–273.

Griffin, B., & Wilson, I. G. (2012). Faking good: Self-enhancement in medical school applicants. *Medical Education, 46*, 485–490.

Griffith, R. L., Chmielowski, T., & Yoshita, Y. (2007). Do applicants fake? An examination of the frequency of applicant faking behavior. *Personnel Review, 36*, 341–355.

Hausknecht, J. P., Halpert, J. A., Di Paolo, N. T., & Moriarty Gerrard, M. O. (2007). Retesting in selection: A meta-analysis of coaching and practice effects for tests of cognitive ability. *Journal of Applied Psychology, 92*, 373–385.

Hogan, R., & Hogan, J. (2007). *Hogan Personality Inventory Manual* (Version 3rd). Tulsa, OK.

Holden, R. R. (2007). Socially desirable responding does moderate personality scale validity both in experimental and in nonexperimental contexts. *Canadian Journal of Behavioural Science. Revue Canadienne Des Sciences Du Comportement, 39*, 184–201.

Holden, R. R., Wood, L. L., & Tomashewski, L. (2001). Do response time limitations counteract the effect of faking on personality inventory validity? *Journal of Personality and Social Psychology, 81*(1), 160–169.

Hunter, J. E., & Schmidt, F. L. (1990). Dichotomization of continuous variables: The implications for meta-analysis. *Journal of Applied Psychology, 75*, 334–349.

Jackson, D. N., Wroblewski, V. R., & Ashton, M. C. (2000). The impact of faking on employment tests: Does forced choice offer a solution? *Human Performance, 13*, 371–388.

Klassen, R. M., Durksen, T., Rowett, E., & Patterson, F. (2014). Applicant reactions to a situational judgment test used for selection into initial teacher training. *International Journal of Educational Psychology, 3*, 104–124.

Levashina, J., & Campion, M. A. (2006). A model of faking likelihood in the employment interview. *International Journal of Selection and Assessment, 14*, 299–316.

Lievens, F., Buyse, T., Sackett, P. R., & Connelly, B. S. (2012). The effects of coaching on situational judgment tests in high-stakes selection. *International Journal of Selection and Assessment, 20*, 272–282.

Lievens, F., Patterson, F., Corstjens, J., Martin, S., & Nicholson, S. (2016). Widening access in selection using situational judgement tests: Evidence from the UKCAT. *Medical Education, 50*, 624–636.

Lievens, F., van Dam, K., & Anderson, N. (2002). Recent trends and challenges in personnel selection. *Personnel Review, 31*, 580–601.

Lievens, F., Sackett, P. R., & Zhang, C. (2021). Personnel selection: A longstanding story of impact at the individual, firm, and societal level. *European Journal of Work and Organizational Psychology, 30*, 444–455.

Lievens, F., & Sackett, P. R. (2017). The effects of predictor method factors on selection outcomes: A modular approach to personnel selection procedures. *Journal of Applied Psychology, 102*, 43–66.

Lind, E. A., & Tyler, T. R. (1988). *The Social Psychology of Procedural Justice.* Springer.

Maurer, T. J., Solamon, J. M., Andrews, K. D., & Troxtel, D. D. (2001). Interviewee coaching, preparation strategies, and response strategies in relation to performance in situational employment interviews: An extension of Maurer, Solamon, and Troxtel (1998). *Journal of Applied Psychology, 86*, 709–717.

Maurer, T. J., Solamon, J., & Troxtel, D. (1998). Relationship of coaching with performance in situational employment interviews. *Journal of Applied Psychology, 83*, 128–136.

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*, 1693–1725.

McFarland, L. A., & Ryan, A. M. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology, 85*, 812–821.

Melchers, K. G., Roulin, N., & Buehl, A.-K. (2020). A review of applicant faking in selection interviews. *International Journal of Selection and Assessment, 28*(2), 123–142.

Messick, S., & Jungeblut, A. (1981). Time and method in coaching for the SAT. *Psychological Bulletin, 89*, 191–216.

Moshinsky, A., Ziegler, D., & Gafni, N. (2017). Multiple Mini-Interviews in the age of the internet: Does preparation help applicants to medical school? *International Journal of Testing, 17*(3), 253–268.

Mueller-Hanson, R., Heggestad, E. D., & Thornton, G. C., 3rd. (2003). Faking and selection: Considering the use of personality from select-in and select-out perspectives. *Journal of Applied Psychology, 88*, 348–355.

Myers, I. B., McCaulley, M. H., Quenk, N. L., & Hammer, A. L. (1998). *MBTI manual: A guide to the development and use of the Myers-Briggs Type Indicator* (3rd ed.). Consulting Psychologists Press.

Patterson, F., Ashworth, V., Kerrin, M., & O'Neill, P. (2013). Situational judgement tests represent a measurement method and can be designed to minimise coaching effects. *Medical Education, 47*, 220–221.

Patterson, F., Cleland, J., & Cousans, F. (2017). Selection methods in healthcare professions: Where are we now and where next? *Advances in Health Sciences Education: Theory and Practice, 22*, 229–242.

Patterson, F., Zibarras, L., Carr, V., Irish, B., & Gregory, S. (2011). Evaluating candidate reactions to selection practices using organisational justice theory. *Medical Education, 45*, 289–297.

Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49–69). Erlbaum.

Paulhus, D. L., Harms, P. D., Bruce, M. N., & Lysy, D. C. (2003). The over-claiming technique: Measuring self-enhancement independent of ability. *Journal of Personality and Social Psychology, 84*, 890–904.

Pauls, C. A., & Crost, N. W. (2005). Effects of different instructional sets on the construct validity of the NEO-PI-R. *Personality and Individual Differences, 39*, 297–308.

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*, 153–172.

Pyburn, K. M., Ployhart, R. E., & Kravitz, D. A. (2008). The diversity–validity dilemma: Overview and legal context. *Personnel Psychology, 61*, 143–151.

Robertson, I. T., & Smith, M. (2001). Personnel selection. *Journal of Occupational and Organizational Psychology, 74*, 441–472.

Sackett, P. R., Lievens, F., Van Iddekinge, C. H., & Kuncel, N. R. (2017). Individual differences and their measurement: A review of 100 years of research. *Journal of Applied Psychology, 102*, 254–273.

Sackett, P. R., & Roth, L. (1996). Multi-stage selection strategies: A Monte Carlo investigation of effects on performance and minority hiring. *Personnel Psychology, 49*, 549–572.

Sackett, P. R., Schmitt, N., Ellingson, J. E., & Kabin, M. B. (2001). High-stakes testing in employment, credentialing, and higher education. Prospects in a post-affirmative-action world. *The American Psychologist, 56*, 302–318.

Salgado, J. F. (2003). Predicting job performance using FFM and non-FFM personality measures. *Journal of Occupational and Organizational Psychology, 76*, 323–346.

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*, 262–274.

Schmidt, F. L., Le, H., & Ilies, R. (2003). Beyond alpha: An empirical examination of the effects of different sources of measurement error on reliability estimates for measures of individual differences constructs. *Psychological Methods, 8*, 206–224.

Sheldon, M. S. (1959). Conditions affecting the fakability of teacher-selection inventories. *Educational and Psychological Measurement, 19*, 207–219.

Snell, A. F., Sydell, E. J., & Lueke, S. B. (1999). Towards a theory of applicant faking: Integrating studies of deception. *Human Resource Management Review, 9*, 219–242.

Truxillo, D. M., & Bauer, T. N. (2011). Applicant reactions to organizations and selection systems. Selecting and developing members for the organizationIn S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology* (Vol. 2, pp. 379–397). American Psychological Association.

Van Iddekinge, C. H., Raymark, P. H., & Roth, P. L. (2005). Assessing personality with a structured employment interview: Construct-related validity and susceptibility to response inflation. *Journal of Applied Psychology, 90*, 536–552.

Whetzel, D. L., McDaniel, M. A., & Nguyen, N. T. (2008). Subgroup differences in situational judgment test performance: A meta-analysis. *Human Performance, 21*, 291–309.

Ziegler, M., & Buehner, M. (2009). Modeling socially desirable responding and its effects. *Educational and Psychological Measurement, 69*, 548–565.

Ziegler, M., MacCann, C., & Roberts, R. D. (2011). Faking: Knowns, unknowns, and points of contention. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 3–16). Oxford University Press.

Chapter 5
Selection Practices and Methods in Other Fields

Abstract The research base for selection into teacher education programs and teaching practice is only recently emerging (Klassen & Kim, 2019; Klassen et al., 2017). In this light, reviewing selection practices and methods used in other fields—especially those where the methods are well-developed and well-researched—provides a lens through which to view and consider teacher selection. Various selection methods have been used to select individuals into educational (training) programs and into employment. Though the methods used in other fields have some degree of overlap with each other, each area also has its own distinct methods and research base that characterize the field. As such, in this chapter, we will review the practices and the evidence base for the methods that are used to select individuals into medical schools, law schools, and into large organizations.

5.1 Selection Practices in Other Fields

Selection into medical schools. Much research has been dedicated to investigating the validity of selection methods in order to identify candidates who are most likely to be successful in medical training and progress to become competent clinicians. One of the distinctive characteristics of medical school selection procedures is the quantity and quality of the evidence used to make the selection decisions. Using evidence-informed selection methods is important for medical school selection researchers and practitioners, at least partly due to the high levels of competition for admission and the evidence-based practice ethos prevalent in modern medicine.

Although selection approaches vary across medical schools and countries, a typical selection procedure consists of the following stages (shown in Fig. 5.1). First, prior academic records are used as a screening tool. Next, applicants' documents and test scores are reviewed, which often includes a review of their personal statements, references, and psychometric test scores. Psychometric tests can include aptitude tests to test academic attributes (cognitive attributes associated with academic skills and abilities), sometimes including tests of numerical ability and verbal ability. Non-cognitive attributes are often also assessed, which are attributes not associated with academic skills and abilities; otherwise known as non-academic competencies, 'soft'

R. M. Klassen and L. E. Kim, *Teacher Selection: Evidence-Based Practices*,
https://doi.org/10.1007/978-3-030-76188-2_5

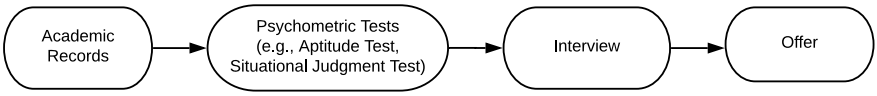


Fig. 5.1 A simplified model of a medical school selection process

skills, or people skills. Situational judgment tests (SJTs; discussed in a later section of the chapter) are increasingly being used at this stage of selection to assess non-cognitive attributes. After first-stage or screening tests, eligible applicants are then invited to the final stage of selection, which is usually an interview of some format such as a multiple mini-interview (MMI), although assessment centers (with multiple activities) are also sometimes used. Each medical school weights the scores from the various stages differently. From this total score, a candidate pool is formed and training places are offered.

Selection into law schools. In some countries, passing the law school selection process is claimed to be the most difficult stage in entering the legal profession (Shultz & Zedeck, 2012). Unlike research in medical school selection where the validity and reliability of a variety of selection methods have been widely explored by multiple researchers, such breadth of research seems to be missing in law school selection. Existing research is largely focused on the academic predictors of first year GPA during law school and not so much on the non-cognitive predictors. Moreover, there is limited research that considers a range of outcomes beyond academic success in law school, such as job performance and job satisfaction.

Unlike the multi-stage process in medical school selection, law school selection often consists of one stage—the assessment of prior academic records and an aptitude test (see Fig. 5.2). Since the 1940s, two scores have largely determined law school selection decisions in the US: undergraduate grade point average (UGPA, as a measure of academic records) and the Law School Admission Test (LSAT, as a measure of aptitudes; Shultz & Zedeck, 2012). Both of these measures are assessments of applicants' academic attributes. Like medical school selection, each law school determines the weighting of each of the scores (i.e., UGPA and LSAT).

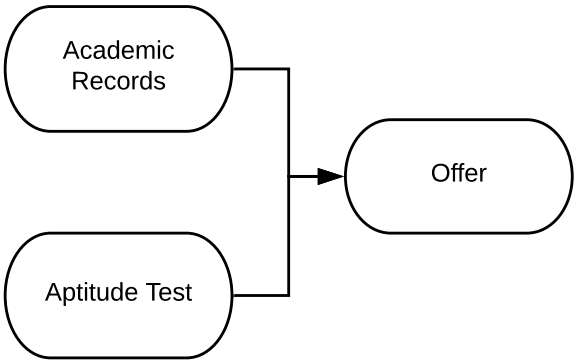


Fig. 5.2 A simplified model of law school selection process

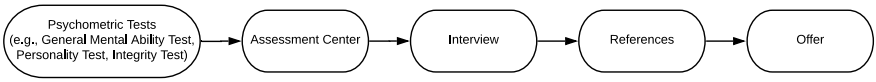


Fig. 5.3 A simplified model of personnel selection process

Personnel selection in large organizations. Selection programs in large organizations often include a wide variety of predictors and a great range of outcome measures. Organizations often tailor their selection methods to their unique organizational factors, such as size, context, and the nature of the tasks required by the various jobs. It is sometimes the case that employers choose their selection methods based on idiosyncratic preferences or tradition rather than the reliability and validity of the methods (Anderson & Witvliet, 2008).

Despite the absence of a commonly agreed set of selection methods, many organizations follow similar stages of practice for selection (Fig. 5.3). After an initial screen of resumés and job applications to assess applicants' suitability for the job, applicants may be invited to complete some psychometric tests, such as those assessing general mental ability, personality, and/or integrity. Applicants may then be invited to assessment centers, in which they demonstrate how they would handle various work situation through a range of activities such as group tasks, group interviews, and individual tasks. Individual interviews often form the last stage of selection, after which applicants' references are checked for cross-verification purposes.

5.2 Selection Methods and Their Evidence

Multiple review articles and meta-analyses have been published on the validity of selection methods used in each field (e.g., Lievens et al., 2021; Patterson et al., 2016; Shultz & Zedeck, 2012). A selection method can be used to select *out* individuals (e.g., a screening of academic records to reduce the pool), select *in* individuals (e.g., conducting structured interviews to make selection decisions), or to verify applicant details (e.g., reference checks). In the next section, we will summarize the evidence behind a range of selection methods that are used frequently in each of the three fields of interest: selection into training programs in medicine and law, and selection into large organizations (personnel selection). A summary of the evidence can be found in Table 5.1.

Academic records. Records of prior academic achievement are frequently considered as part of selection decision-making for educational programs and employment. Forms of academic records used for selection include grades from secondary school, if applying for undergraduate programs; and grades from both secondary school and from undergraduate study, if applying for graduate school programs or jobs.

Academic records are generally used for selection in two ways. First, academic records are used as an indicator of cognitive ability or used in conjunction with other information (e.g., resumés) that highlight relevant experience. Here, a record

Table 5.1 An interpretation of the wider literature on various selection methods

Selection method	Summary of evidence
Academic records	Most evidence is from educational program selection, with moderate predictive validity. Issues of fairness for minority ethnic groups
Aptitude tests	Most evidence is from education program selection. Good evidence of predictive validity and reliability. Issues of fairness for minority ethnic groups
General Mental Ability (GMA) tests	Most evidence is from personnel selection. Good evidence of predictive validity and reliability. Issues of fairness for minority ethnic groups
Personality tests	Most evidence is from personnel selection. Good evidence of predictive validity, reliability, and fairness
Integrity tests	Most evidence is from personnel selection. Good evidence of predictive validity and reliability
Situational judgment tests	Most evidence is from medical education and personnel selection. Good evidence of predictive validity and fairness. Low-to-moderate reliability given their multidimensionality
Interviews (including MMIs)	Structured interviews (particularly MMIs) have strong evidence of predictive validity, reliability, and fairness
Assessment centers	Most evidence is from personnel selection. High face validity but little incremental validity. Some evidence of reliability. Issues of fairness for minority ethnic groups
Reference checks	Used mostly for face validity. Limited evidence of predictive validity and fairness. Low reliability

of minimum level of academic achievement is used for screening. Second, academic records are used to give priority to applicants with higher levels of academic achievement and used as an indicator of the quality of the accepted applicants.

For medical school selection, most studies indicate that applicants' academic records are a good predictor of academic and clinical performance. For example, studies found that applicants' prior academic records reliably predicted various measures of success during and after medical training, including medical school academic grades (McManus et al., 2013), licensing examination scores (Julian, 2005), internship performance (Ferguson et al., 2002), and career progression (McManus et al., 2003). Moreover, a meta-regression study found that secondary school academic records in the UK (A-levels) were a stronger predictor of medical students' first year academic performance than aptitude test scores (McManus et al., 2013). However, the strength of prediction declined throughout the undergraduate and postgraduate years, though it was still statistically significant.

First year academic performance in law school is the criterion outcome that a majority law school selection studies use to assess the validity of selection tools. Studies have found that undergraduate academic performance was a relatively good predictor of performance in first year of law school, with correlations ranging from

0.26 to 0.29 (Law School Admission Council, 2019). Such empirical relationships are unsurprising since it is established that an academic measure is moderately associated with future measures of academic success (e.g., Geiser & Santelices, 2007). Unlike research in medical school selection research, limited research has been conducted on how law students' UGPA predicts outcomes beyond law school. An exception is a study by Lempert et al. (2000), which found that law students' UGPA did not predict post-law school measures of success (i.e., income, career satisfaction, and service contributions).

Academic records are sometimes used for personnel selection. Although sparsely researched, there is evidence indicating that academic records can add useful information for selection. A meta-analytic study by Roth et al. (1996) found that academic records were associated with job performance with a corrected correlation in the mid 0.30. Moreover, the relationship was stronger the closer to the time job performance was measured to the time of the GPA, which indicate that academic records are reasonable proximal predictors (i.e., they are better at predicting outcomes the closer they are to each other temporally). However, it has been noted that using college academic performance for selection may create an adverse impact given issues of ethnic group differences in GPA scores (Roth & Bobko, 2000; see Chap. 4 for discussion on adverse impact).

In sum, academic records seem to be most helpful for predicting outcomes that are academic in nature, though care must be taken for selection use as it can disadvantage certain minority groups. Furthermore, since the sizes of the associations tend to be stronger the closer the timing is between the measurement of the academic record and the criterion outcome, considering the time lag between two measurements should be considered.

Psychometric tests. Using psychometric tests can be an efficient way to screen applicants before more resource-intensive selection methods (e.g., interviews or assessment centers) are used to further select individuals to the next stages of the selection process. Psychometric tests can be relatively easy to administer and score, especially when they are computer-administered. There are generally two types of psychometric tests: tests that assess one's academic attributes (e.g., aptitude tests, general mental ability tests) and tests that assess one's non-academic attributes (e.g., personality tests, integrity test, and situational judgment tests). The evidence behind each of these methods will be outlined below.

Aptitude tests. Aptitude tests are assessments that examine qualities important for a skill, job, and/or field. In personnel selection, aptitude tests that examine specific aptitudes (e.g., verbal ability and numerical ability) as well as general aptitudes (i.e., general mental ability; discussed in a later section of the chapter) are used. In medical and law school selection, aptitude tests tend to measure multiple aptitudes, including subject-specific areas. In this section, we will focus on the aptitude tests used in medical school and law school selection.

In medical school selection, the aptitude tests aim to assess the academic and non-cognitive attributes associated with medical school academic and clinical performance. The type of aptitude test that applicants must sit depends on the country in which they wish to study medicine, and sometimes the level of entry. For example,

applicants for British medical schools must sit either the University Clinical Aptitude Test (UCAT) or the Biomedical Admissions Test (BMAT), whereas applicants for Australian medical schools must sit either the Undergraduate Medicine and Health Sciences Admissions Test (UMAT) for undergraduate medical programs or the Graduate Medical Schools Admissions Test (GAMSAT) for graduate medical programs. In North America, the Medical College Admissions Test (MCAT) is most commonly used for admissions. These tests share common ground in assessing a range of similar aptitudes but also have varying emphases on different aptitudes. For example, the UCAT includes subtests of *verbal reasoning*, *decision making*, *quantitative reasoning*, *abstract reasoning,* and *situational judgment*, but no specific scientific knowledge (UCAT, 2021). The MCAT includes questions that assess *critical analysis and reasoning skills,* but also those that assess knowledge of concepts and principles associated with medicine (e.g., biology, biochemistry, and psychobiology; AAMC, 2021).

Despite the wide usage of aptitude tests to select individuals into medical schools, the evidence of their validity is mixed. Some studies report that aptitude tests do predict academic and clinical performance (e.g., Puddey & Mercer, 2014) and do so above and beyond prior academic records (e.g., McManus et al., 2013; Sartania et al., 2014). On the other hand, some studies report that aptitude tests do not predict academic and clinical performance (e.g., Yates & James, 2010). Discrepancies in the predictive validity of medical school entrance exams indicate that closer examinations of the content of aptitude test sections may be needed.

Research on students' perceptions of the fairness and usefulness of medical school selection aptitude tests seem to report mixed findings. Studies have reported that students viewed aptitude tests as neither fair nor useful (Dhar et al., 2012), some viewed them as generally useful and suitable for selection (Cleland et al., 2011), while others viewed only particular sections of the aptitude tests (e.g., logical reasoning and problem solving, or interpersonal understanding) as useful and well-designed (Stevens et al., 2014). In a summary of evidence of medical school selection tools, Patterson et al. (2016) reported that evidence on the fairness and usefulness of aptitude tests seemed mixed. They advised that closer examination may be necessary for each section of the aptitude tests as well as for specific types of aptitude tests (e.g., UCAT, BMAT) rather than generalizing to all aptitude tests. Nevertheless, they concluded that aptitude tests are good screeners for medical school selection.

For law school selection, the type of aptitude test used varies by country and institution. In the US, Canada, and some universities in Australia, the Law School Admission Test (LSAT) is a required component of the admissions process. In the UK, the Law National Aptitude Test is used for admissions by some law schools. As LSAT is the most researched type of law school aptitude test, the LSAT will be the focus of this section.

The LSAT has been in use since 1948 (LaPiana, 2004) and consists of sections testing *analytical and logical reasoning skills*, *reading comprehension skills*, and *writing skills* (Law School Admissions Council, 2021). The LSAT was designed to predict first year GPA (LaPiana, 2004), which is supported by findings that the correlation between average LSAT and first year GPA ranges between 0.34 and 0.41,

compared to the correlation between UGPA and first year GPA that ranges between 0.26 and 0.29 (Law School Admissions Council, 2019). Although academic attributes are known to predict job performance and job knowledge in general, there is limited evidence that LSAT scores predict performance in law school beyond the first year in the program.

The issue of fairness in using LSAT scores to make law school selection decisions has been a source of debate (Holmquist et al., 2014), largely due the findings that LSAT scores seem to differ between the ethnicities of the applicants. For example, findings from the LSAT administered between 2007–2008 and 2013–2014 indicate that Caucasian test-takers consistently had the highest mean scores, followed by Asian/Pacific Islander test-takers, with Puerto Rican test-takers consistently had the lowest (Dalessandro et al., 2014). Associated with this issue is the finding that LSAT scores tend to overpredict first year GPA performance for minority test-takers, indicating that LSAT scores may have differential powers in predicting academic performance depending on the test-takers' race/ethnicity (see Kidder, 2001 for review of studies). As such, heavy emphasis on the LSAT for admissions has been identified as a factor that is restricting the diversity of individuals in legal education and in the legal profession. Including assessments of non-academic attributes as part of the aptitude test suite has been identified as a way to reduce these adverse effects (Holmquist et al., 2014).

General mental ability tests. Psychometric tests of general mental ability (GMA; otherwise known as general cognitive ability or intelligence) aim to assess a composite of multiple cognitive abilities. GMA has commonly been used in personnel selection, although aspects of GMA are included in admissions tests for medical schools and law schools. Multiple meta-analyses (e.g., Lang et al., 2010) indicate that GMA predicts both overall job performance and specific job performance dimensions (e.g., task performance, contextual performance, counterproductive work behavior). However, GMA is not as predictive for less complex jobs (Gottfredson, 1997). Nevertheless, tests of GMA seem to be generalizable and valid predictors of a range of outcomes, including academic performance, career potential, creativity, and job performance (Kuncel et al., 2004). The reliability of GMA tests is one of the highest of the selection methods, often in the 0.80 to 0.90 (Ones et al., 2012). Using multiple cognitive ability tests (e.g., numerical ability, literacy skills) to capture GMA is encouraged rather than a reliance on a single test (Lubinski, 2000). Overall, GMA tests seem to be useful for inclusion in selection programs given that cognitive ability is a good predictor of multiple outcomes.

Personality tests. Personality tests aim to examine the underlying non-cognitive attributes influencing thoughts, feelings, and behaviors (John et al., 2008). In the context of medical school selection, whether personality tests should be used for selection has been a source of debate (Patterson et al., 2016). On the one side, the 'explicit' nature of many personality tests is concerning. As applicants can often know what each of the items are measuring, issues such socially desirable response patterns arise (e.g., Lievens & Sackett, 2017). However, there is evidence that medical school students' personality is associated with their outcomes during medical training and beyond. For example, McLarnon et al. (2017) examined the utility of including

personality in addition to the traditional predictors for medical school selection (i.e., MCAT, GPA, and scores from a semi-structured panel interviews). They found that personality predicted medical school academic performance and clinical performance above and beyond these traditional predictors and the regression coefficients were stronger. Furthermore, personality was the only significant predictor of clinical performance, providing validity evidence for the use of personality for selection. Given the empirical relationship between personality and important outcomes in medicine, more research into their potential use in medical school selection is warranted.

Personality tests have rarely been included in research on law school admissions. An exception is a study by Shultz and Zedeck (2011), who found that the personality test subscale scores correlated with more lawyer effectiveness criterion outcomes than the LSAT and UGPA scores, and the size of the correlations was bigger. This finding suggests that personality tests could potentially play a role in predicting important lawyer outcomes that are not adequately accounted for by academic attributes alone, and hence worth considering as part of a suite of law school selection tests (Holmquist et al., 2014).

In contrast to research in medical school and law school selection, personnel selection has a long history of using personality tests as part of the selection procedure. Its use is supported by numerous meta-analytic findings that conscientiousness, a domain of the Big Five personality framework, consistently predicts job performance (e.g., Judge et al., 2013). Contrary to common belief though, the relationship between personality and job performance may not be linear but curvilinear (Le et al., 2011). This finding suggests that very high scores on personality tests are not necessarily optimal for higher performance, which has implications for how personality scores are used for selection. Another consideration is the potential for adverse impact. Though there are only modest variations of personality across racial groups (Hough et al., 2001), there can still be adverse impact depending on how personality scores are used (Risavy & Hausdorf, 2011), which warrants further research.

Integrity tests. Integrity tests are used more often for personnel selection than for selection into medical and law schools. Integrity tests measure applicants' levels of honesty, which can be considered a hybrid of personality factors including conscientiousness, agreeableness, and adjustment (Ones & Viswesvaran, 1998). There are two types of integrity tests: overt tests and personality-based tests (Wanek et al., 2003). Overt tests explicitly examine test-takers' levels of honesty, their attitudes about theft, and their past cases of theft. Personality-based tests are a less explicit assessment of antecedent qualities associated with dishonest behaviors, assessing qualities such as dependability, trouble with authority, and hostility. A meta-analysis indicated that overt integrity tests are more closely associated with job performance than personality-based integrity tests (Van Iddekinge et al., 2012). The focus of research on integrity tests has primarily been on their relationship with counterproductive work behaviors (negative behaviors at work, such as theft and lying), with results showing moderate correlations with job performance, training performance,

and turnover. A review on integrity tests found that applicants seem to report reasonably positive reactions to the test (Berry et al., 2007). Thus, integrity tests seem to be particularly useful for personnel selection.

Situational judgment tests. Situational judgment tests (SJTs) are a scenario-based method used to measure a range of attributes, especially non-cognitive (non-academic) attributes such as interpersonal skills, contextualized judgment, and other work-related 'soft skills'. SJTs present a series of context-related scenarios and ask applicants for their responses to the scenarios (Motowidlo et al., 1990). There are a range of design possibilities, including in the response instructions (e.g., asking what the applicants *would do* or *should do),* and response format (e.g., multiple-choice options or Likert-scale options; e.g., de Leng et al., 2018). Increasingly, digitalized and gamified versions are becoming more available (Gkorezis et al., 2021) with the increasing uptake of digital selection procedures (see Woods et al., 2020 for a review).

In medicine, SJTs are widely used for selection purposes with substantial evidence to support their use (Webster et al., 2020). SJTs assessing the ability to regulate emotions have been shown to predict academic performance in medical training, above and beyond cognitive ability and conscientiousness (Libbrecht et al., 2014). Similarly, an SJT assessing interpersonal skills was shown to predict academic and clinical performance in medicine, above and beyond cognitive ability (Lievens, 2013). Both text-based and video-based SJT formats have been trialed in medicine and video-based formats have demonstrated higher predictive validity (Lievens & Sackett, 2006) and generally received more favorable evaluations (Kanning et al., 2006).

Limited research has been dedicated to exploring how SJTs can be used for law school selection. One study, however, found that an SJT measuring non-cognitive qualities associated with lawyer effectiveness correlated with 23 of 26 lawyer effectiveness criterion outcomes, with correlations ranging from 0.11 to 0.21 (Shultz & Zedeck, 2011). Here, SJT scores were associated with more outcomes than LSAT scores and the size of the correlations appeared to be bigger, suggesting that research in this area may be worth conducting.

Large organizations have used SJTs for personnel selection for decades, using the method to assess a wide range of constructs, including interpersonal skills, leadership, personality, and heterogeneous composites of multiple constructs. A meta-analysis reported that the associations between SJTs and job performance ranged from 0.19 (for job knowledge and skills) to as high as 0.43 (for personality composites; Christian et al., 2010). In line with medical school selection research findings, video-based SJTs demonstrated stronger criterion-related validity than text versions (Christian et al., 2010). Given the multidimensional nature of SJTs, their internal consistency is often low to moderate (Lievens et al., 2008). However, there are minimal group differences (including gender and race) on SJT performance (Whetzel et al., 2008). More information on SJTs, especially pertaining to teacher selection, can be found in Chaps. 7 and 9.

Interviews. Interviews are face-to-face interactions between one or more interviewers with one or more interviewees. There are two broad types of interview formats: structured and unstructured. Structured interviews consist of pre-determined

questions with set scoring keys that are consistent across applicants. Unstructured interviews have no set format: the questions can differ between applicants and the scoring of the applicants are often overall impressions with no set scoring key. Panel interviews are an interview format with multiple interviewers and, although they are often perceived to have higher reliability and validity than individual interviews, evidence for this seems to be inconclusive (Dixon et al., 2002).

Interviews have been long used for medical school selection. The interviewers for medical school selection may be faculty members of the medical school and/or community members who are given interview training by medical school selection staff. Despite the popular use of interviews, especially in the final stage of medical school selection, past research shows that the reliability of structured interviews used for medical school selection can be low to moderate (Kreiter et al., 2004). This finding is surprising because interviews have high face validity and are widely used across numerous medical schools globally.

Interviews are infrequently used in law school selection, but they are one of the most frequently used methods for personnel selection (Anderson & Witvliet, 2008). Structured interviews show higher validity and reliability than unstructured interviews (Posthuma et al., 2002). There are only small inter-group differences when using structured interviews, with particularly low differences for interviews focusing on behavior (Moscoso, 2000). Applicants perceive interviews as a fairer method of selection than other methods (Hausknecht et al., 2004), with applicants typically expecting that interviews will take place as part of the selection process (Lievens et al., 2003). From the employers' perspective, interviews are seen as a chance to assess the applicants' social and communication skills.

Multiple mini-interviews (MMIs). A specific type of interview, multiple mini-interviews (MMIs; see Chap. 8 for discussion on applications in teacher selection), are increasingly being adopted to replace traditional interviews for medical selection (Patterson et al., 2016). The underlying assumption of MMIs is that a greater sampling of behaviors provides more information about the suitability of applicants and increases the reliability of the interview process. In MMIs, applicants rotate through 5 to 12 stations for brief interviews (typically 5–10 min) and to complete various tasks (Gafni et al., 2012). These stations differ in the type of interview administered. For example, one station may require applicants to describe what they would do in certain situations (situational judgment stations), another station might require applicants to describe what they did in the past in a particular situation (behavioral interview station), and another might ask the applicant to engage in a conversation with an actor playing the role of a patient. Of these types, behavioral interview stations seem to best differentiate among applicants (Eva & Macala, 2014).

Investigating MMI's overall validity in medical school selection, Eva et al. (2012) found that those who were rejected at an institution because of low MMI scores also received lower scores on the Canadian national licensing examinations than those who were not rejected given their MMI scores. Furthermore, the magnitude and the direction of MMI's concurrent validity with a national medical school aptitude test seemed to differ depending on the content of the aptitude test. Specifically, there was a small positive correlation with a section on reasoning in humanities and

social sciences (0.26) and a small negative correlation with a section on reasoning in biological and physical sciences (−0.15), suggesting a multi-faceted relationship with different aptitudes (Roberts et al., 2008).

MMIs seem to be moderately to highly reliable (Rahim & Yusoff, 2016) and are favorably perceived by both applicants and examiners (Eva et al., 2004). Nevertheless, sufficient interviewer training and consistent scoring scheme is necessary given the inevitable subjective nature of interviews in general. All in all, MMIs appear to be a promising interview format for medical school selection with the potential for application in other fields, including teacher education.

Assessment centers. Assessment centers (ACs) are a suite of individual and group exercises that aim to assess a variety of applicant attributes that relate both directly and indirectly to a particular training or job opportunity. ACs can consist of group exercises, written/in-tray tasks, oral presentations, and interactive exercises. The AC format has not been widely explored for entrance into law school or medical education; however, some similar formats have been used for selection into postgraduate medical training (e.g., Randall, Davies et al., 2006; Randall, Stewart et al., 2006). Further evidence is needed before the use of ACs could be widely endorsed for medical school selection (Patterson et al., 2016).

Since their introduction over 50 years ago, ACs have been extensively used for personnel selection given their high face validity (Sackett & Lievens, 2008). A meta-analysis reported that ACs most commonly assess six dimensions of cognitive and non-cognitive attributes: *consideration of others*, *communication skills*, *motivation*, *persuasive power*, *organization and planning skills*, and *problem-solving skills* (Arthur et al., 2003). One of the barriers to using ACs is the high cost associated with administering this type of selection method. Job analyses are typically conducted to identify important dimensions of jobs, followed by training of multiple assessors who would rate applicants' performance in each of the identified dimensions (International Taskforce on Assessment Center Guidelines, 2015). As with interviews, highly trained assessors are needed to ensure that the selection method is useful in making predictions about future behaviors and outcomes. ACs tend to have weak measurement properties, especially in relation to the reliability of the scoring (Jackson et al., 2016). Moreover, in Schmidt and Hunter's (1998) meta-analysis, AC methods did not demonstrate significant incremental validity when combined with GMA in predicting overall job performance (a gain of 0.02 points). Furthermore, a meta-analysis reported that the standardized differences between ethnic subgroups (d) can be large, with White applicants scoring higher than Black applicants ($d = 0.52$) and Hispanic applicants ($d = 0.28$; Dean et al., 2008). Thus, more research to ensure the rigorousness and fairness of the method and its scoring may be necessary.

Reference checks. References provided by applicants typically include letters (or contact details) from individuals who know the applicant from a personal or professional context. These are used to verify the character (e.g., teamwork skills, organization skill) and work history of the applicants.

Reference checks are widely used in medical school selection (Kuncel et al., 2014). However, evidence of their predictive validity is mixed. Some researchers have found evidence supporting their predictive validity (e.g., DeZee et al., 2014)

but other researchers found inconsistent or no evidence in predicting medical school performance (e.g., Poole et al., 2009). However, it difficult to differentiate applicants using references. Most references provided are exclusively or mostly positive, with any negative indicators given 'in code'—a referee may have sent a 'hidden message' but is missed by the assessor, which raises questions as to the appropriateness of including reference checks as part of medical school selection procedure (Stedman et al., 2009).

There is scant research examining how law schools use references as part of their selection system. We can, however, draw from the literature on how references are used for college and graduate admission generally. A meta-analysis on the relationship between references and undergraduate, graduate, and medical school performance found that there are modest correlations and little incremental validity over traditional academic predictors (Kuncel et al., 2014).

Reference checks add very little to predicting overall job performance when combined with GMA (Schmidt & Hunter, 1998). Their low predictive validity may be the result of uncertainty about what content should be assessed and how the content, typically positive, can be used to differentiate between applicants. Furthermore, low inter-rater reliability (around 0.40) means that there is a problem in the trustworthiness of the data from reference letters (Kuncel et al., 2014). This issue is in addition to limited research on adverse impact for particular groups and limited evidence supporting the use of reference letters except for their face validity. Thus, if references are used for selection, they may be useful for cross-checking or confirming applicant details at the final stages of selection rather than as a criterion for progression to a subsequent stage of selection.

5.3 What Can We Learn from Selection in Other Fields?

We have reviewed the common methods used in medical school selection, law school selection, and personnel selection. What can we learn from these fields that we can apply to teacher selection?

1. Use multiple selection methods. There is no magical selection method that can do everything. That is, one selection method should not be used as the only method to choose individuals into teacher training or employment. Rather, a carefully designed selection procedure consisting of multiple methods is best. Using multiple selection methods can: (a) allow assessments of multiple important constructs, (b) allow certain selection methods to 'select in' or 'select out' applicants at various stage of the selection procedure, (c) reduce costs as certain selection methods can be used to mass-screen a large pool of applicants so that a smaller pool of applicants can be invited to undergo more intensive (and expensive) selection methods, and (d) increase predictive validity by using multiple predictors rather than a single predictor.

2. Use evidence-informed selection methods. It is easy to include selection methods that have been used in the field for a long time and that are easy to administer. However, as we have reviewed in this chapter, there is varying strength of

evidence for each selection method. It is best to use selection methods that have the strongest evidence behind them (see Chap. 4 for issues to consider in decision-making). Examining the evidence that is available within and outside of the target field helps develop a strong theoretical and empirical rationale for the use of particular selection methods.

3. Distinguish between selection method and selection constructs. Before a selection method is chosen, consideration should be given to decisions about which constructs the selection process is targeting for assessment. After the constructs of interest are identified, one should assess which methods can best assess these constructs. For example, if communication skills are the target construct, one should consider a variety of methods that can assess this construct, such as MMIs and SJTs, and compare the evidence behind each of the methods. After choosing the selection method(s), one should make sure that the construct is indeed featured as a criterion in the procedure. For example, if communication skills are the construct of interest and interviews were chosen as a method to assess this, communication skills need to be explicitly featured in the structure of the interviews and the scoring criteria.

4. Include structure in interviews. Interviews are one of the most popular selection methods, with considerable evidence supporting the use of structured interviews over unstructured interviews. Unfortunately, structured interviews are less often used in practice for personnel selection (Lievens & De Paepe, 2004). Some of the explanations given by human resources personnel chime with why unstructured interviews are more frequently used in teacher selection; for example, interviewers' desire to establish an informal contact with the interviewees and to have greater discretion over the interview questions. Furthermore, conducting highly structured interviews can be quite costly both in time and money as they require additional processes, such as a job-analysis to form the interview questions, formulation of scoring procedures, and interviewer training to ensure standard scoring procedures. In contrast to common concerns, there is considerable flexibility when creating structured interviews (Levashina et al., 2014). Although it may be easy for interviewers to feel that unstructured interviews give a greater 'feel' for the applicants, the evidence is clear: structured interviews have greater predictive validity evidence than unstructured interviews. In the future, MMIs may become a more common method of implementing structured interviews in teacher selection (see Chap. 8).

5.4 Chapter Summary

To improve teacher selection methods, it is important to examine research from multiple contexts, including from fields outside of education, where there are existing, and often stronger, research foundations. This chapter reviewed the selection methods used for entrance into medical schools, law schools, and for employment in organizations (personnel selection). We outlined that the research base for different selection methods varies widely—from those with little evidence (e.g., reference checks) to those with stronger evidence (e.g., SJTs and MMIs). The combination and weighting

of the methods included in a selection program should be carefully considered and scrutinized against the needs of the program or organization, and resources that are available. In the next chapter, we will turn specifically to the selection of prospective teachers and explore historical and current research and practices in this field.

References

AAMC (2021). *What's on the MCAT exam?* Retrieved from https://students-residents.aamc.org/applying-medical-school/article/whats-mcat-exam/

Anderson, N., & Witvliet, C. (2008). Fairness reactions to personnel selection methods: An international comparison between the Netherlands, the United States, France, Spain, Portugal, and Singapore. *International Journal of Selection and Assessment, 16*, 1–13.

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*, 125–153.

Berry, C. M., Sackett, P. R., & Wiemann, S. (2007). A review of recent developments in integrity test research. *Personnel Psychology, 60*, 271–301.

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*, 83–117.

Cleland, J. A., French, F. H., Johnston, P. W., & Scottish Medical Careers Cohort Study Group. (2011). A mixed-methods study identifying and exploring medical students' views of the UKCAT. *Medical Teacher, 33*, 244–249.

Dalessandro, S. P., Stilwell, L. A., & Reese, L. M. (2014). *LSAT performance with regional, gender, and racial/ethnic breakdowns: 2007–2008 through 2013–2014 testing years (TR 14–02).* Law School Admission Council Research Rep. No. TR-00–01. Retrieved from http://www.lsac.org/docs/default-source/research-(lsac-resources)/tr-14-02.pdf

Dean, M. A., Roth, P. L., & Bobko, P. (2008). Ethnic and gender subgroup differences in assessment center ratings: A meta-analysis. *Journal of Applied Psychology, 93*(3), 685–691.

de Leng, W. E., Stegers-Jager, K. M., Born, M. P., & Themmen, A. P. N. (2018). Influence of response instructions and response format on applicant perceptions of a situational judgement test for medical school selection. *BMC Medical Education, 18*(1), 282.

DeZee, K. J., Magee, C. D., Rickards, G., Artino, A. R., Jr., Gilliland, W. R., Dong, T., & Durning, S. J. (2014). What aspects of letters of recommendation predict performance in medical school? Findings from one institution. *Academic Medicine: Journal of the Association of American Medical Colleges, 89*, 1408–1415.

Dhar, D., Perry, W. R. G., & Poole, P. (2012). Students' perceptions of the Undergraduate Medicine and Health Sciences Admissions Test (UMAT). *The New Zealand Medical Journal, 125*(1361), 29–36.

Dixon, M., Wang, S., Calvin, J., Dineen, B., & Tomlinson, E. (2002). The Panel Interview: A Review of Empirical Research and Guidelines for Practice. *Public Personnel Management, 31*, 397–428.

Eva, K. W., & Macala, C. (2014). Multiple mini-interview test characteristics: 'tis better to ask candidates to recall than to imagine. *Medical Education, 48*(6), 604–613.

Eva, K. W., Reiter, H. I., Rosenfeld, J., Trinh, K., Wood, T. J., & Norman, G. R. (2012). Association between a medical school admission process using the multiple mini-interview and national licensing examination scores. *Journal of the American Medical Association, 308*, 2233–2240.

Eva, K. W., Rosenfeld, J., & Reiter, H. I. (2004). An admissions OSCE: The multiple mini-interview. *Medical Education, 38*, 314–326.

Ferguson, E., James, D., & Madeley, L. (2002). Factors associated with success in medical school: Systematic review of the literature. *BMJ, 324*(7343), 952–957.

Gafni, N., Moshinsky, A., Eisenberg, O., Zeigler, D., & Ziv, A. (2012). Reliability estimates: Behavioural stations and questionnaires in medical school admissions. *Medical Education, 46*, 277–288.

Geiser, S., & Santelices, M. V. (2007). Validity of high-school grades in predicting student success beyond the freshman year: High-school record vs. standardized tests as indicators of four-year college outcomes. *UC Berkley: Center for Studies in Higher Education.* Retrieved from https://escholarship.org/uc/item/7306z0zf

Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence, 24*, 79–132.

Gkorezis, P., Georgiou, K., Nikolaou, I., & Kyriazati, A. (2021). Gamified or traditional situational judgement test? A moderated mediation model of recommendation intentions via organizational attractiveness. *European Journal of Work and Organizational Psychology, 30*, 240–250.

Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology, 57*, 639–683.

Holmquist, K., Shultz, M., Zedeck, S., & Oppenheimer, D. (2014). Measuring merit: The Shultz-Zedeck research on law school admissions. *Journal of Legal Education, 63*, 565–584.

Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection and amelioration of adverse impact in personnel selection procedures: Issues, evidence and lessons learned. *International Journal of Selection and Assessment, 9*(1–2), 152–194.

International Taskforce on Assessment Center Guidelines. (2015). Guidelines and ethical considerations for assessment center operations. *Journal of Management, 41*(4), 1244–1273.

Jackson, D. J. R., Michaelides, G., Dewberry, C., & Kim, Y.-J. (2016). Everything that you have ever been told about assessment center ratings is confounded. *Journal of Applied Psychology, 101*(7), 976–994.

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big-Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (pp. 114–158). Guilford Press.

Judge, T. A., Rodell, J. B., Klinger, R. L., Simon, L. S., & Crawford, E. R. (2013). Hierarchical representations of the five-factor model of personality in predicting job performance: Integrating three organizing frameworks with two theoretical perspectives. *Journal of Applied Psychology, 98*(6), 875–925.

Julian, E. R. (2005). Validity of the Medical College Admission Test for predicting medical school performance. *Academic Medicine: Journal of the Association of American Medical Colleges, 80*(10), 910–917.

Kanning, U. P., Grewe, K., Hollenberg, S., & Hadouch, M. (2006). From the subjects' point of view. *European Journal of Psychological Assessment, 22*(3), 168–176.

Kidder, W. (2001). Does the LSAT mirror or magnify racial and ethnic differences in educational attainment?: A study of equally achieving "Elite" college students. *California Law Review, 89*(4), 1055–1124

Klassen, R. M., Durksen, T. L., Kim, L. E., Patterson, F., Rowett, E., Warwick, J., Warwick, P., & Wolpert, M.-A. (2017). Developing a proof-of-concept selection test for entry into primary teacher education programs. *International Journal of Assessment Tools in Education, 4*, 96–114.

Klassen, R. M., & Kim, L. E. (2019). Selecting teachers and prospective teachers: A meta-analysis. *Educational Research Review, 26*, 32–51.

Kreiter, C. D., Yin, P., Solow, C., & Brennan, R. L. (2004). Investigating the reliability of the medical school admissions interview. *Advances in Health Sciences Education: Theory and Practice, 9*(2), 147–159.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology, 86*(1), 148–161.

Kuncel, N. R., Kochevar, R. J., & Ones, D. S. (2014). A meta-analysis of letters of recommendation in college and graduate admissions: Reasons for hope. *International Journal of Selection and Assessment, 22*(1), 101–107.

Lang, J. W. B., Kersting, M., Hülsheger, U. R., & Lang, J. (2010). General mental ability, narrower cognitive abilities, and job performance: The perspective of the nested-factors model of cognitive abilities. *Personnel Psychology, 63*, 595–640.

LaPiana, W. P. (2004). Merit and diversity: The origins of the law school admissions test. *Saint Louis University Law Journal, 48*(3), 955–990.

Law School Admissions Council. (2021). *About the LSAT.* Retrieved from http://www.lsac.org/jd/lsat/about-the-lsat

Law School Admission Council. (2019). *2017–2019 LSAT correlation study results.* https://www.lsac.org/system/files/inline-files/lsat-correlation-results_combined_2017_2018_2019_0.pdf

Le, H., Oh, I.-S., Robbins, S. B., Ilies, R., Holland, E., & Westrick, P. (2011). Too much of a good thing: Curvilinear relationships between personality traits and job performance. *Journal of Applied Psychology, 96*, 113–133.

Lempert, R. O., Chambers, D. L., & Adams, T. K. (2000). Michigan's minority graduates in practice: The river runs through law school. *Law & Social Inquiry, 25*(2), 395–505.

Levashina, J., Hartwell, C. J., Morgeson, F. P., & Campion, M. A. (2014). The structured employment interview: Narrative and quantitative review of the research literature. *Personnel Psychology, 67*(1), 241–293.

Libbrecht, N., Lievens, F., Carette, B., & Côté, S. (2014). Emotional intelligence predicts success in medical school. *Emotion, 14*(1), 64–73.

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

Lievens, F., De Corte, W., & Brysse, K. (2003). Applicant perceptions of selection procedures: The role of selection information, belief in tests, and comparative anxiety. *International Journal of Selection and Assessment, 11*(1), 67–77.

Lievens, F., & De Paepe, A. (2004). An empirical investigation of interviewer-related factors that discourage the use of high structure interviews. *Journal of Organizational Behavior, 25*(1), 29–46.

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. (2006). Video-based versus written situational judgment tests: A comparison in terms of predictive validity. *Journal of Applied Psychology, 91*(5), 1181–1188.

Lievens, F., & Sackett, P. R. (2017). The effects of predictor method factors on selection outcomes: A modular approach to personnel selection procedures. *Journal of Applied Psychology, 102*(1), 43–66.

Lievens, F., Sackett, P. R., & Zhang, C. (2021). Personnel selection: A longstanding story of impact at the individual, firm, and societal level. *European Journal of Work and Organizational Psychology, 30*, 444–455.

Lubinski, D. (2000). Scientific and social significance of assessing individual differences: "sinking shafts at a few critical points." *Annual Review of Psychology, 51*(1), 405–444.

McLarnon, M. J. W., Rothstein, M. G., Goffin, R. D., Rieder, M. J., Poole, A., Krajewski, H. T., & Mestdagh, T. (2017). How important is personality in the selection of medical school students? *Personality and Individual Differences, 104*, 442–447.

McManus, I. C., Dewberry, C., Nicholson, S., Dowell, J. S., Woolf, K., & Potts, H. W. W. (2013). Construct-level predictive validity of educational attainment and intellectual aptitude tests in medical student selection: Meta-regression of six UK longitudinal studies. *BMC Medicine, 11*(1), 1–21.

McManus, I. C., Smithers, E., Partridge, P., Keeling, A., & Fleming, P. R. (2003). A levels and intelligence as predictors of medical careers in UK doctors: 20 year prospective study. *BMJ, 327*, 139–142.

Moscoso, S. (2000). Selection interview: A review of validity evidence, adverse impact and applicant reactions. *International Journal of Selection and Assessment, 8*(4), 237–247.

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.

Ones, D.S. and Viswesvaran, C. (1998) Integrity testing in organizations. In R.W. Griffin, A. O'Leary-Kelly and J.M. Collins (eds) *Dysfunctional Behavior in Organizations: Violent and Deviant Behavior*. JAI Press.

Ones, D. S., Dilchert, S., & Viswesvaran, C. (2012). Cognitive Abilities. In N. Schmitt (Ed.), *The Oxford Handbook of Personnel Assessment and Selection* (pp. 1–92). Oxford University Press.

Patterson, F., Knight, A., Dowell, J., Nicholson, S., Cousans, F., & Cleland, J. (2016). How effective are selection methods in medical education? *A Systematic Review. Medical Education, 50*(1), 36–60.

Poole, P. J., Moriarty, H. J., Wearn, A. M., Wilkinson, T. J., & Weller, J. M. (2009). Medical student selection in New Zealand: looking to the future. *The New Zealand Medical Journal, 122*(1306), 88–100.

Posthuma, R. A., Morgeson, F. P., & Campion, M. A. (2002). Beyond employment interview validity: A comprehensive narrative review of recent research and trends over time. *Personnel Psychology, 55*, 1–81.

Puddey, I. B., & Mercer, A. (2014). Predicting academic outcomes in an Australian graduate entry medical programme. *BMC Medical Education, 14*(31), 1–12.

Rahim, A. F. A., & Yusoff, M. S. B. (2016). Validity evidence of a Multiple Mini Interview. *Education in Medicine Journal, 8*(2), 49–63.

Randall, R., Davies, H., Patterson, F., & Farrell, K. (2006). Selecting doctors for postgraduate training in paediatrics using a competency based assessment centre. *Archives of Disease in Childhood, 91*(5), 444–448.

Randall, R., Stewart, P., Farrell, K., & Patterson, F. (2006). Using an assessment centre to select doctors for postgraduate training in obstetrics and gynaecology. *The Obstetrician & Gynaecologist, 8*(4), 257–262.

Risavy, S. D., & Hausdorf, P. A. (2011). Personality testing in personnel selection: Adverse impact and differential hiring rates. *International Journal of Selection and Assessment, 19*(1), 18–30.

Roberts, C., Walton, M., Rothnie, I., Crossley, J., Lyon, P., Kumar, K., & Tiller, D. (2008). Factors affecting the utility of the multiple mini-interview in selecting candidates for graduate-entry medical school. *Medical Education, 42*(4), 396–404.

Roth, P. L., BeVier, C. A., Switzer, F. S., III., & Schippmann, J. S. (1996). Meta-analyzing the relationship between grades and job performance. *Journal of Applied Psychology, 81*(5), 548–556.

Roth, P. L., & Bobko, P. (2000). College grade point average as a personnel selection device: Ethnic group differences and potential adverse impact. *Journal of Applied Psychology, 85*(3), 399–406.

Sackett, P. R., & Lievens, F. (2008). Personnel selection. *Annual Review of Psychology, 59*, 419–450.

Sartania, N., McClure, J. D., Sweeting, H., & Browitt, A. (2014). Predictive power of UKCAT and other pre-admission measures for performance in a medical school in Glasgow: A cohort study. *BMC Medical Education, 14*(116), 1–10.

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), 262–274.

Shultz, M. M., & Zedeck, S. (2011). Predicting lawyer effectiveness: Broadening the basis for law school admission decisions. *Law & Social Inquiry: Journal of the American Bar Foundation, 36*(3), 620–661.

Shultz, M. M., & Zedeck, S. (2012). Admission to law school: New measures. *Educational Psychologist, 47*(1), 51–65.

Stedman, J. M., Hatch, J. P., & Schoenfeld, L. S. (2009). Letters of recommendation for the predoctoral internship in medical schools and other settings: Do they enhance decision making in the selection process? *Journal of Clinical Psychology in Medical Settings, 16*(4), 339–345.

Stevens, L., Kelly, M. E., Hennessy, M., Last, J., Dunne, F., & O'Flynn, S. (2014). Medical students views on selection tools for Medical School-a mixed methods study. *Irish Medical Journal, 107*(8), 229–231.

UCAT (2021). *UCAT test format.* Retrieved from: https://www.ucat.ac.uk/ucat/test-format/

Van Iddekinge, C. H., Roth, P. L., Raymark, P. H., & Odle-Dusseau, H. N. (2012). The criterion-related validity of integrity tests: An updated meta-analysis. *Journal of Applied Psychology, 97*(3), 499–530.

Wanek, J. E., Sackett, P. R., & Ones, D. S. (2003). Towards an understanding of integrity test similarities and differences: An item-level analysis of seven tests. *Personnel Psychology, 56*, 873–894.

Webster, E. S., Paton, L. W., Crampton, P. E. S., & Tiffin, P. A. (2020). Situational judgement test validity for selection: A systematic review and meta-analysis. *Medical Education, 54*(10), 888–902.

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.

Woods, S. A., Ahmed, S., Nikolaou, I., Costa, A. C., & Anderson, N. R. (2020). Personnel selection in the digital age: A review of validity and applicant reactions, and future research challenges. *European Journal of Work and Organizational Psychology, 29*, 64–77.

Yates, J., & James, D. (2010). The value of the UK Clinical Aptitude Test in predicting pre-clinical performance: A prospective cohort study at Nottingham Medical School. *BMC Medical Education, 10*(55), 1–9.

Part III
Teacher Selection: Past, Present, and Future

Chapter 6
Teacher Selection: History and Current Practices

Abstract In this chapter, we show that choosing the right people to teach has been an ongoing challenge in education, with researchers, practitioners, and policymakers pondering two key questions for as long as teachers have been appointed: What are the key personal characteristics related to teacher effectiveness? and How can these characteristics be assessed in a valid way when selecting prospective teachers? These two questions have received long-standing attention in education, but with little systematic research carried out to provide guidance to selectors. This chapter begins with an overview of historical issues in teacher selection, and then examines current practices for selecting teachers into ITE and into employment in a range of jurisdictions. The second half of the chapter reports a review of the research exploring the links between teacher selection practices and teacher effectiveness based on a recent meta-analysis that closely examined the research (Klassen & Kim, 2019).

In Chap. 5 we examined the selection practices in a range of fields outside of teaching and teacher education, including selection into medical education, legal education, and various kinds of jobs in organizational contexts. In this chapter we turn our attention towards the selection of teachers, first taking a historical perspective, and then critically reviewing current practices.

At the heart of teacher selection is the prediction of short-term and long-term teaching effectiveness. The question for selecting teachers for initial teacher education (ITE) or into employment is, at its essence, the same: *Will this person be, or develop into, an effective teacher*? Teachers become more effective, more reflective, and more knowledgeable about teaching as they gain experience (Antoniou et al., 2015; Atteberry et al., 2015), and the 'art of selection is to sample and evaluate personal attributes and behaviors that are believed to predict future effectiveness in the classroom. However, predicting teacher effectiveness is remarkably difficult, because teaching is a complex, multi-faceted job that is influenced by a host of interacting environmental and personal factors (e.g., Rimm-Kaufman & Hamre, 2010). Furthermore, predicting future behaviors from a brief sample of carefully curated behaviors during the selection process will always carry a measure of error. Choosing selection methods that are reliable, valid, and fair can improve the likelihood that we will make the best possible decisions about prospective teachers.

R. M. Klassen and L. E. Kim, *Teacher Selection: Evidence-Based Practices*,
https://doi.org/10.1007/978-3-030-76188-2_6

In this chapter, we show that choosing the right people to teach has been an ongoing challenge in education, with researchers, practitioners, and policymakers pondering two key questions for as long as teachers have been appointed: *What are the key personal characteristics related to teacher effectiveness?* and *How can these characteristics be assessed in a valid way when selecting prospective teachers?* These two questions have received long-standing attention in education, but with little systematic research carried out to provide guidance to selectors. This chapter begins with an overview of historical issues in teacher selection, and then examines current practices for selecting teachers into ITE and into employment in a range of jurisdictions. The second half of the chapter reports a review of the research exploring the links between teacher selection practices and teacher effectiveness based on a recent meta-analysis that closely examined the research (Klassen & Kim, 2019).

6.1 Historical Perspective on Teacher Selection

The question of how to select the most effective teachers has been asked for nearly a century. In 1922 F. B. Knight asked the questions that remain at the heart of teacher selection:

> *What facts concerning a candidate for a teaching position are of prognostic value? Of a hundred graduates of a normal college quite probably some will make excellent teachers, a larger number will do well, and a few will fail. By what system of interviewing can a superintendent increase his chances of picking more successful teachers and fewer failures than pure chance would account for? What qualities possessed by a candidate and ascertainable by a prospective employer are correlated highly enough with teaching success to be worth considering in a sound selective technique? (*Knight, 1922*, p. 207)*

Knight's work was an attempt to improve the likelihood of making good decisions about selection through establishing "statistically dependable facts to teacher selection" (p. 207). His study assessed a wide range of potential predictors: handwriting, age, experience, intelligence, ranking in teacher education program, amount of additional 'professional study' (defined as summer school and Saturday work in educational courses), and a 'trade test' assessing candidates' knowledge about teaching. No significant relationship with teacher effectiveness was found for handwriting, age, teaching experience, intelligence, amount of professional development, or standing in a teacher education program. The study found a statistically significant relation between teacher effectiveness and a tailor-made 'trade test', which was designed to measure knowledge about teaching practices. Knight concluded by questioning whether school district superintendents were able to reliably identify teacher effectiveness using their own intuition and called for "a genuinely scientific procedure of teacher selection" (p. 216).

Other early twentieth century educationalists recognized the importance of selection, with Tubbs (1928) leading the charge in promoting more systematic research on teacher selection: "Upon this one thing (i.e., teacher selection) more than any

other depends on success or failure in the education and training of future citizens" (p. 332). Tubbs' strong beliefs about the importance of teacher selection stemmed from teacher shortages of the time, partly due to the loss of teachers and potential teachers in WWI, and also from high rates of teacher attrition rates. Tubbs stated that attrition rates should be "regarded with alarm by everyone interested in American education" (p. 323), due to the one-third to one-half of the teaching population that were leaving the profession (or changing jobs within the profession) each year. He proposed five criteria for teacher selection: (a) educational background, (b) *experience*, which, according to Tubbs, builds adaptability, (c) *health* ("teachers are under a moral obligation to protect others from any possible contagion or infection" p. 328), a focus perhaps not surprising after the ravages of the influenza epidemic of the preceding decade), (d) *character* (described as the 'greatest' of the requisite qualities), and (e) *personality*, the lack of which "greatly handicaps the quality of service which a superior teacher should give" (p. 329).

Tubbs described the ability to "see below surface indications" in the selection of teachers as a gift without which "no (superintendent) can meet with more than a modicum of success" (p. 329). The 'problem' of teacher selection has historical roots—and current employers and ITE providers continue to focus on identifying the attributes associated with future effectiveness, and how to measure these attributes in a way that is reliable, valid, and fair.

6.2 Need for Teacher Selection

A process for teacher selection is needed when the number of applicants is greater than the number of available ITE places or jobs, when there is a need to identify unsuitable applicants ('selecting out') before beginning training or employment, and when there is a benefit in generating profiles of applicant strengths and weaknesses for future development. Systems for teacher selection are built on data gathered from existing records (e.g., evaluation of academic transcripts) and from new sources (e.g., face-to-face interviews, personality tests, teaching demonstrations) that are determined by employers or teacher education programs. Although selection methods have been the subject of in-depth research attention in some professions—especially medicine and business—the knowledge gained has nor often spilled over into education.

A selection process—for training or employment—is a predictive exercise that involves three steps: first, identification of the attributes needed for success in the endeavor, second, development of a method for assessing these attributes, and third, an assessment of the relationship between measured attributes and some kind of criterion or outcome measure. In order to make these predictions, selectors gather evidence that they believe can help them make valid selection decisions. Most selection methods will focus on the three personal characteristics discussed in Chap. 2: cognitive attributes, such as subject area knowledge; non-cognitive attributes, such as beliefs, motives, traits, and dispositions; and background experience, including previous relevant experiences. Higher-performing school systems tend to have more

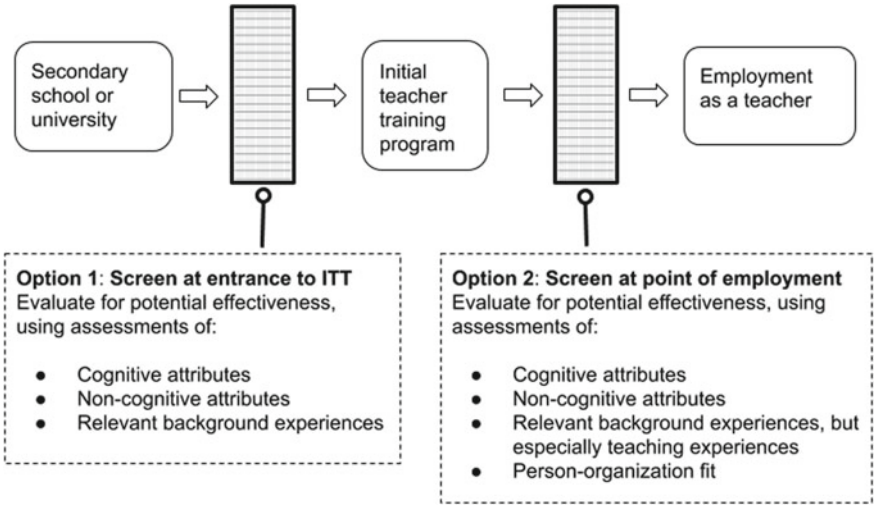


Fig. 6.1 Options for screening applicants

sophisticated selection systems (i.e., with multiple stages assessing multiple factors), with explicit recognition that poor selection methods influence the quality of the teaching workforce (Barber & Mourshed, 2007). Whatever selection process is implemented, the purpose of teacher selection is universal: using the best possible data to make the best possible decisions about prospective teachers.

Prospective teachers can be selected, or screened, at two key points: at entry into ITE or entry into employment. Figure 6.1 shows two potential screening points in the teacher selection process. Option 1 shows screening at the entrance to ITE, where cognitive attributes, non-cognitive attributes, and relevant background factors are evaluated. Option 2 shows that screening at the point of employment includes evaluation of the same factors, but with the addition of consideration of 'person-organization fit', where consideration is given to how well the applicant might fit into the school or school district based on additional, possibly non-evaluated factors. Higher performing education systems tend to have more effective processes to select candidates for ITE. Barber and Mourshed (2007) show that countries that perform well in international comparisons, such as Finland and Singapore, have selection procedures that are systematic, test a wide range of attributes, and filter applicants at the point of entry into ITE, rather than at the point of entry into the profession.

6.3 Selecting Candidates for ITE

Little research evidence is available supporting the predictive validity of selection methods in teacher education (e.g., Casey & Childs, 2011), and some of the methods

used, such as letters of reference and interviews have been shown to be biased against certain groups of candidates (Patterson et al., 2016). Not very much is known about the effectiveness of selection procedures into ITE, and what we do know suggests an arbitrariness in selection methods (e.g., Casey & Childs, 2017; Denner et al., 2001).

Selection into ITE also varies according to the structure of the ITE program. Two models of ITE are generally endorsed: a *consecutive* model in which students first complete an undergraduate degree in a particular subject, and then enroll in an ITE program. A *concurrent* model involves students studying a combination of a particular subject (or subjects) alongside courses involving pedagogical and theoretical instruction coupled with practical experience in teaching. In many countries, universities (and other initial teacher-education providers) offer different selection methods for undergraduate and postgraduate entrance for ITE programs. For example, in Finland, the two-phase selection process for direct entry 'class-teacher' education programs involves a nationwide literacy test (VAKAVA) which assesses the cognitive attributes of memorization, understanding, and the ability to apply knowledge from articles to practice (Malinen et al., 2012). The second phase of selection involves an 'aptitude test' developed by individual universities, and aimed at evaluating applicants' suitability, motivation, and commitment to teaching. The aptitude test varies across universities but may include an individual interview and a group discussion task (Malinen et al., 2012). Selection methods in Finland have recently been under review, with a consideration of alternative selection methods, including situational judgment tests (SJTs) and multiple mini-interviews (MMIs), currently underway (personal correspondence, R. Metsäpelto, May 2020).

In the UK, selection for ITE programs usually takes place at the postgraduate level. A survey of 74 university-based initial teacher education providers in England and Wales was conducted to understand how cognitive and non-cognitive attributes were assessed for selection (Klassen & Dolan, 2015). Cognitive attributes were assessed in multiple ways: through the use of a government mandated professional skills (literacy and numeracy) test, through evaluation of academic qualifications such as A levels, GCSE grades in English and Math, and through evaluation of university degree performance or 'class' (i.e., 1st, 2:1, 2:2, etc.). Non-cognitive attributes were assessed through individual and group interviews (97%), assessment of social behaviours through group activities (62%), and formal personality tests (3%); however, the survey revealed no evidence of the robustness of assessment practices, and published research on the topic is rare (Klassen & Kim, 2019).

The range of selection methods used for ITE programs varies across and within countries. In Table 6.1 we report how a sample of international ITE programs assess cognitive, non-cognitive, and background factors for selection into their programs. Most of the ITE programs included in the sample evaluate cognitive factors through an assessment of achievement level of the completed degree (e.g., minimum qualification standards such as degree class (in the UK) or GPA (in American settings). In some jurisdictions, cognitive attributes are further assessed at selection; for example, in Singapore, entrance proficiency tests are used to test subject knowledge in some subjects, and in the U.S., scores from a basic skills test in math, reading, and writing

Table 6.1 Selection criteria for initial teacher education programs (content accurate as of June 2021)

ITE program	Assessment of cognitive factors	Assessment of non-cognitive factors	Assessment of background factors/experience
Australia (University of New South Wales) • 1-year Master of Teaching (Primary) Postgraduate	• Completed UG degree with a major in one of six key learning areas	• Personal statement (150 words) outlining "reasons for wanting to be a teacher and why you are suited for that role" (scored on a 2-point scale)	Not listed
Canada (University of British Columbia) • 11-month BEd program (Elementary) Postgraduate	• Completed UG degree • 65% GPA on last 60 credits of post-secondary coursework, including English, laboratory science, mathematics, Canadian history or geography, Canadian studies	• Personal profile essay (maximum 900 words) responding to prompts: "Why do you want to become a teacher? Describe the kind of teacher you want to be." • Two reference letters "speaking to the applicant's experiences, interests, and abilities relevant of the teaching profession"	• A minimum of 100 hours of practical experience working with groups of 10 or more children or youth • Two reference letters "speaking to the applicant's experiences, interests, and abilities relevant to the teaching profession"
Finland (University of Helsinki) • 1-year STEP program) Postgraduate	• Completed Master's degree in teaching subject and Pedagogical Studies for Teachers	• Individual interview to evaluate suitability to work as a teacher (educability, motivation, and commitment). A 'multiple mini-interview is now being used in some Finnish contexts'	Not listed

(continued)

Table 6.1 (continued)

ITE program	Assessment of cognitive factors	Assessment of non-cognitive factors	Assessment of background factors/experience
Korea (Seoul National University) • 4-year Bachelor of Primary Education • 2.5-year Master of Primary Education	• Excellent high school grades (mark not specified but based on competitiveness) • Completed UG degree • Entrance proficiency test	• Individual interview with panel • (Letter of recommendation from school and letter of recommendation from a teacher) • Individual interview with panel	Not listed
Malawi (national selection for primary teacher education) Post-secondary	• Secondary school academic record • 'Aptitude test' at selection including tests of reasoning, mathematics, English	• Situational judgment test at selection assessing non-cognitive attributes (empathy, communication, organization, resilience, adaptability, integrity, community relationships, motivation, commitment, reflection, creativity, autonomy	Not listed
Singapore (National Institute of Education) • 1-year diploma in education Postgraduate	• Completed UG degree • Entrance proficiency tests (for some subjects)	• Interview (writing exercise; role play; individual interview with panel) • Evaluation of suitability during compulsory 4-month 'untrained teaching stint'	• Compulsory 4-month teaching stint (if selected for appointment) that must be passed before students continue in the program
United Kingdom (University of Newcastle) • 1-year certificate in education Postgraduate	• Completed UG degree at 2.2 level or above • Written and mathematics task on selection day	• Selection day activities include an individual interview assessing motivation, self-awareness, educational issues, classroom management	• 'Some experience of the English school system, usually through school observation visits'

(continued)

Table 6.1 (continued)

ITE program	Assessment of cognitive factors	Assessment of non-cognitive factors	Assessment of background factors/experience
United Kingdom (Teach First)	• A 2.1 degree or above • Grade C in GCSE Maths and English	• Two stage process. First, online application that assesses seven competencies (e.g., motivation, leadership, resilience). An SJT is used to provide 'a preview of life as a Teach First trainee'. Second, a 'development center' that includes a 1-1 competency-based interview, a group exercise with self-evaluation, and a teaching demonstration with self-evaluation	Not listed
USA (University of Washington) • 1-year Master's in Teaching program (Elementary) Postgraduate	• Completed UG degree with 3.0 GPA • WEST-B Basic Skills Test (Math, Reading, Writing) • ACT or SAT scores • NES (state-mandated content knowledge test) • 3.0 GPA from 60 credits	• Goal statement highlighting passion for teaching underserved populations and an interest in social justice + completion of 'character and fitness form' • Two letters of recommendation	• Minimum of 40 hours of documented education-relevant experience • Resumé detailing work and academic experience + prerequisite courses on education of an ethnic group, developmental psychology, and math for elementary teachers

Note Sources are publicly available ITE program websites except for Malawi, which is from Klassen et al. (2018)

provide further data beyond the information provided in undergraduate degree transcripts. The theme running through all selection processes is that decisions are made based on the evaluation of cognitive and non-cognitive attributes and sometimes including an evaluation of relevant background factors.

6.4 Selecting Teachers into Employment

The methods used to select prospective teachers into employment have been described as "ad hoc" (p. 24, Goldhaber et al., 2014) and "information poor" (p. 324, Liu & Johnson, 2006), with weak empirical and theoretical foundations supporting their use. The lack of research analyzing the effectiveness of selection methods is surprising in light of the importance of teachers in achieving societal goals of social equality and improving knowledge levels, and in light of the knowledge about selection we have accrued in other fields.

The kinds of methods chosen for selection depends on the volume of applicants and the degree of centralization of the hiring process. In more centralized systems with a large volume of applicants (e.g., Austria, Italy, Korea, Malaysia, Malta, Mexico, Singapore, Spain, and Turkey), selection methods tend to be standardized, with the specific methods set by central bodies. In less centralized systems (e.g., Belgium, Bulgaria, Hungary, Norway, Poland), schools, and especially school principals, have considerable autonomy in hiring. Other countries used a combined system where a central office may screen applicants at an initial stage, but individual schools make final hiring decisions (e.g., Australia, Canada, United States). Whether the system is centralized or less centralized, the methods chosen for selection typically target a combination of cognitive attributes, non-cognitive attributes, and background factors (e.g., teaching experience).

Based on data from PISA 2012, Han (2018) found that decentralized selection processes (i.e., school-based hiring) are associated with greater variance in the distribution of teacher quality across schools, and a greater gap in achievement between low- and high-SES students. Although Han's data did not speak to specific selection methods used, research from organizational psychology suggests that the methods used by smaller employing units (e.g., schools that might be hiring one or two teachers) tend to be more idiosyncratic and less reliable than more systematic and structured selection methods used by larger organizations.

In decentralized systems, individual school principals play a key role in deciding the elements assessed during the selection process. Engel and Finch (2015) interviewed 31 principals of urban schools in Chicago about their decision-making processes when making new teacher hiring decisions. Evaluation of cognitive attributes, especially subject area knowledge, was typically done through collaboration with school colleagues, but the attributes targeted during selection tended to be determined non-systematically, e.g., "We sort of sit down and talk a little bit about what we are looking for… You know, to decide, what kind of person do we want to have here?" (p. 32). Systematic differences were found in the strategies principals

used to recruit and hire new teachers: principals in lower achieving schools were more likely to hire substitute or student teachers than principals in higher achieving schools (who accessed larger social networks in hiring), and principals in primary schools tended to work more autonomously throughout the hiring process than principals in secondary schools.

Two case studies: hiring teachers in the U.S. (NYC) and Australia (NSW). The methods used for selecting teachers for employment are similar across contexts. The New York City Department of Education Hiring Guide (2018–2019) outlines the use of individual interviews to target cognitive attributes such as *content knowledge* (How would you make your content area relevant to daily life?) and *instructional practice* (What specific strategies do you use for classroom management?). The assessment of non-cognitive attributes includes individual interviews targeting *beliefs and strengths* (Why did you become a teacher? What are three words to describe yourself as a teacher?), *collaboration* (How do you feel about collaborative teaching?) and *student understanding* (Does a student's background influence his or her achievement?). Assessment of background factors, and especially teaching skills, is optionally assessed through a demonstration lesson where students are evaluated on their "poise and comfort in front of a group" and on how well applicants test for student understanding.

The New South Wales Department of Education is the biggest employer of teachers in Australia, with over 2,200 schools ranging from very remote to very urban settings. Current government policy requires teacher education providers to select teachers based on both cognitive and non-cognitive attributes to ensure suitability for teaching (Sheridan et al., 2021). The 'New standards for NSW's teachers' document (NSW government, 2018) highlights five main criteria required for graduates to be considered for teaching positions:

- A minimum credit grade point average
- Sound practical knowledge and ability
- Superior cognitive and emotional intelligence measured by psychometric assessment
- Commitment to the values of public education displayed in an interview
- Preference for face-to-face teaching degrees over online degrees

Both cognitive and non-cognitive attributes are assessed through Teacher Suitability Assessments (https://www.teach.nsw.edu.au/becomeateacher/approval-to-teach/faqs) that include measures of verbal reasoning (ACER Advanced Test), abstract reasoning (ACER APTS Abstract Reasoning Organisational), and emotional intelligence (Genos Emotional Intelligence Inventory). During the COVID-19 pandemic, online interviews were used to assess **knowledge** (including pedagogy and syllabus content), **critical experiences** (demonstration of actions that have contributed to student progress and wellbeing), and **skills and capabilities**. Targeted cognitive attributes are aligned with the Australian Professional Standard for Teachers, e.g., *Know students and how they learn; Know the content and how to teach it; Plan for and implement effective teaching and learning.* An assessment of

professional experience (and/or practicum reports) is used to identify readiness for success in the classroom.

6.5 How Valid Are Current Selection Methods?

In education, there has been little systematic research examining the efficacy of selection methods (Bowles et al., 2014; Liu & Johnson, 2006). Many of the existing selection methods are based on ad hoc decisions with little evidence supporting their use. A recent study conducted in the UK (Davies et al., 2016) explored how selection methods were developed for teacher education programs, with the finding that selectors emphasized their intuition when making selection decisions: "Really, you do only need one question and a bit of conversation to find out what you are looking for" (p. 298), with selectors tending to rely on a "gut feeling" to identify the "X factor" (p. 298). No evidence was gathered to support the selection methods used: "I wouldn't have any statistics… after they've left us," (p. 297).

Most people are confident that they can accurately judge personality and other personal characteristics through interviews (Dana et al., 2013), but research tells us otherwise. Research from organizational psychology suggests that interviewers, especially when conducting unstructured interviews, suffer from unreliable judgment and are influenced in the decisions they make by unconscious biases based on race, age, and appearance (Cook, 2009). In education, selection methods may pay lip service to well-developed teaching standards frameworks that reflect multiple competencies and values (e.g., Casey & Childs, 2017; Denner et al., 2001), but the methods chosen for selection may not reliably assess these competencies. In any field, selection methods require regular evaluation of their reliability (consistency over time, validity (evidence of predictive utility), and fairness for all applicants, regardless of age, gender, ethnicity, sexuality, and socio-economic status.

6.6 Reviews of Research on Teacher Selection Practices

Research on the efficacy of teacher selection methods is under-developed in comparison to research in other professional fields, such as medicine or business. In Chap. 5, we reviewed selection practices in other fields, and considered the evidence available for the selection methods used in these fields. We saw that systematic research on selection into employment and training is particularly well developed in medical education, where a systematic program of research has been conducted into the reliability, validity, and fairness of selection methods. However, much less research attention has been given to the methods used for teacher selection. Two review studies sum up the field: Metzger and Wu (2008) and, more recently, Klassen and Kim (2019).

Metzger and Wu's 2008 meta-analysis. In 2008, Metzger and Wu reviewed and meta-analyzed 24 studies that examined the predictive validity of one teacher

selection tool, the Gallup Teacher Perceiver Interview (TPI). The review proves a useful starting point in investigating selection practices in education because it examined the use of what was, in the 1980s and 1990s, one of the most widely used teacher selection tools in the United States. Metzger and Wu's meta-analysis used validity data from studies published from 1975 to 2003, largely from dissertations ($n = 16$), but also from reports from the Gallup Organization ($n = 7$) and from one journal article. Most of the studies included in their review (20/24) were published before 2000, with 6 studies from the 1970s, 10 studies from the 1980s, and 4 studies from the 1990s. Of the four post-2000 studies, one was a dissertation, (Buresh, 2003), one was a journal article (Young & Delli, 2002), and two were released by test companies (i.e., the Gallup organization). Overall, the authors found a range of −0.12 to 0.87 for the correlation between TPI scores and indicators of teaching effectiveness, with a weighted mean of $r = 0.28$, and a median r of 0.22, considered by the authors to be a moderate effect size. Although Metzger & Wu's meta-analysis provided a valuable snapshot of one selection tool at a particular point in time, more work is needed to provide a fuller, more accurate, and more up-to-date picture of the teacher selection landscape.

Klassen and Kim's 2019 meta-analysis. Klassen and Kim extended Metzger & Wu's, 2008 review in 2019 (2019) in order to broaden the coverage of all teacher selection methods in use, and to provide a more up-to-date look at selection (most of the studies Metzger and Wu included were published before 2000). The goal stated by Klassen and Kim was to examine the methods used for the selection of teachers for employment and prospective teachers entering ITE. Four key questions were posed in their review:

1. What is the predictive validity of the methods used to select teachers and teacher candidates?
2. Are there differences in the predictive validity of tests assessing cognitive and non-cognitive attributes?
3. Are there differences in the predictive validity of the methods used for selection into employment and for selection into ITE programs?
4. What is the relationship between cost and benefit (predictive validity of selection methods?

Method. The key indicator of effect size for the meta-analysis was Pearson's r, which is a measure of the size of relation between selection method and teacher effectiveness, and which can be interpreted as an indication of predictive validity. In educational research $r = 0.10$ describes a small effect, $r = 0.20$ describes a medium effect, and $r = 0.30$ describes a large effect. Coe (2002) proposed that an effect size of $d = 0.10$ (roughly $r = 0.05$) can result in important educational outcomes if the effect can be applied to all students (i.e., as in an effect involving teachers) and is cumulative over time. For context, predictive validity coefficients in other fields are as follows: 0.18 to 0.43 in dentistry (Patterson et al., 2012), 0.37 in medicine (Lievens & Patterson, 2011), between 0.06 to 0.50 in business (Christian et al., 2010), and 0.34 across multiple occupation groups (McDaniel et al., 2001).

The goal of the meta-analysis was to analyze studies that: (a) reported primary research in the form of journal articles, dissertations, and published research reports published between 2000 and 2017, (b) included participants who were job applicants or ITE candidates in the K-12 system, (c) included a selection measure (cognitive or non-cognitive) administered at the point of selection, and (d) included a measure of teacher effectiveness using an *external* source (i.e., not self-reported), either observation scores (from supervisor or principal) or classroom-level student achievement gains. The authors excluded (in contrast to Metzger and Wu) un-verified data provided by test companies in support of their commercial products. The search of relevant databases and key journals resulted in 1306 records which were then screened for relevance to the study. A series of further screens left a pool of 32 studies that met the criteria for inclusion, and which were included in further analyses.

Results. Table 6.2 presents a summary of the research questions, the results, and the implications for research and practice. An overall effect size of $r = 0.12$ was found for the relationship between selection method and teacher effectiveness. Out of the 32 studies, 28 showed positive effect sizes, and 4 showed negative effect sizes, but only 10 studies reported statistically significant findings, all positive. The moderator analyses, conducted in order to break down the relations between predictors and outcomes, showed that cognitive predictors ($r = 0.13$) were significantly more predictive of teacher effectiveness than non-cognitive predictors ($r = 0.10$). Methods to select candidates for ITE programs were nominally more predictive than selecting candidates into employment, but the difference was not significant. There was no indication that paying more for selection methods (in money and time) resulted in better outcomes.

Conclusions of the study. There are several key conclusions that can be drawn from the study. First of all, there has been much less research and development attention paid to selection methods in education than in other fields, with the result that the current methods are not as effective as in other professions. The existing methods are, in general, not very effective at predicting which candidates will be successful in ITE programs or as teachers in schools. Notwithstanding the fact that small validity coefficients can be usefully applied at the systems-level, there are several possible explanations for the lack of predictive validity of current selection methods in education. We know that in fields where selection methods are closely studied, there is growth and development in the methods that are used; for example, in medicine, the relatively extensive body of research has led to new selection methods—e.g., SJTs and MMIs—being developed, tested, and implemented. In education, most jurisdictions that were studied used commercial tools that have little published evidence of validity, or in-house methods that have been developed 'organically' but again, neither reflect best practices in current selection research, nor have a base of evidence supporting their use. Looking outside of teacher education and educational psychology to medical education and organizational psychology, where research on selection methods is extensive, is one way to refresh the current moribund state of teacher selection.

Table 6.2 Results from meta-analysis of research on teacher selection tools

Research question	Results	Implications
What is the overall predictive validity of the methods used to select teachers and teacher candidates?	Overall $r = 0.12$ (small effect)	Predictive validity of current teacher selection methods, on the whole, is modest, especially in comparison with other professions
Are there differences in the predictive validity of tests assessing cognitive and non-cognitive attributes?	Cognitive predictors: $r = 0.13$ Non-cognitive: $r = 0.10$ (Sig. difference)	Cognitive predictors are slightly better at predicting teacher effectiveness than non-cognitive predictors. Finding valid methods to evaluate prospective teachers' non-cognitive attributes remains a challenge
Are there differences in the predictive validity of the methods used for selection into employment and for selection into ITT programs?	Employment: $r = 0.11$ ITT programs: $r = 14$ (No sig. difference)	Although the validity of selection methods for ITT is nominally higher, the difference is not significant. The methods used to select teachers for employment and for training need further research and development
What is the relationship between cost and benefit (predictive validity) of selection methods?	Mean cost: US $104 per candidate Relationship between validity and cost: $r = -0.12$	Spending more money on teacher selection methods is no guarantee of success; the commercial methods currently available seem no better than 'in-house' methods
What are the overall conclusions?		Three main conclusions can be drawn: (a) there is a lack of research on teacher selection methods in comparison to other professions, and (b) the current methods are largely stagnant, and not very good, and (c) more attention should be paid to methods used in other fields

Note Data from Klassen and Kim (2019)

6.7 Chapter Summary

Questions about how to select the best possible teachers have been asked for at least a century, with researchers and practitioners in the early 1900s pondering the challenges of teacher selection. In this chapter we explored various models used to select prospective teachers into ITE and into employment and found that there was a general agreement that both cognitive and non-cognitive methods were necessary for

successful selection. Reviews of research on teacher selection methods found that there was less research in education than in other fields, and perhaps consequently, the methods used for selection were not very effective and were not reflective of leading-edge research and practice found in other professional and research fields. In the following chapter, we explore new approaches to teacher selection that have emerged in the last few years. In particular, we look at how situational judgment tests (SJTs) have been developed to identify prospective teachers who are most likely to experience success in ITE programs and in teaching jobs.

References

Antoniou, P., Kyriakides, L., & Creemers, B. P. M. (2015). The Dynamic Integrated Approach to teacher professional development: Rationale and main characteristics. *Teacher Development, 19*, 535–552.

Atteberry, A., Loeb, S., & Wyckoff, J. (2015). Do first impressions matter? Predicting early career teacher effectiveness. *AERA Open, 1*(4), 1–23.

Barber, M., & Mourshed, M. (2007). *How the world's best-performing schools systems come out on top*. McKinsey & Company.

Bowles, T., Hattie, J., Dinham, S., Scull, J., & Clinton, J. (2014). Proposing a comprehensive model for identifying teaching candidates. *The Australian Educational Researcher, 41*, 365–380.

Buresh, R. J. (2003). *The predictive validity of the Teacher Perceiver Interview in selecting effective elementary teachers in a mid-sized midwestern school district (Unpublished doctoral dissertation)*. University of North Dakota.

Casey, C., & Childs, R. (2011). Teacher education admission criteria as measure of preparedness for teaching. *Canadian Journal of Education, 34*, 3–20.

Casey, C., & Childs, R. A. (2017). 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.

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*, 83–117. https://doi.org/10.1111/j.1744-6570.2009.01163.x

Dana, J., Dawes, R., & Peterson, N. (2013). Belief in the unstructured interview: The persistence of an illusion. *Judgment and Decision Making, 8*, 512–520.

Davies, P., Connolly, M., Nelson, J., Hulme, M., Kirkman, J., & Greenway, C. (2016). 'Letting the right one in': Provider contexts for recruitment to initial teacher education in the United Kingdom. *Teaching and Teacher Education, 60*, 291–302.

Denner, P. R., Salzman, S. A., & Newsome, J. D. (2001). Selecting the qualified: A standards-based teacher education admission process. *Journal of Personnel Evaluation in Education, 15*, 165–180.

Engel, M., & Finch, M. A. (2015). Staffing the classroom: How urban principals find teachers and make hiring decisions. *Leadership and Policy in Schools, 14*, 12–41.

Goldhaber, D., Grout, C., and Huntington-Klein, N. (2014). *It's selective but is it effective? Exploring the predictive validity of teacher selection tools*. CEDR Policy Brief 2014–9. University of Washington, Seattle, WA.

Han, S. W. (2018). School-based teacher hiring and achievement inequality: A comparative perspective. *International Journal of Educational Development, 61*, 82–91.

Klassen, R. M., & Dolan, R. (2015, September). *Selection for teacher education in the UK and the Republic of Ireland: A proposal for innovation*. Presented at the meeting of the European Conference on Educational Research, Budapest, Hungary.

Klassen, R. M., & Kim, L. E. (2019). Selecting teachers and prospective teachers: A meta-analysis. *Educational Research Review, 26*, 32–51.

Knight, F. B. (1922). Qualities related to success in elementary school teaching. *The Journal of Educational Research, 5*, 207–216.

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*, 927–940.

Liu, E., & Johnson, S. M. (2006). New teachers' experiences of hiring: Late, rushed, and information poor. *Educational Administration Quarterly, 42*, 324–360.

Malinen, O.-P., Vaisanen, P., & Savolainen, H. (2012). Teacher education in Finland: A review of a national effort for preparing teachers for the future. *The Curriculum Journal, 23*, 567–584.

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*, 730–740.

Metzger, S. A., & Wu, M.-J. (2008). Commercial teacher selection instruments: The validity of selecting teachers through beliefs, attitudes, and values. *Review of Educational Research, 78*, 921–940. https://doi.org/10.3102/0034654308323035

NSW government (2018). *New standards for NSW's teachers.* Retrieved from: https://www.nsw.gov.au/news/new-standards-for-nsws-teachers

Patterson, F., Ashworth, V., Mehra, S., & Falcon, H. (2012). Could situational judgement tests be used for selection into dental foundation training? *British Dental Journal, 213*, 23–26.

Patterson, F., Knight, A., Dowell, J., Nicholson, S., Cousans, F., & Cleland, J. (2016). How effective are selection methods in medical education? A systematic review. *Medical Education, 50*, 36–60.

Rimm-Kaufman, S. E., & Hamre, B. K. (2010). The role of psychological and developmental science in efforts to improve teacher quality. *Teachers College Record, 112*, 2988–3023. http://www.tcrecord.org

Sheridan, L., Coleman, B., & Durksen, T. L. (2021). Important non-academic attributes in Australian initial teacher education. *The Australian Educational Researcher*. https://doi.org/10.1007/s13384-021-00443-x

Tubbs, E. V. (1928). The selection of teachers. *Peabody Journal of Education, 5–6*, 323–332.

Young, I. P., & Delli, D. A. (2002). The validity of the Teacher Perceiver Interview for predicting performance of classroom teachers. *Education Administration Quarterly, 38*, 586–612. https://doi.org/10.1177/0013161X02239640

Chapter 7
Situational Judgment Tests and Their Use for Teacher Selection

Abstract In this chapter we will sharpen our focus to look closely at situational judgment tests (SJTs), typically used for large-scale screening of applicants to training programs. Although SJTs have a solid research foundation and are commonly used for selection into training and employment in diverse professional fields, especially health-related fields, they have rarely been implemented in teacher education. In this chapter we look at the research and theory behind the use of SJTs for selecting the best possible teachers and provide some direction for how this method can be developed for use by ITE programs and other education organizations.

In Chap. 6, we took a historical view of teacher selection methods, and also examined methods that are currently implemented around the world. In this chapter we will sharpen our focus to look closely at situational judgment tests (SJTs), typically used for large-scale screening of applicants to training programs. Although SJTs have a solid research foundation and are commonly used for selection into training and employment in diverse professional fields, especially health-related fields, they have rarely been implemented in teacher education. In this chapter we look at the research and theory behind the use of SJTs for selecting the best possible teachers and provide some direction for how this method can be developed for use by ITE programs and other education organizations.

7.1 Situational Judgment Tests

SJTs have become increasingly popular in the last two decades because they show higher levels of predictive validity than other screening measures assessing non-cognitive attributes, are easy to use, and are well-received by applicants (e.g., Klassen et al., 2014). SJTs are a measurement method well suited for measuring judgment in challenging situations, and usually consist of a 'stem' and a series of response options.

R. M. Klassen and L. E. Kim, *Teacher Selection: Evidence-Based Practices*,
https://doi.org/10.1007/978-3-030-76188-2_7

The stem consists of a range of contextualized work-related scenarios presented in text or video format, and the response options provide a list of possible courses of action, usually preceded with the phrase, 'What *should* you do?' (procedural knowledge) or 'What *would* you do?' (behavioral intentions). Early versions of SJTs were created by military psychologists to select soldiers to join the armed forces in World War Two where the tests included a series of detailed and realistic scenarios that described challenging situations likely to be encountered in military settings. The early military SJTs were useful in several ways: they gave potential candidates a taste of what life as a soldier might hold, and they gave recruiters insight in how judgment was displayed by potential officers (Lievens & De Soete, 2015).

Theory underpinning SJTs. The theory underpinning SJTs—implicit trait policy—refers to an individual's implicit beliefs about the effectiveness of expressing particular personality traits in particular situations (Whetzel & McDaniel, 2009). SJTs do not explicitly measure personality or other non-cognitive attributes but are designed to capture useful information about personality traits *indirectly* by asking people to evaluate work-related scenarios and then to judge the effectiveness of response options. Some additional theoretical foundations for SJTs can be located in Sternberg's theory of successful intelligence, whereby procedural knowledge in complex situations is often tacit (e.g., Elliott et al., 2011), and in Boyatzis and Kelner's (2010) theory of links between attributes as the behavioral manifestation of implicit motives. However, the theory of *implicit trait policy* is most often used to describe the framework supporting the use of SJTs.

SJTs are considered a measurement method, and as such can be designed to capture a range of non-cognitive attributes, including Big Five personality traits such as agreeableness, extraversion, and conscientiousness (e.g., Hooper et al., 2004). SJTs can be designed to measure other related non-cognitive domains—including motivation, resilience, professional integrity, and empathy—derived from a careful job analysis of workplace demands (e.g., Patterson, Lievens, et al., 2013). A growing body of research shows that SJTs are a reliable and valid approach to make selection decisions in a range of professional contexts (Whetzel et al., 2020).

Context-specific or context-general SJTs? SJTs can be constructed to reflect a particular context, for example, scenarios representing a school environment for teacher selection SJTs, or to reflect more general situations that one might encounter in daily life. There is divided opinion on the importance of contextualization of SJTs. Lievens and Motowidlo (2016) argued that SJTs tap general domain knowledge that requires an understanding of the utility of expressing certain traits across a range of work settings. In such a model, SJTs are designed to measure specific constructs in a clear and explicit way, without reference to a particular situation. Other SJT theorists and researchers disagree. Harris et al. (2016) countered the 'situation-free' SJT approach by noting that general domain knowledge is always contingent on the use of contextual and situational cues. In trait activation theory (TAT), it is the interaction between person and situation that explains behaviour; a trait will only be expressed when a trait-relevant situation demands the activation of that particular trait in that particular situation (Harris et al., 2016).

In a similar fashion, Bandura's social cognitive theory proposes that personal characteristics are not formed and expressed in isolation; rather, it is the reciprocal interaction between personal characteristics, behaviour, and the environment that forms the basis of human agency (Bandura, 1999). Fan et al. (2016) argue that although general domain knowledge is an important feature of SJTs, it is the ability to understand when and how to express certain traits that separates SJTs from disembodied measures of personality or other interpersonal attributes. A series of studies recently published by Freudenstein et al. (2020) tested the importance of situation construal in SJTs. The authors found that test-takers' perceptions of the situation predicted responses even after controlling for personality, emotion recognition, and mental ability, and that situation construal plays a pivotal role in determining SJT responses. For SJTs used in teacher selection, providing job-relevant situational cues may be essential to understand how particular attributes are activated in authentic classroom environments.

Research on the use of SJTs for selection. The use of SJTs as an alternative to conventional selection tests for entry into professional training has received considerable recent research attention. The surge in interest is due to the effectiveness of SJTs for predicting job performance (e.g., Christian et al., 2010): SJTs have been shown to be better predictors of job performance than conventional personality tests (e.g., Shultz & Zedeck, 2012), and when tailored to specific contexts, are useful for selection purposes in a wide range of fields (Patterson et al., 2015). In addition, SJTs tend to display stronger face and content validity than conventional non-academic measures due to their close correspondence to the work-related situations that they describe (Whetzel & McDaniel, 2009). In addition, SJTs constructed by researchers working in collaboration with expert practitioners are less susceptible to coaching effects and faking than other kinds of selection tests (e.g., conventional personality tests).

Recent empirical studies and meta-analyses show that SJTs administered as selection tools at the beginning of training programs can be reliable and robust predictors of subsequent job performance (Patterson, Lievens, et al., 2013). SJTs have been used for selection into training programmes in a range of professions, including dentistry, law and medicine (see Chap. 5 for more detail). In medicine, SJTs have been successfully validated for use in selection into foundation year training in the UK (Patterson, Tavabie, et al., 2013) and are widely used across medical schools for selection purposes in the UK. In the United States, Shultz and Zedeck (2012) reported that SJTs were a better predictor of lawyer effectiveness than the conventional tests used for selection into highly competitive law schools, and furthermore, were less prone to inter-group differences (i.e., gender, SES, and ethnicity) than conventionally-used selection metrics (i.e., Law School Admissions Test and grade point average). SJTs show less inter-group bias than other selection methods such as tests of cognitive ability and interviews and are perceived to be fair by candidates (Patterson et al., 2015). In Table 7.1, we present a brief summary of research on SJTs in diverse professional fields, along with their validity evidence.

Table 7.1 SJTs Used for selection in various disciplines

Study	Context	Validity Evidence
Lievens and Sackett (2012)	Admission into **medical school**	SJTs (procedural knowledge about interpersonal behavior) predicted internship performance and job performance ($r = 0.21$) 9 years after SJT administration
Koczwara et al. (2012)	Admission into **advanced medical training**	SJTs were the best single predictor of performance in selection center
Patterson et al. (2012)	Admissions into **advanced dentistry training**	SJTs showed significant correlations ($r = 0.43$) with entrance interview (concurrent validity)
Shultz et al. (2012)	Developing new measures for **law school admissions**	SJTs showed significant correlations with 23 of 26 lawyering effectiveness factors
Bateson et al. (2014)	Selection of **service employees**	SJTs showed predictive validity for selection of service employees at the start of the recruitment process
Klassen et al. (2020)	Admissions into teacher education	SJTs used for selection showed predictive validity for performance in teaching placements (rs 0.24−0.30, $p <$ 0.01)

Reliability and predictive validity of SJTs. Most SJTs tend to have lower internal consistency than other tests measuring non-cognitive attributes, largely due to their multidimensional nature, with a review of SJT reliability showing a weighted corrected coefficient α of 0.46 (Catano et al., 2012). Kasten and Freund (2016) found that SJT internal consistency was higher for low stakes tests, for tests using theoretical, versus expert-based or empirical scoring, and for tests using Likert-type response scales, rather than 'pick-best' scoring approaches. The authentic situations on which SJTs are built tend to be complex, 'messy', and reflective of more than one non-cognitive attribute, even when an individual scenario is designed to represent a single attribute.

SJTs tend to be constructed to cover multiple domains, explaining their sometimes-lower internal consistency (but higher predictive validity) than single construct measures (e.g., cognitive ability or personality). However, reliability is dependent on test length and item heterogeneity. Internal consistency (i.e., alpha) may not be the best reliability index if item heterogeneity is high in the SJT (Catano et al., 2012), and some form of test–retest reliability or split-half estimates (e.g., Whetzel et al., 2020) will supplement reliability estimations. Internal consistency of teacher selection SJTs using a rating approach was shown to be acceptable (e.g., α

= 0.78 in Klassen et al., 2020; $\alpha = 0.79$ in Klassen et al., 2017; and $\alpha = 0.70$ in brief video format, Bardach et al., 2020).

Factor analysis of SJT content typically results in ambiguous factor structures unless cross-loadings are allowed. Factor analytic approaches need to explicitly model the multidimensionality of SJTs at the item level, not just for the test as a whole. The development strategy for SJTs represents a trade-off in assessment of non-cognitive attributes. An inductive or 'bottom-up' approach (e.g., using critical incidents) may result in a more predictive test, but one with lower internal consistency, whereas an SJT developed using a deductive, 'top-down' approach targeting specific constructs may result in a more internally consistent measure.

Meta-analytic research indicates that SJTs generally have good predictive validity (corrected $r = 0.34$; McDaniel et al. 2001). What has been difficult with SJTs in general has been establishing exactly which constructs are being measured and they are sometimes criticized as a 'black box' measurement method. A study investigated whether SJTs predicted job performance above and beyond cognitive ability, job experience, job knowledge, and conscientiousness in three samples (Clevenger et al., 2001), with SJTs predicting job performance in all three samples.

Scoring options for SJTs. Scoring for SJTs involves consideration of several options. Typically, SJTs are scored by comparing applicants' judgments with the judgments expressed by subject matter experts (SMEs). The judgment tasks in an SJT are designed to assess contextualized judgment and are based on the notion that situation-specific judgments and responses reflect implicit personality traits that have a causal effect on job performance. In contrast, conventional personality tests ask individuals to describe themselves directly, opening up the likelihood that candidates will choose responses that portray their personality in the best possible light, but possibly inaccurately. A second scoring option is empirical scoring, where scoring is determined either by consensus (i.e., 'crowd wisdom') or by examining applicant data and specifically the correlations of each response option to a criterion score (e.g., performance score on an important criterion). A third option, especially for SJTs that are more purely 'construct-driven' is to use a theoretically derived scoring key whereby the scoring pattern is determined by reference to, and interpretation of, the construct underlying the SJT content (e.g., Tiffin et al., 2020).

Traditional vs. construct driven SJTs. The traditional approach to developing SJT content is a 'bottom-up' or inductive approach, where SMEs (usually led by a psychologist or consultant) gather together to elicit 'critical incidents' (*Think of a time when a new trainee faced a challenging classroom situation requiring careful judgment. What did this trainee do? Was the response appropriate in your view? What other options might s/he have considered?*). Responses to these critical incident questions form the basis of SJT scenarios, with content reflecting authentic workplace situations and challenges. Following this path of test development typically leads to an SJT that is contextualized, accepted by candidates, predictive of the job it is built around, but hampered by conceptual and psychometric issues, and in particular, lack of a clear factor structure. There are clear benefits to developing SJTs using a traditional inductive approach, but some disadvantages, and in particular, problems with relating the content to specific domains and constructs.

In contrast, developing a construct-driven SJT depends on a 'top-down' approach, typically led by a psychologist or team of psychologists who focus more on a single target trait (e.g., integrity), and less on a particular context. The scenario is designed to elicit a particular trait, and the response options represent degrees of the target trait. Reliability coefficients tend to be higher for single-construct SJTs, and test scores tend to correlate more highly with personality measures. Predictive validity of construct driven SJTs is still emerging, although tests developed using this approach tend to show high correlations with other measures of the target trait. A multimedia SJT assessing emotion management showed similar validity to conventional measures of emotion management, although the test was not used to predict workplace performance. Construct-driven SJTs may be more prone to faking, similar to the case with conventional personality tests, where the appropriate course of action may be easier to detect when a series of scenarios all focus on the same construct with a detectable theme (Tiffin et al., 2020). Some promising work on construct driven SJTs relevant to teacher selection is currently being conducted by Bostwick and Durksen at the University of New South Wales in Australia in collaboration with the Teacher Selection Project (K. Bostwick, personal communication, October 2020), with development of a prototype SJT assessing growth mindset in teachers recently piloted.

Recent studies (e.g., Klassen et al., 2020) have used an integrated 'construct-informed' approach, in which target attributes are developed before the scenario development process, and critical incidents are mapped onto these attributes. A combined inductive-deductive approach allows for a priori non-cognitive attributes to be identified early on in the test development process (i.e., using a deductive approach), while allowing for 'bottom-up' or inductively derived attributes to emerge during the development process. Using an integrated construct-informed approach leads to building the scenarios and response options of the SJT on a foundation of identifiable domains, although factor analyses typically show the existence of single overarching factors representing judgment about effective behaviors in particular contexts (Patterson et al., 2015).

Applicant reactions to SJTs. Research on applicant reactions to selection processes has been based on evaluation of *procedural justice*, or the perceived fairness of the methods used for decision making, and *distributive justice,* or the fairness of the outcome of the selection process (Patterson et al., 2011). Measures of non-cognitive attributes such as interpersonal skills and empathy may not be viewed as positively as 'fact-based' assessments; indeed, Patterson et al. (2011) found that for selection into specialist medical training, a clinical skills test was more favourably received by candidates than an SJT assessing non-clinical judgment, even though the SJT was a better predictor of subsequent outcomes. Most studies show that SJTs are favourably received by applicants, and that video-based SJTs are preferred to text-based SJTs. Bardach et al. (2020) found that video-based SJTs were more engaging than text-based SJTs, but applicants did not rate the video format as fairer or more job-related than the text version. Klassen et al. (2014) evaluated applicant reactions to an SJT for admission into ITE using a mixed methods approach and found that applicants were generally favorable about the relevance of the SJT content and the

appropriateness of its use for selection, but raised questions about procedural justice issues related to (in)experience: "It's hard to judge (the scenarios) if you've never been in a certain situation before", and that the SJTs measured skills that "should be taught during teacher training" (p. 116). However other researchers have shown that SJTs that are contextualized and show higher fidelity to the job are preferred by candidates due to their relevance and the realistic nature of the scenarios (Whetzel & McDaniel, 2009).

7.2 Situational Judgment Tests for Teacher Selection

SJTs have rarely been implemented for selection into teacher education or for entry into the profession, but their use is increasing in the last decade. In a review of teacher selection methods (Klassen & Kim, 2019), the authors found that it was rare for teacher selection programs to be built on methods with a published evidence base (i.e., using SJTs, MMIs, or other evidence-based methods). Research on using SJTs for selecting teachers is slowly emerging, with most studies in the last decade coming from the Teacher Selection Project group in the UK (https://www.teacherselect.org/), and some from the TCAT group in Australia (e.g., Bowles et al., 2014). In 2014, the Teacher Selection Project group published one of the first articles on teacher selection using SJTs (Klassen et al., 2014), describing applicant reactions to SJTs used in parallel to other selection methods (see Fig. 7.1 for a sample teacher selection SJT). Results showed that SJTs administered to primary and secondary teaching applicants were generally well received, with primary applicants expressing more favorable opinions of the test than secondary applicants.

The research base investigating the use of SJTs to predict teaching performance has grown at pace in the last 10 years, including research on the predictive validities

Situational Judgment Test Sample Item

You are walking into school when the parent of one of your students, Mr Andrews, asks if he can speak to you. He informs you that his son, Callum, was hit on the playground at lunchtime yesterday by another student, Jack, and came home very upset. You have been unaware of the incident until now, and you donít know whether similar incidents have occurred in the past.

Rate the appropriateness of each of the options in terms of what you should do as a first-year teacher (Inappropriate to Appropriate)

- Look in both of the students' school records to establish if similar incidents have occurred before
- Reassure Mr. Andrews that the incident will be investigated
- Ask Jack's parents to come in for a meeting to discuss Jack's behavior

Fig. 7.1 Example text-based SJT from the Teacher Selection Project

of SJTs and other selection methods in a range of settings. In Klassen et al., 2020, a 25-item SJT was administered online along with two competency-based essays in order to screen candidates for invitation to an 'assessment center interview day'. The assessment center included a 1–1 interview, a 'case study' worked on in small groups, and a teaching demonstration. The SJT was correlated $r = 0.42$, $p < 0.01$ with the teaching demonstration and $r = 0.46$, $p < 0.01$ with the assessment center total (but not with the group case study). The other screening measures (i.e., the competency-based essays) were not significantly correlated with the teaching demonstration ($rs = 0.14$ to 0.20, $ps = ns$) but were correlated with overall assessment center performance ($rs = 0.24, 0.22$). Hierarchical regression showed that scores on the SJT contributed unique variance to the prediction of assessment center performance. The conclusion of this study was that SJTs were a useful screening method—predictive and efficient—in comparison to the other methods used.

A recent study explored how SJTs used for selection into primary and secondary ITE programs predicted teaching performance during a major teaching placement approximately six months after selection (Klassen & Rushby, 2019). The study showed that the 'conventional' methods employed during the selection process—math and English tests, a group problem-solving task, and a 1–1 interview—were not significant predictors of teaching performance on the teaching placement six months after selection (primary program, $rs = -0.13$ to 0.09; secondary program $rs = -0.14$ to 0.15). In contrast, SJTs were significantly predictive of teaching performance both at the primary level ($r = 0.24$, $p < 0.05$) and secondary level ($r = 0.35$, $p < 0.05$). The results from this study showed that an SJT tailored for teacher selection purposes may provide an effective way to systematically evaluate large numbers of applicants to ITE programs.

Video SJTs. Video formats of SJTs are intuitively appealing, with the opportunity to provide applicants with engaging animations or live action (with human actors) as they work through an SJT. Video SJTs, whether using animation or live actors, provide a higher level of realism and allow the test designer to add details (e.g., facial expressions and body language) not easily represented in text. However, video SJTs are expensive to produce, and revisions made to item content can be time-consuming and costly. Recent research has compared video- and text-based SJTs in a range of settings. Lievens and Sackett (2006) compared the predictive validity of video and text SJTs and found that video-based SJTs measuring interpersonal skills had significantly higher predictive and incremental validity than the text SJTs using the same content. The authors suggested that the video format provided extra sources of information, leading to higher accuracy and fidelity. In addition, the text-based SJTs were correlated more strongly with cognitive predictors than were the video SJTs, suggesting that the video format may be a better one for use in assessing non-cognitive variables.

Bardach et al. (2020) from the Teacher Selection Project group recently compared video and text formats for a teacher selection SJT. Prospective teachers were randomly assigned to one of three SJT conditions: 3D animated video with accompanying text, 3D animated video without text, and text only (see Fig. 7.2 for example image). The authors examined how the format of SJTs would be associated with

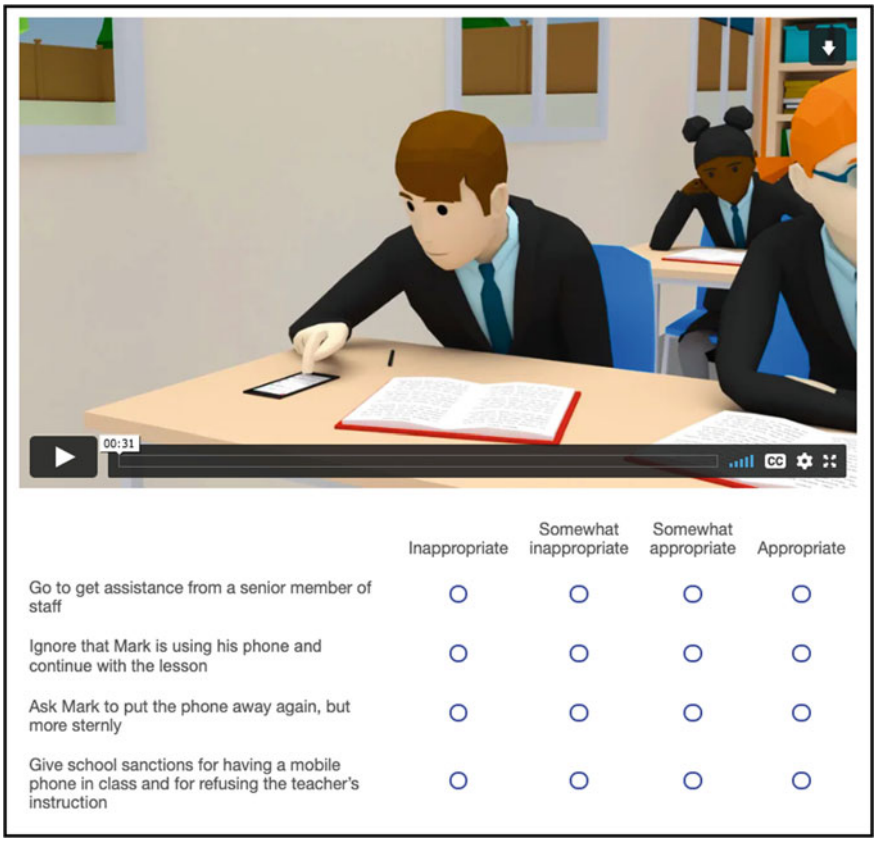


Fig. 7.2 Example video-based SJT from the Teacher Selection Project

applicant reactions and subgroup (gender and ethnicity) differences. No differences in scores between the three formats were found, but participants found the two video conditions more engaging than the text format. Females scored significantly higher than males in the text format SJT, consistent with much of the SJT literature, but that difference disappeared in the video format SJTs, consistent with the findings by Bruk-Lee et al. (2016). Ethnicity effects (participants from majority groups scoring higher than minority groups) were consistent in all three SJT formats, although the mean score differences were not large, typically about 3 points (e.g., for video with text condition, $M_{majority} = 146.4$, $SD = 6.2$, $M_{minority} = 143.5$, $SD = 4.8$). The study concluded that the benefits of video SJTs pertaining to applicant engagement and reduced gender effects should be weighed against the resources (time, money, expertise) needed to produce SJTs in this format. In addition, further exploration was needed to understand the persistence of ethnic group scoring differences.

7.3 Developing SJTs: A Collaborative Approach

Developing SJTs involves creating scenarios and writing response options, with content generated through a collaboration between test-developers and 'subject matter experts', or SMEs. Research has shown that a collaborative approach to item-writing improves the conceptual and psychometric characteristics of test items, and in particular, the item discrimination and reliability of items (Abozaid et al., 2017). For 'traditional' SJTs (i.e., measuring overall situational judgment rather than specific constructs), development typically involves creating scenarios using a 'critical incident' approach, where SMEs (including job incumbents, supervisors, trainees) describe past experiences of specific problems faced during work, the action taken or considered to address the problem, and the outcome of the actions. The construction of each scenario is thus highly contextualized and authentic, ensuring fidelity with the actual workplace. For teacher selection SJTs, SMEs might include experienced teachers, teacher educators, and school leaders who have experience in working with novice teachers. The authenticity of SJTs depends on the 'real-life' experiences of those involved in developing the test content.

One approach to building content for SJTs is to use a 'workshop approach' in which test developers and SMEs work collaboratively to develop test content. The workshops consist of a gathering of experienced educators, primarily teachers, principals, and teacher educators who are brought together to determine target attributes, and to develop and test content. An important consideration when setting up the workshops is to ensure that the SMEs invited to participate reflect the diversity of the potential applicants; teacher workforces are frequently unrepresentative of the general population of the students they teach (Carter Andrews et al., 2019). Inviting a diverse SME group helps ensure that the content of the selection methods represents the target population, and by extension, the teaching workforce. Although organizational psychologists frequently develop SJT content through paired writing where SMEs collaborate with item-writers in dyads, a workshop with multiple participants and open communication can deliver higher quality content—at least at the early stages—for complex, multi-faceted professions like teaching, where contextual differences are marked. The development of SJTs for teacher selection is carried out in three phases—identifying target attributes, creating content, and pilot-testing and administration—with eight steps (see Fig. 7.3 for the proposed framework). Each of the steps is carried out through close collaboration between education experts and the test developers.

Step 1: Test specification. The first step in developing an SJT for teacher selection is to specify the purpose of the test, the feasible length of the test, the item types and response formats, how the test will be administered, and the kinds of information that need to be generated by test administration. Important considerations at this step are delivery method (online, paper-and-pencil, proctored, unsecured at home), and response formats (rating, ranking, best and worst, etc.). Delivery methods have become more streamlined in recent years, and online test delivery, whether the test is administered on-site or remotely, is infinitely more desirable for reasons of data

	Identifying target attributes		Creating content			Pilot testing and administration		
Task	**Step 1:** Test specification	**Step 2:** Select and define key attributes	**Step 3:** Review of existing items in target context	**Step 4:** Generate new items with review	**Step 5:** Determine scoring key.	**Step 6:** Design and pilot SJT	**Step 7:** Psychometric analysis	**Step 8:** Development of item bank
Who?	Program leaders	Workshop	Workshop	Workshop	Review panel	Remote delivery	SJT consultants	SJT consultants

Fig. 7.3 Proposed framework for selection of teachers for training and professional practice

storage, scoring, and general ease-of-use. The response format issue has been the subject of considerable research, with the overall finding that a rating format is preferable over ranking or 'pick best, pick worst' formats (Arthur et al., 2014).

Step 2: Identify and define key attributes. In most cases, the key purpose for implementing SJTs for teacher selection is to evaluate applicants' non-cognitive attributes—the 'soft' skills that are so difficult to assess in a reliable and valid fashion at interview. SJTs can be designed to measure a range of attributes depending on how they are developed, but most researchers agree that SJTs measure an individual's awareness and judgment about effective behavior in specific situations. Assessing applicant judgment in a fair and reliable way is the hallmark of SJTs, but which attributes are best targeted when constructing SJTs? An important development step is to identify a set of key attributes on which to build scenario content. A brainstorming session can be part of SJT development; in a workshop environment, the questions are *What are the key non-cognitive* (or *non-academic*) *attributes of novice teachers? What are the attributes that are necessary for the success of trainees? Are there particular attributes that are especially important in this context?* The results from the brainstorming session (see Fig. 7.4 below) are collated and assessed through a content analysis of collected data, with a frequency analysis indicating the agreed upon attributes on which to build scenarios. It is suggested that between 3–6 attributes or attribute clusters be targeted for scenario development.

Step 3 (optional): Review of existing items in target context. Using an 'off-the-shelf' SJT for selecting teachers offers certain advantages, such as proven psychometric properties, efficiencies of cost and time, and content that has been shown to evoke positive applicant reactions. However, adaptations to existing items may be necessary if the target context differs in significant ways from the context in which the test was originally developed. The level of adaptations depends on the degree of contextual differences: a review of scenarios, response options, and scoring will usually highlight any changes needed, which may range from terminology (replacing 'headteacher' for 'principal', for example when adapting a UK-developed SJT for use in Canada) to revisions of scenarios, response options, and scoring if the cultural distance is considerable. In the section 'Adapting content for new settings', we provide a more thorough description of the adaptation process when cultural distance is large.

Fig. 7.4 Results from brainstorm session on key non-cognitive attributes for novice teachers

Step 4: Generate new items using a 'critical incidents' approach. An SJT item comprises a scenario that describes a realistic work-place situation, and a set of plausible options for responding to the scenario. Developing new items for teacher selection SJTs demands a knowledge of the relevant context, a knowledge of typical challenges facing new teachers, and the relevant experience needed to understand the pros and cons of various response options. Scenarios are typically built using a critical incidents approach (e.g., Buyse & Lievens, 2011) in which instructions are sent to workshop participants along the lines of 'We are developing a teacher selection tool that focuses on the non-academic attributes associated with successful teaching (e.g., empathy, conscientiousness, adaptability). In our upcoming workshop, we will ask you to share scenarios of incidents that are related to these attributes. The scenarios should reflect situations that novice teachers encounter and should be related to one of the target attributes'. During the workshop, participants generate scenarios and response options that are subsequently reviewed by test developers (to eliminate errors, inappropriate and redundant items, and items that do not map onto the target attributes).

Step 5. Conduct a review panel to set scoring. The purpose of the review panel is to carefully scrutinize the scenarios and response options generated in Step 4, and to answer the questions *Are the items set in the correct context? Are the response options feasible and set at an appropriate level for a novice teacher? Does the content depend too heavily on specific procedural knowledge?* The reviewed and revised items are then tested in a concordance panel (which can be delivered remotely) in which experienced teachers complete the prototype SJT to determine the level of consensus of scoring of the SJT and to provide additional feedback on the items. Items with high consensus are retained for use in next steps; items with low consensus are revised and assessed in an iterative manner.

Step 6. Design and pilot SJT. Items that fare well in Step 5 are tabulated against the target attributes, and a representative selection of items are included in the pilot SJT and administered either to (a) 'incumbents' (i.e., students already admitted to ITE programs, or beginning teachers who are working in schools) or to (b) applicants for ITE programs. When pilot SJTs are delivered to applicants, the pilot tests are prefaced with a statement indicating that completion is voluntary and will not affect their application status. Applicant reaction data is importantly collected at this stage, and concurrent validity data (i.e., interview scores, academic data, teaching performance data if incumbents). After administration, scoring keys may be adjusted based on psychometric analysis in Step 7.

Step 7: Psychometric analysis. Once the data is collected, scoring keys are set using one of four approaches: rational, theoretical, empirical, or integrated. A rational approach uses an SME consensus approach as discussed in Step 5. A theoretical approach builds a scoring key based on what theory suggests is the 'best' course of action in a situation. An empirical approach is determined by evaluations of the relations between applicant responses and an external criterion (e.g., other interview scores, teaching ratings). An integrated approach sets the initial scoring key using a rational (or theoretical) approach, and then revises the key based on empirical results.

Analysis of SJT data typically includes a measure of reliability (internal consistency, test–retest, or split-half), item difficulty, and concurrent, construct, and/or predictive validity. Reliability estimates (consistency of measurement) are notoriously fickle for SJTs due to their multidimensional nature, and reliability indices other than conventional Cronbach's alpha are recommended (e.g., test–retest, split-half using the Spearman-Brown prophecy formula), and if alpha is used, should be considered a lower bound of the reliability estimate (Whetzel et al., 2020).

Step 8: Development of item bank. Items that are not included in the development of the SJT will be retained in an item bank, with some items rated as 'good' (items with scoring consensus and acceptable psychometric properties) that can be included in future test versions, and some items rated as 'needing further work' that can be revised and re-piloted or discarded. A functioning item bank is important to develop future iterations of the test and is important to bolster test security.

7.4 Adapting Content for New Settings

We saw in Chap. 3 how culture—the shared beliefs, goals, and values that guide the way we think and behave—influences the education environment and even influences the way the personal characteristics deemed necessary for novice teachers' success are viewed. For teacher selection methods, it is not a case of *one-size-fits-all* when it comes to using off-the-shelf selection methods, and although methods such as SJTs or MMIs that are effective in one setting may prove effective in other settings, the *content* of these methods needs to reflect the cultural, social, and educational context. In a study exploring the feasibility of using an American-developed SJT assessing integrity in a Spanish context, Lievens et al. (2015) found that most of the scenarios (84%) were deemed to be realistic by Spanish test-takers, with similar relations to external criteria. Herde et al. (2019) tested the measurement invariance of five SJTs testing 'twentieth century skills' (e.g., achieving objectives, adapting to change) across multiple countries in Europe and Latin America, and found the same latent factorial structure (and similar internal consistency coefficients) across regional groups, suggesting participants interpreted the SJT scenarios and response options in the same way. Nevertheless, merely translating SJT content into a new language is likely to be insufficient especially with high levels of 'cultural distance'; a deep knowledge of the target culture is needed to ensure that items reflect the social, cultural, and educational norms that may be different from those in the context where the test was first developed.

Especially in the case where cultural distance is great, SJT development requires socio-cultural awareness and collaboration with partners in the target setting. A true partnership is needed: developers of selection tests who are serving as consultants will have only moderate knowledge of the cultural setting; education experts in the target setting may have only moderate knowledge of test-writing. Ryan and Brunfaut (2016) used a case study approach to better understand how to conduct (language) test development work in cultural settings that are unfamiliar to the test developers. They offered several suggestions to maximize the chance of effective test-writing. First, the test developers benefit from preliminary work to increase familiarization with the target language and culture, ensuring a level of basic knowledge about the language and socio-cultural and educational context. In parallel, the education experts in the target culture benefit from preliminary work on familiarization with test-development principles. In the case of SJT writing, education experts can be provided with sample items, critical incident prompt materials, and a description of key psychometric considerations including item discrimination, reliability, item difficulty, and predictive validity.

Researchers at the Teacher Selection Project have developed a range of teacher selection tools across cultural contexts, not only in the UK, but in settings that are culturally distant from their original work in England (e.g., Bulgaria, Lithuania, Malawi, and Morocco). The process of adaptation began with a determination of appropriateness of key non-cognitive attributes. A cross-national comparison

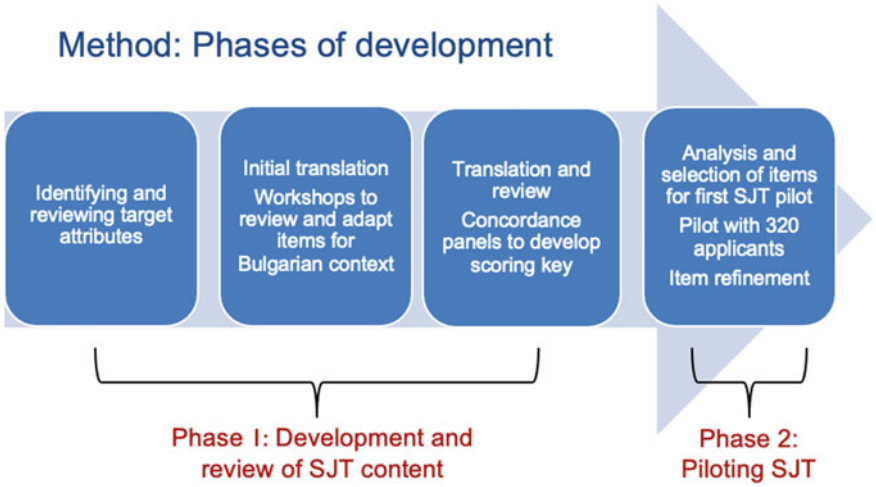


Fig. 7.5 Adapting SJTs for Bulgarian context

(Klassen et al., 2018) found that core non-cognitive attributes (i.e., communication, adaptability, organization) identified in England were endorsed across culturally disparate settings (i.e., Finland, Malawi, Oman), but each non-English setting also proposed additional non-cognitive attributes believed to be essential to successful novice teaching. In line with the 2018 study, some non-cognitive attributes of successful novice teachers seem to be universal, and some appear to be context- and culture-specific.

Case study: Bulgaria. An existing English-language SJT originally developed in the UK was adapted for use in Bulgaria (Rushby & Klassen, 2019). Figure 7.5 presents the two phases of SJT development, with Phase 1 involving the development and review of SJT content, and Phase 2 including the pilot-testing and revision of content. After reviewing and confirming target attributes established in previous work (i.e., empathy and communication, organization and planning, resilience and adaptability), the translated SJTs were adapted. The translation of existing items included surface changes to scenarios (school contexts, teaching roles) and deeper changes to item responses (desired and undesired options for responses).

A review panel to establish the scoring key that included 28 subject matter experts (SMEs) was conducted to review the existing scoring and to identify how expert teachers in Bulgaria evaluated scenarios and scoring originally developed in England. In cases where scenarios were not substantially changed (i.e., apart from names and job titles), most of the scoring (57.6%) of the responses was the same across contexts, about one-third of responses (28.6%) were one position away (e.g., 'appropriate' in the UK setting; 'somewhat appropriate' in Bulgaria), 6.2% of responses were two positions away, and one response was three positions away (e.g., 'inappropriate' in UK; 'appropriate' in Bulgaria). The scenario that showed the greatest cross-cultural

difference was a classroom situation where a teacher assistant was routinely interrupting and correcting a teacher's lessons: confronting the assistant in front of the students was deemed 'appropriate' by most Bulgarian SMEs, but 'inappropriate' by most UK SMEs. Identifying and discussing these differences in interpersonal relationships was an important part of the adaptation process in this context.

Case study: rural and remote Australia. Durksen and Klassen (2018) developed an SJT to promote the key characteristics needed for rural and remote Australian settings. In Australia, the turnover rate for teachers in rural and remote settings is up to six times higher than in city schools, with many new teachers in remote regions leaving their posts before the end of their contracted teaching (Lyons, 2006). In this project sponsored by the New South Wales Department of Education, the authors began by evaluating the key clusters of non-cognitive attributes that had been developed in the UK: resilience and adaptability, organization and planning, and empathy and communication. A review panel of experienced teachers concluded that the non-cognitive attributes previously identified in the UK were universally salient in a remote and rural setting, but that a new attribute cluster—'culture and context' was necessary to capture the unique qualities needed for success in the target culture. The new cluster was defined as 'The capability to adapt to remote settings, recognize the importance of building relationships and maintaining professional behavior in all aspects of community life. Demonstrates sensitivity to cultural knowledge and practice'. After identifying the key attributes needed for success in remote settings, an item-writing workshop resulted in 37 new NSW-specific items and some minor revisions to the existing 32 UK items trialed in earlier stages. Although the cultural distance between the UK and Australia was not as great as between the UK and Bulgaria, important cultural differences were identified, and SJT development needed to reflect these differences.

7.5 Chapter Summary

Current teacher selection tools are not always reflective of the most recent research in organizational or educational psychology, but recent studies have shown that SJTs provide an evidence-supported alternative for organizations looking to select effective teachers (Klassen & Kim, 2019). In this chapter we first explored the use of SJTs for selection in detail, and considered how SJTs could be built deductively, i.e., using a top-down, construct-driven approach, or inductively, i.e., using a bottom-up, inductive approach in collaboration with subject matter experts. A blueprint for developing SJTs was provided alongside a discussion on how to adapt SJTs to new contexts. In the next chapter we examine another research-supported method for selection—multiple mini-interviews, or MMIs—that are now being implemented and tested for teacher selection in the UK and Finland.

References

Abozaid, H., Park, Y. S., & Tekian, A. (2017). Peer review improves psychometric characteristics of multiple choice questions. *Medical Teacher, 39*, S50–S54.

Arthur, W., Jr., 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*, 1–11.

Bandura, A. (1999). A social cognitive theory of personality. In L. Pervin & O. John (Eds.), *Handbook of personality* (2nd ed., pp. 154–196). Guilford Publications.

Bardach, L., Rushby, J. V., Kim, L. E., & Klassen, R. M. (2020). Using video- and text-based situational judgement tests for teacher selection: A quasi-experiment exploring the relations between test format, subgroup differences, and applicant reactions. *European Journal of Work and Organizational Psychology*. https://doi.org/10.1080/1359432X.2020.1736619

Bateson, J. E., Wirtz, J., Burke, E. F., & Vaughan, C. J. (2014). Sifting to efficiently select the right service employees. *Organizational Dynamics, 43*, 312–320.

Bowles, T., Hattie, J., Dinham, S., Scull, J., & Clinton, J. (2014). Proposing a comprehensive model for identifying teaching candidates. *The Australian Educational Researcher, 41*, 365–380.

Boyatzis, R. E., & Kelner, S. P. (2010). Competencies as behavioral manifestation of implicit motives. In O. C. Schultheiss & J. C. Brunstein (Eds.), *Implicit motives* (pp. 488–509). Oxford University Press.

Bruk-Lee, V., Lanz, J., Drew, E. N., Coughlin, C., Levine, P., Tuzinski, K., & Wrenn, K. (2016). Examining applicant reactions to different media types in character-based simulations for employee selection. *International Journal of Selection and Assessment, 24*, 77–91.

Buyse, T., & Lievens, F. (2011). Situational judgment tests as a new tool for dental student selection. *Journal of Dental Education, 75*, 743–749.

Carter Andrews, D. J., Castro, E., Cho, C. L., Petchauer, E., Richmond, G., & Floden, R. (2019). Changing the narrative on diversifying the teaching workforce: A look at historical and contemporary factors that inform recruitment and retention of teachers of color. *Journal of Teacher Education, 70*, 6–12.

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*, 334–346.

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*, 83–117. https://doi.org/10.1111/j.1744-6570.2009.01163.x

Clevenger, J., Pereira, G. M., Wiechmann, D, Schmitt, N., & Harvey, V. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology, 86*, 410–417. http://dx.doi.org/https://doi.org/10.1037/0021-9010.86.3.410

Durksen, T. L., & Klassen, R. M. (2018). The development of a situational judgment test (SJT) of personal attributes for quality teaching in rural and remote Australia. *The Australian Educational Researcher, 45*, 255–276.

Elliott, J. G., Stemler, S. E., Sternberg, R. J., Grigorenko, E. L., & Hoffman, N. (2011). The socially skilled teacher and the development of tacit knowledge. *British Educational Research Journal, 37*, 88–103. https://doi.org/10.1080/01411920903420016

Fan, J., Stuhlman, M., Chen, L., & Weng, Q. (2016). Both general domain knowledge and situation assessment are needed to better understand how SJTs work. *Industrial and Organizational Psychology, 9*, 43–47.

Freudenstein, J.P., Schäpers, P., Roemer, L., Mussel, P., & Krumm, S. (2020). Is it all in the eye of the beholder? The importance of situation construal for situational judgment test performance. *Personnel Psychology,* DOI: https://doi.org/10.1111/peps.12385.

Harris, A. M., Siedor, L. E., & Listyg, & Carter, N. T. . (2016). In defense of the situation: An interactionist explanation for performance on situational judgment tests. *Industrial and Organizational Psychology, 9*, 23–28.

Herde, C. N., Lievens, F., Solberg, E. G., Harbaugh, J. L., Strong, M. H., & Burkholder, G. J. (2019). Situational judgment tests as measures of 21st century skills: Evidence across Europe and Latin America. *Journal of Work and Organizational Psychology, 35*, 65–74.

Hooper, A. C., Jackson, H. L., & Motowidlo, S. J. (2004). *Situational judgment measures of personality and work-relevant performance.* Paper presented at the 112th annual meeting of the American Psychological Association, Honolulu, HI.

Kasten, N., & Freund, P. A. (2016). A meta-analytical multilevel reliability generalization of situational judgment tests (SJTs). *European Journal of Psychological Assessment, 32*, 230–240.

Klassen, R. M., Durksen, T. L., Al Hashmi, W., Kim, L. E., Longden, K., Metsäpelto, R.-L., Poikkeus, A. M., & Györi, J. (2018). Cultural context and teacher characteristics: Exploring the non-cognitive attributes of prospective teachers in four countries. *Teaching and Teacher Education, 72*, 64–74.

Klassen, R. M., Durksen, T. L., Kim, L. E., Patterson, F., Rowett, E., Warwick, J., Warwick, P., & Wolpert, M. A. (2017). Developing a proof-of-concept selection test for entry into primary teacher education programs. *International Journal of Assessment Tools in Education, 4*, 96–114.

Klassen, R. M., Durksen, T. L., Rowett, E., & Patterson, F. (2014). Applicant reactions to a situational judgment test used for selection into initial teacher training. *International Journal of Educational Psychology, 3*, 104–125. https://doi.org/10.4471/ijep.2014.07

Klassen, R. M., & Kim, L. E. (2019). Selecting teachers and prospective teachers: A meta-analysis. *Educational Research Review, 26*, 32–51.

Klassen, R. M., Kim, L. E., Rushby, J., & Bardach, L. (2020). Can we improve how we screen applicants for initial teacher education? *Teaching and Teacher Education, 87,* 102949.

Klassen, R. M., & Rushby, J. V. (2019, January). Can we predict preservice teachers' performance in teaching placements? The validity of ITE selection methods. *TSP Working Paper.* https://bit.ly/2SeQkVv

Koczwara, A., Patterson, F., Zibarras, L., Kerrin, M., Irish, B., & Wilkinson, M. (2012). Evaluating cognitive ability, knowledge tests and situational judgement tests for postgraduate selection. *Medical Education, 46*(4), 399–408.

Lievens, F., & De Soete, B. (2015). Situational judgment tests. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (2nd ed., Vol. 22, pp. 13–19). Elsevier.

Lievens, F., & Motowidlo, S. J. (2016). Situational judgment tests: From measures of situational judgment to measures of general domain knowledge. *Industrial and Organizational Psychology, 9*, 3–22.

Lievens, F., & Sackett, P. R. (2006). Video-based versus written situational judgment tests: A comparison in terms of predictive validity. *Journal of Applied Psychology, 91*, 1181–1188.

Lievens, F., & Sackett, P. R. (2012). The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance. *Journal of Applied Psychology, 97*, 460–468.

Lievens, F., Corstjens, J., Sorrel, M. A., 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*, 361–372.

Lyons, T. (2006). Science, ICT and mathematics education in rural and regional Australia, *SiMERR National Research Centre [Report].* http://apo.org.au/node/2732.

Patterson, F., Ashworth, V., Mehra, S., & Falcon, H. (2012). Could situational judgement tests be used for selection into dental foundation training? *British Dental Journal, 213*, 23–26.

Patterson, F., Lievens, F., Kerrin, M., Munro, N., & Irish, B. (2013). The predictive validity of selection for entry into postgraduate training in general practice: evidence from three longitudinal studies. *British Journal of General Practice,* November, e734-e741. doi: https://doi.org/10.3399/bjgp13X674413

Patterson, F., Tavabie, A., Denney, M, Kerrin, M., Ashworth, V., Koczwara, A., & MacLeod, S. (2013). A new competency model for general practice. *British Journal of General Practice*, May, e331–e338. doi: https://doi.org/10.3399/bjgp13X667196

Patterson, F., Zibarras, L., & Ashworth, V. (2015). Situational judgement tests in medical education and training: Research, theory and practice: AMEE Guide No. 100. *Medical Teachers, Aug 27,* 1–15 E-pub ahead of print.

Patterson, F., Zibarras, L., Carr, V., Irish, B., & Gregory, S. (2011). Evaluating candidate reactions to selection practices using organisational justice theory. *Medical Education, 45*, 289–297.

Rushby, J. V., & Klassen, R. M. (2019, September). *The cross-cultural transferability of a situational judgment test used for initial teacher selection.* Paper presented at the British Psychological Society Psychology of Education Section annual meeting: Manchester, UK.

Ryan, E., & Brunfaut, T. (2016). When the test developer does not speak the target language: The use of language informants in the test development process. *Language Assessment Quarterly, 13*, 393–408.

Shultz, M. M., & Zedeck, S. (2012). Admission to law school: New measures. *Educational Psychologist, 47*, 51–65. https://doi.org/10.1080/00461520.2011.610679

Tiffin, P. A., Paton, L. W., O'Mara, D., MacCann, C., Lang, J. W., & Lievens, F. (2020). Situational judgement tests for selection: Traditional vs construct-driven approaches. *Medical Education, 54*(2), 105–115.

Whetzel, D. L., & McDaniel, M. A. (2009). Situational judgment tests: An overview of current research. *Human Resource Management Review, 19*, 188–202. https://doi.org/10.1016/j.hrmr.2009.03.007

Whetzel, D. L., Sullivan, T. S., & McCloy, R. A. (2020). Situational judgment tests: An overview of development practices and psychometric characteristics. *Personnel Assessment and Decisions, 6*, 1–16.

Chapter 8
Developing Multiple Mini-Interviews for Teacher Selection

Abstract Although SJTs are useful for screening large numbers of applicants, many selection programs also need more intensive methods that can be used for decision-making on a smaller scale. Earlier in the book we were introduced to multiple mini-interviews (MMIs), a method that incorporates a circuit of structured, independent interview stations to assess the non-cognitive attributes of applicants in a range of fields. In this chapter we will look more closely at MMIs: how they have been used outside of education, how they are used for teacher selection, and how MMIs might be developed for ITE programs.

In the previous chapter, we looked at situational judgment tests (SJTs) and how they can be developed and used for teacher selection in a range of settings. SJTs are useful for screening large numbers of applicants, but many selection programs also need more intensive methods that can be used for decision-making on a smaller scale. Earlier in the book we were introduced to multiple mini-interviews (MMIs), a method that incorporates a circuit of structured, independent interview stations to assess the non-cognitive attributes of applicants in a range of fields. In this chapter we will look more closely at MMIs: how they have been used outside of education, how they are used for teacher selection, and how MMIs might be developed for ITE programs.

8.1 The Problem with Conventional Interviews

Selecting prospective teachers for training or employment often involves a face-to-face interview to provide a detailed look at applicants' cognitive and non-cognitive attributes. In some settings (such as the UK), 'traditional' interviews are one of the most commonly used methods of selection (Davies et al., 2016), and range from unstructured to highly structured. In terms of reliability and validity, a highly structured interview is desirable (Patterson et al., 2016). A serious problem in teacher selection interviews is that interviews are not highly structured, and *intuition* about a teachers' future effectiveness often plays a critical role (e.g., Davies et al., 2016),

R. M. Klassen and L. E. Kim, *Teacher Selection: Evidence-Based Practices*,
https://doi.org/10.1007/978-3-030-76188-2_8

even when the evidence for the predictive role of intuition making selection decisions is not strong (Kausel et al., 2016). In some contexts, an applicant's success when applying for teacher education depends on successfully navigating the selection interview, but there is almost no empirical evidence that current teacher selection interviews are reliable and valid.

Many selectors in education are confident in their abilities to detect teaching potential (e.g., Davies et al., 2016), but Dana et al. (2013) called this confidence in interviewing the "persistence of an illusion". Kausel et al. (2016) studied confidence and decision-making in selection practices and found that although selectors were often confident that they were making good decisions, the predictive validity of their decisions was low. Furthermore, selectors frequently possess an *illusion of understanding* (see Kahneman, 2011), and lack the metacognitive awareness to understand the accuracy, or inaccuracy, of their selection decisions. Such overconfidence in selection means that selection decisions may be compromised, unfair, and biased. In addition, some interviewers tend to be stricter than others: a particular interviewer may have a tendency or bias to issue high scores (or low scores) to all applicants they encounter, known as the 'hawk and dove' phenomenon.

The hawk and dove phenomenon has the potential to influence an applicant's chances of success solely based on random assignment of interviewer (Kiraly et al., 2020). With only one interview, an applicant's score is largely due to chance: a lucky applicant is randomly assigned to an interviewer who is an easy rater (a 'dove') and who has a similar background and experiences. An unlucky applicant faces a hard rater (a 'hawk') who is not familiar with the applicant's background and experiences. A 'strong hawk' or 'strong dove' can also influence panel interviews, thus influencing the fairness of selection decisions (Kiraly et al., 2020). Over the last 20 years, new interview approaches have been developed to help alleviate the unreliability of conventional interviews. MMIs are an interview method that are at the forefront of the efforts to improve the fairness of selection interviews, and after being used almost exclusively for selection in health-related fields, are beginning to be used for selection in teacher education.

8.2 Multiple Mini-Interviews (MMIs)

The push to develop a better interview process for selection into medical school was advanced by the work of Eva et al. (2004), who developed and tested MMIs in an attempt to improve the reliability, validity, and especially, the *fairness* of selection interviews. For many years, selection panels in medical schools relied on a combination of academic performance and other cognitive measures (such as the Medical College Admissions Test; MCAT) and face-to-face interviews to assess non-cognitive attributes, either with a single interviewer or a panel of interviewers. Although conventional interviews for entrance into medical school showed high face validity, there were several problems with the process. For example, the interviewers were typically practitioners or medical school faculty who unconsciously

(or consciously) favored applicants who 'looked like them', resulting in a lack of heterogeneity in the pool of successful candidates (Burke, 2004). Bias in interviews is a longstanding problem, and conventional interviews tend to be variable and highly subjective (Lemay et al., 2007). MMIs were developed to alleviate the inherent unreliability of conventional interviews.

The distinguishing features of MMIs are that: (a) they are objective (i.e., highly structured), (b) each station is brief (typically 4 to 8 min), (c) the stations are independent (i.e., each station is a 'new start' and interviewers are blind to performance on previous stations), (d) there are multiple stations (ranging from 3–10 depending on time and resources), and (e) each station targets a key non-cognitive attribute (e.g., integrity, empathy or other attribute) determined by the organization to be essential. Reliability and validity studies have shown that MMIs used for medical school admissions show moderate-to-high reliability (e.g., Pau et al., 2013), with internal consistency coefficients ranging from 0.69 to > 0.90. Studies examining validity have shown that MMIs for medical selection predict patient interaction in medical school ($r = 0.65$), 'communication, cultural, legal, and ethical aspects of medicine' ($r = 0.44$), and score on a licensing exam ($r = 0.36$) five years after admission (Eva et al., 2009). In a study investigating MMIs used for selection into occupational therapy training, Thomas et al. (2017) found that MMIs were more predictive of trainee performance than traditional interview scores. Research has shown that MMIs do not correlate highly with academic qualifications such as GPA or admissions tests (Pau et al., 2013), suggesting that MMIs are measuring attributes conceptually and empirically separate from the cognitive attributes assessed at selection.

MMIs used for selection in health professions. In the last few years, interest in implementing MMIs has increased rapidly in the health professions, with the methodology used for selection into a range of fields including medicine, midwifery, nursing, pharmacy, and dentistry. In the early work on MMIs, Eva and colleagues developed 10 brief stations modeled on the clinical skills assessment, the 'OSCE' (objective structured clinical examination) which has been in use for medical training since the 1970s. The original MMI was focused on four domains—critical thinking, ethical decision making, communication skills, and knowledge of the health care system—assessed across 10 stations, including one station presenting a 'standard interview' question (i.e., *Why do you want to be a physician? Discuss this question with the interviewer*). Other stations variously presented ethical dilemmas (*Should homeopathic treatments be recommended if patients believe they work?*), knowledge of health care (*Should GP visit fees be introduced to deter unnecessary visits?*), and interactions with actors in role-plays to assess communication skills. Most of the research on MMIs has been conducted in medical schools, beginning in Canada (Eva et al., 2004), but now the method is widely used across the world, and in a range of different fields such as midwifery, nursing, and dentistry.

In the UK, Callwood et al. (2014) developed an eight-station MMI for the selection of midwife trainees, with stations representing compassion and empathy, respect for diversity, integrity, intellectual curiosity, advocacy, respect for privacy, team working, and motivation to become a midwife, with communication assessed at each station. Internal consistency was high (Cronbach's alpha 0.91 to 0.97), and in light of the

positive results, the midwifery training program adopted the MMIs for 'live' selection after piloting. In nursing, Perkins et al. (2013) assessed applicant and interviewer reactions to the implementation of an MMI, with 7% of applicants (5% of interviewers) rating the MMI as a 'worse experience' than other methods, 27% (24% of interviewers) rating it as 'neither better nor worse', and 65% (71% of interviewers) rating the MMI as a better experience overall. In the field of dentistry, McAndrew and Ellis (2011) analyzed the free text responses of 190 dental school applicants who had completed an MMI, and identified four key themes: (a) lack of control (applicants reported not being able to display their best side), (b) anxiety and nervousness (anxiety about poor performance at one station influenced performance on subsequent stations), comparisons with traditional interviews (generally positive about the chance to rebound from a 'bad station'), and preparedness (difficulty in preparing for the MMI). Overall, most applicants (64.8%) found the MMIs to be preferable to traditional interviews, with 25% ambivalent, and 10% rating MMIs worse than other, more conventional interviews that they had encountered.

Virtual or online MMIs. In some cases, face-to-face MMIs are not feasible for all applicants. For example, during the Covid-19 crisis of 2020–2021, some institutions completed some or all of their selection processes virtually, with applicants completing stations through video-conferencing software, and admissions decisions depending solely on virtual interactions between interviewers and applicants. Cleland et al. (2020) noted that although MMIs are inherently a face-to-face method, it is possible to adapt the method for online delivery. In their online MMI for medical school selection in Singapore, they dropped stations involving role-playing with simulated patients, and reduced the number of stations from eight to five. They used the 'breakout room' function on video-conferencing software Zoom to imitate the station format of their onsite MMIs. The authors suggested that communication was a critical factor for the smooth running of the process, with detailed technical and contingency guidance provided to interviewers and applicants. Similarly, Ungtrakul et al. (2020) found that a virtual MMI was feasibly implemented during Covid-disrupted MMIs for medical school selection in Thailand, with only minor adaptations necessary for station content. One question not yet considered in the Covid-related MMI research is that of applicant reactions. Blacksmith et al. (2016) found that job applicants prefer face-to-face over virtual interviews, due to reduced opportunity to demonstrate social skills in the virtual environment. Applicant reactions to MMIs are generally positive, but further work is needed to ensure that virtual MMIs are sustainable in the long-term and provide a positive applicant experience.

Applicant reactions to MMIs. Research on applicant reactions to MMIs shows a generally favorable response to the method. In the systematic review conducted by Kelly et al. (2018), applicants generally found MMIs as fair, relatively free of bias, and providing reasonable scope for presenting their abilities and strengths. The authors found that applicants who had experienced both traditional interviews and MMIs tended to prefer MMIs because of the independence of stations (interviewers and content), and the opportunity to 'rebound' if they performed poorly at one station. Similarly, O'Brien et al. (2011) compared the perceptions of applicants who had experienced both interview formats. In general, applicants expressed

negative reactions to traditional interviews (because they could not show their true potential; it was difficult to understand the questions) and positive reactions to the MMIs (they were less threatening, more relaxed, and more accurate representation of capabilities), although they reported both formats as 'fair and accurate'. Although reliability, validity, and fairness are critically important when designing selection methods, application reactions are also important, and evidence from medical and other health-related programs suggests MMIs are generally well-received. In education, these same factors are important when choosing selection methods, but the MMI research base is much less well-developed.

8.3 MMIs for Selection in Teacher Education

There are important advantages to implementing MMIs for teacher selection: they are fairer, more valid, and better-received by applicants than traditional interviews. However, MMIs are a resource-intensive selection method demanding careful planning and consideration of the logistics of implementation. Rosenfeld et al. (2008) compared the infrastructure requirements of MMIs to conventional interviews. They found that MMIs were more reliable and more predictive than traditional interviews, but acknowledged that feasibility of use was a key consideration when deciding whether or not to implement MMIs. Implementing MMIs requires greater planning and preparation and more rooms for interviewing than conventional interviews, but these 'costs' are offset by requiring fewer person-hours of effort. Demands on staff involvement are heavy for most kinds of interview formats, and the multi-station format of MMIs allows for the involvement of community and incumbent student assessors to reduce the interviewing load on program staff. Including community and student assessors has been successfully implemented in medical education (e.g., Dowell et al., 2014), although to our knowledge has not yet been trialed in teacher education programs.

Implementing MMIs in Finland. Teacher education programs have only recently begun to implement MMIs for selection, with two main centers of research: Finland and the UK. Metsäpelto et al. (2020) designed a five-station MMI for selecting applicants into a competitive ITE program at a university in central Finland. The five highly structured stations (assessing social skills, cultural competence, motivation for teaching, managing emotions, and collaboration) each lasted five minutes, with a three-minute turnaround time between stations. Interviewers received a four-hour training session focusing on the aims of the MMI, and training on administration and scoring the stations. Results showed that the interviewer effect was modest on 3/5 stations with intra-class correlations < 0.10, indicating that the station variance was attributable to applicant differences rather than interviewer effects.

Applicants and interviewers in Finland perceived the MMI to be a fair means of assessment, with interviewers supporting the MMI format as easier to implement than previously used panel interviews. Applicant and interview reactions were generally positive (e.g., mean scores around 4.0 on a 1–5 scale) for face validity and fairness,

whereas scores for perceived predictive validity of the methodology were lower for both applicants and interviewers ($M = 2.66$ for applicants, and 2.90 for interviewers). The study confirmed the importance of rigorous development of selection methods in general and supported the further development of MMIs for teacher selection in the Finnish context. Subsequent to the research reported in this article, it was determined that MMIs would be implemented nationally for selection into all of the ITE programs in Finland (personal correspondence, Metsäpelto, 2020).

Implementing MMIs in the UK. Recent work by the Teacher Selection Project (https://www.teacherselect.org/) in the UK has resulted in MMIs developed and implemented for teacher selection at two large ITE programs. At Site A, researchers and ITE selection staff worked together on MMI design, logistics, staffing, and station resources. Three stations were developed in collaboration with ITE staff, piloted with students, and then implemented:

- Values and beliefs (three teacher profiles)
- Diversity and social justice (school play)
- Professional awareness (card sort professionalism)
- (Communication, assessed at all 3 stations)

Scoring for the MMIs was completed by assessors on tablets or laptop computers with scores delivered within 24 h to the ITE program. Results from applicants were generally positive, although feedback from assessors was more mixed (see detailed description in the next section).

At Site B six domains were developed with program staff, piloted with a small group of current students, and implemented for selection. The five stations were designed to assess:

- Motivation and commitment
- Learning expectations
- Reflective approach
- Integrity and ethics
- Intellectual curiosity
- Communication (assessed at all five stations).

The MMIs at Site B used a 'low tech' approach, with pencil-and-paper used to record and collate scores. Applicants were positive about the MMI approach to interviewing, and expressed supportive views about the ITE program's implementation of the method: *Nice to have a fresh start at each station… It was a speedy process; I didn't have time to feel nervous*, and *It was good to have a range of attributes that could be assessed as this made it feel like a more thorough process.*

At both sites, the development of the MMI followed the same eight-step process which began with the identification of key attributes and ended with implementation and review of the method (Fig. 8.1). The steps for developing MMIs are similar to those followed in SJT development, i.e., with both methods built on a foundation of agreed-upon key attributes, but with the added complication of consideration of logistical issues involving an assessment of available resources: time, space, and people.

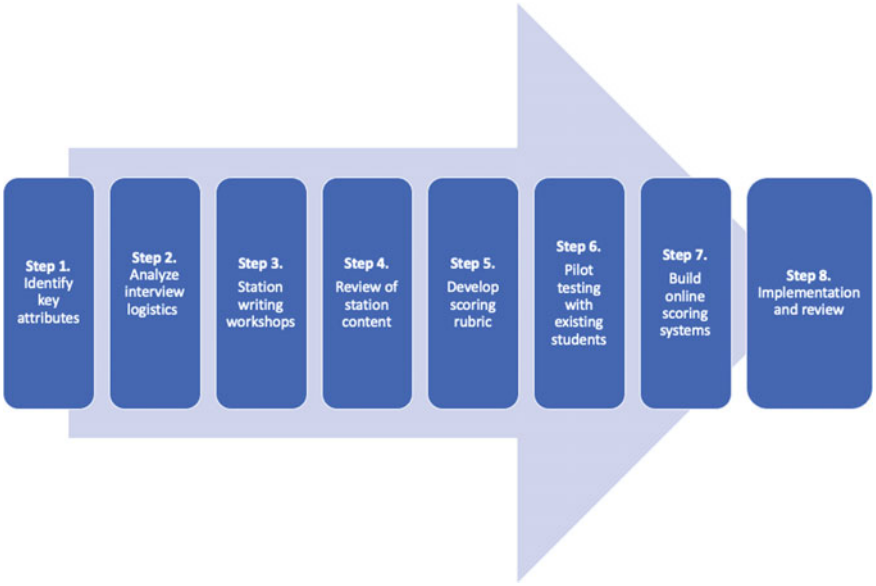


Fig. 8.1 Development steps for teacher selection MMIs

8.4 Developing MMIs for Teacher Selection in the UK

In this section, we present the development and results for a three-station (with four domains) MMI administered to almost 600 applicants over four interview dates. The development of the MMI took place during 2018–2019, with 'live' administration of the MMI during the 2019–2020 academic year.

Step 1: Identify key attributes. The MMI is built on key attributes that are identified and developed using a process similar to that used in SJT development (discussed in Chap. 7). An initial workshop was conducted with 28 teacher educators from the ITE program where the MMIs were to be implemented. A range of attributes were identified and debated based on the aims of the programme and the skills required for effective teaching, and after refinement, included *communication*, *respect for diversity and social justice*, *values and beliefs in relation to teaching*, *professional awareness*, *problem solving*, and *reflective thinking*.

Step 2. Consider interview logistics. The first workshop also included consideration of logistical factors such as the number of applicants and available resources (availability of interviewers and rooms). After modelling a range of administrative options (projected number of candidates; number of interviewers needed; available space), it was determined that only three stations could be included in the MMI, largely due to the pressure put on availability of interviewers. The 3-station model allowed up to 180 candidates to be assessed during each interview day. Although many MMI models developed in other fields have used a higher number of stations with longer durations (see e.g., Rees et al., 2016 for overview), Dodson et al. (2009)

found that shorter MMI station duration did not have a significant impact on reliability or interview outcomes. The final 3-station MMI model included a 20-min MMI circuit, which consisted of three 4-min stations, with a break of two minutes between each station for assessors to complete scoring and for candidates to move to the next station and prepare for the next station task.

Step 3. Station writing workshops. The researchers worked with a six-member expert panel to develop six potential MMI stations for piloting. The station tasks were trialed and reviewed by the panel that included education advisors and researchers. After review, three stations were selected and were deemed to appropriately encapsulate the target attributes. In line with Callwood et al.'s (2014) MMI model, communication was highlighted as an underlying construct relevant to all station tasks and thus was assessed at every station. Further revisions were made to the three stations based on feedback from the expert panel, before conducting a workshop to review the final content with teacher education staff at the university.

Step 4. Review of station content. The researchers introduced assessment staff to the MMI process, the content of the station tasks and the assessment criteria. Staff participants were divided into groups of four and assigned one of the three stations to review. Each group completed a role-play of the station task, with participants taking it in turns to enact the role of the interviewer or interviewee to trial the station content. Participants were then asked to review and discuss the MMI stations in terms of (a) the suitability of the language and terminology used, (b) the relevance of the interview task to the teaching profession and the programme, (c) the difficulty level, and (d) the appropriateness of the assessment criteria. Participants then supplied feedback through group discussion. The three station tasks are described in Table 8.1.

Step 5. Scoring guide development. A detailed scoring guide was developed by the researchers and program staff (see Fig. 8.2). A detailed station description included introduction to the station, a brief summary of the station task, a list of specific questions to be asked, additional prompt questions, and a rubric of positive and negative indicators.

Step 6. Pilot-testing with existing students. In Step 6 the three-station MMI was piloted with a small group of existing (i.e., incumbent) students in the program, with changes to the content and logistics based on the student feedback. In particular, instructions for the stations were improved with more clarity about the tasks.

Step 7. Building an online scoring platform. An online scoring platform was built using Qualtrics survey software. Interviewers entered applicant details, selected positive and negative indicators, and provided an overall attribute and communication score for each applicant. The positive and negative indicators were used to support the generation of a feedback report that was available to download after the interview day was completed. In addition, interviewers were able to write any additional comments, including 'red flag' comments that would be available for further consideration after the MMI was completed. After the interviewer completed the scoring, they submitted the score and were presented with a screen for the next applicant.

Step 8. Implementation and review. The MMI was piloted in 2019–2020 with 572 applicants (86% female); see Fig. 8.3 for a schematic of the process. After administration of the MMI was completed, the entire MMI process was reviewed,

Table 8.1 Attributes assessed at each station

Station	Activity	Description
Station 1 Values and beliefs	*Teacher profiles:* picture profile stimulus and discussion	The task requires candidates to show an understanding of a range of approaches to teaching, an appreciation of the value of diversity in teaching, and a commitment to exploring and establishing their own teacher identity
Station 2. Diversity and social justice	*School play*: Discussion with picture and text stimuli	The task assesses candidates' understanding of issues related to social justice, equality and inclusion in day-to-day teaching
Station 3. Professionalism	*Always, sometimes, never*: Card-sorting activity	The task evaluates candidates' awareness and understanding of a range of teacher professional values and behaviors
Communication	Assessed at all three stations	Defined as the ability to articulate well-reasoned arguments and to respond effectively to new information

with changes proposed to improve the process and station content for the following year.

Results. The total mean score on the MMI (scored out of 90) was 68.04, with an approximately normal distribution (skewness and kurtosis scores in the acceptable range). Reliability (internal consistency) for the overall MMI using intraclass correlations was 0.76, similar to MMI reliabilities reported in the literature. There were no significant sex differences. A factor analysis with the six station scores (i.e., three attributes and three communication scores) showed that each attribute/communication dyad formed a separate factor, explaining a total of 94.0% of the variance.

Applicant reactions. Applicant reactions to the three-station MMI were collected at the end of each session, with applicants generally positive about the experience. Most applicants agreed that *the MMI process is an appropriate selection method* (96% agreement), that *the MMI process is a fair method of selection* (95% agreement), and *The MMI stations were relevant to teaching* (98% agreement). Open-ended questions about the MMI resulted in 176 responses, with six key themes identified:

- Preferred MMIs over other interview methods, e.g., *I found the MMI enjoyable and a lot less daunting than sitting in front of a traditional interview panel*

Attribute Scoring Guide	
1-3 Poor	Candidateís responses are not directly relevant to the question, are one-dimensional, and do not address most of the key issues or alternative perspectives. Arguments are not always reasoned or accurate, and candidate displays very limited reflection.
4-6 Borderline	Candidateís responses show some understanding of key issues or alternative perspectives, and are mainly relevant to the question, but important factors are omitted. Arguments show limited reasoning, and the candidate engages in some reflection but this is cursory and predominantly descriptive.
7-9 Satisfactory	Candidate shows understanding of key issues and presents a somewhat reasoned argument. Responses are relevant to the question but not all of the important factors or issues have been considered. Some evidence of self-reflection but this is more descriptive than analytical.
10-11 Good	Candidateís responses show an appreciation of most of the key issues or alternative perspectives. Arguments are reasoned, and candidate shows evidence of engaging in reflection with some critical analysis.
12-13 Very good	Candidate shows an insightful appreciation of the key issues or alternative perspectives. Arguments are well reasoned, and candidate presents original answers that show critical self-reflection and analysis.
14-15 Outstanding	Candidate shows insightful appreciation of key issues. Arguments are comprehensive, showing creativity, original thought, and critical analysis. Candidate demonstrates strong evidence of critical self-reflection and analysis.

Fig. 8.2 Scoring guide for attributes

- Relaxed environment, e.g., *Feels a lot more relaxed than previous times I have applied and the gap in between interviews allow you to calm your nerves and reset yourself*
- Process was fair, e.g., *Lovely experience, very fair; everyone has the same opportunity*
- Engaging experience, e.g., *Actually very fun, gives you an idea of various scenarios*
- Provides a 'fresh start', e.g. *It was helpful to know that if you didn't feel like a section went well, you have another opportunity, making it a fair and thorough process*

Fig. 8.3 MMI design and stations

- Realistic content, e.g., *The content of the questions was really thought-provoking and relevant to teaching*

The key criticism from applicants reflected the desire for **more time needed** (e.g., *There could be a minute or two longer at each station; otherwise, an excellent method of interview*), and **more opportunities to talk freely about experience and to show personality** (e.g., *Doesn't take into account the personality or experiences of applicants; You don't get the opportunity to really 'sell yourself'*).

Interviewer reactions. Feedback from interviewers was generally positive, but less so than that from applicants. Most agreed that *The MMI process is an appropriate selection method* (78% agreement), *The MMI process is a fair method of selection* (76% agreement), and *The MMI process helps choose candidates suitable for teaching* (76% agreement). An open-ended question (*Do you have any comments about the MMI (e.g., format, usefulness, fairness?*) resulted in multiple comments about the timing of the process: *It was exhausting for the assessors; If a candidate struggled with a task (e.g., through nerves), their talk time was really limited; It's too quick to allow a valid judgment to be made; The format did not allow for professional dialogue to develop fully - felt rushed and impersonal.* Other interviewers were more positive about the efficiency of the process: *The interviews were well organized and efficient; It's a good way to assess a lot of candidates quickly.* The overall feedback from interviewers, while generally positive, was that the stations were too short, and that more time was needed between stations to reflect on each applicant's performance.

8.5 Key Suggestions for MMI Implementation

The idea behind MMIs is simple: replace traditional panel interviews with multiple, independent stations with systematic scoring in order to increase reliability. However, the implementation raises a series of logistical complexities, and possibly, reluctance from interviewers who may have less time with each candidate, and less autonomy in their interviewing sessions. Our own experiences in implementing MMIs showed that those applicants who experienced conventional interviews and MMIs generally preferred the MMI format. Building on their considerable experience as originators of MMIs, Eva et al. (2019) proposed a number of suggestions to maximize the quality of the MMI process, a number of which are relevant for developing MMIs for teacher selection.

1. There is no such thing as 'the MMI'. The number of key non-cognitive attributes that *could* be assessed is nearly infinite, and MMIs vary considerably in the number of stations, the attributes assessed, and the overall purpose of the method.
2. MMIs should reflect an integrated selection system which can assist in identifying the priorities of the curriculum and solidify the aims and identity of the training program. The targeted non-cognitive attributes are not chosen randomly but reflect the values of the program.
3. If reducing bias and increasing diversity of successful ITE candidates is a goal, then make every effort to construct a diverse team of MMI designers/writers and interviewers from the very beginning of the development process. Involving interviewers at an early stage increases the transparency of the process. Consider using a diverse group of incumbent students as interviewers in at least one station.
4. Think very carefully about the content of each station and how the stations interrelate in terms of targeted non-cognitive attributes, station content, and the tasks given to applicants. Consider the 'optics' of the station: what are the implicit messages sent? How much does the station require 'insider knowledge' that might disadvantage certain applicants? In addition, be sure to submit stations to rigorous review in a process that includes staff, the public, and incumbent students.
5. As much as possible, maximize the number of stations in order to increase the reliability of the process. When implementing teacher selection MMIs, we found that a five-station MMI was preferred by applicants and interviewers to a three-station MMI, and most medical school MMIs have upwards of seven or eight stations.
6. Train interviewers rigorously. Some interviewers will have a tendency to be more severe in marking ('hawks') and some more lenient ('doves'), and although MMIs lessen the effects of individual assessor bias, they do not eliminate the 'hawk and dove' effect. Pre-interview training and post-interview analysis can help reduce interviewer effects (e.g., Kiraly et al., 2020).
7. Review, revise, and improve. Whatever MMI process is implemented, there will be a need to review, revise, and improve the stations. Some stations will

work better than others, and some stations may result in unforeseen scoring patterns that disadvantage certain groups of applicants. Careful analysis of scoring data and feedback collected from applicants and interviewers will ensure that a process of continuous improvement can be implemented. For test security purposes, the same stations should not be used annually, with applicant sharing of station content and selection processes almost a certainty.

8.6 Next Steps and Chapter Summary

In this chapter we examined the research on the development and implementation of MMIs for teacher selection. MMIs provide an in-depth selection method that for large applicant intakes is best implemented as a second stage of the selection process, following a screening process, perhaps using SJTs and other selection methods. We examined teacher selection MMIs used in two locations—Finland and the UK—where MMIs show promising results, with advantages over 'traditional' interviews, but coupled with logistical complexities that need careful consideration. Research on MMIs in education is in its infancy, and further work is needed to establish how best to develop and implement the method. In Chap. 9, we look at how to 'put it all together' when designing a teacher selection program, with suggestions and recommendations for implementation.

References

Blacksmith, N., Willford, J. C., & Behrend, T. S. (2016). Technology in the employment interview: A meta-analysis and future research agenda. *Personnel Assessment and Decisions, 2*, 12–20.

Burke, K. (2004). Medical schools continue to favour white, middle class candidates. *British Medical Journal, 328*(7455), 1518.

Callwood, A., Cooke, D., & Allan, H. (2014). Developing and piloting the multiple mini-interview in pre-registration student midwife selection in a UK setting. *Nurse Education Today, 34*(12), 1450–1454.

Cleland, J., Chu, J., Lim, S., Low, J., Low-Beer, N., & Kwek, T. K. (2020). COVID 19: Designing and conducting an online mini-multiple interview (MMI) in a dynamic landscape. *Medical Teacher, 42*, 776–780.

Dana, J., Dawes, R., & Peterson, N. (2013). Belief in the unstructured interview: The persistence of an illusion. *Judgment and Decision Making, 8*, 512–520.

Davies, P., Connolly, M., Nelson, J., Hulme, M., Kirkman, J., & Greenway, C. (2016). 'Letting the right one in': Provider contexts for recruitment to initial teacher education in the United Kingdom. *Teaching and Teacher Education, 60*, 291–302.

Dodson, M., Crotty, B., Prideaux, D., Carne, R., Ward, A., & De Leeuw, E. (2009). The multiple mini-interview: How long is long enough? *Medical Education, 43*(2), 168–174.

Dowell, J., Lynch, B., Till, H., Kumwenda, B., & Husbands, A. (2014). The multiple mini-interview in the UK context: 3 years of experience at Dundee. *Medical Teacher, 34*, 297–304.

Eva, K. W., Macala, C., & Fleming, B. (2019). Twelve tips for constructing a multiple mini-interview. *Medical Teacher, 41*, 510–516.

Eva, K. W., Reiter, H., Trinh, K., Wasi, P., Rosenfeld, J., & Norman, G. R. (2009). Predictive validity of the multiple mini-interview for selecting medical trainees. *Medical Education, 43*, 767–775.

Eva, K. W., Rosenfeld, J., Reiter, H. I., & Norman, G. R. (2004). An admissions OSCE: The multiple mini-interview. *Medical Education, 38*, 314–326.

Kahneman, D. (2011). *Thinking fast and slow*. Farrar, Straus, and Giroux.

Kausel, E. E., Culbertson, S. S., & Madrid, H. P. (2016). Overconfidence in personnel selection: When and why unstructured interview information can hurt hiring decisions. *Organizational Behavior and Human Decision Processes, 137*, 27–44.

Kelly, M. E., Patterson, F., O'Flynn, S., Mulligan, J., & Murphy, A. W. (2018). A systematic review of stakeholder views of selection methods for medical schools admission. *BMC Medical Education, 18*, 139.

Kiraly, L., Dewey, E., & Brasel, K. (2020). Hawks and doves: Adjusting for bias in residency scoring interview scoring. *Journal of Surgical Education.* https://doi.org/10.1016/j.jsurg.2020.08.013

Lemay, J., Lockyer, J. M., Collin, V. T., & Brownell, A. K. (2007). Assessment of non-cognitive traits through the admissions multiple mini-interview. *Medical Education, 41*, 573–579.

McAndrew, R., & Ellis, J. (2011). An evaluation of the multiple mini-interview as a section tool for dental students. *British Dental Journal, 212*, 331–335.

Metsäpelto, R., Utriainen, J., Poikkeus, A., Muotka, J., Tolvanen, A., & Warinowski, A. (2020, November 23). Multiple Mini Interview as a Selection Tool for Initial Teacher Education Admissions. Retrieved from psyarxiv.com/ph5t8

O'Brien, A., Harvey, J., Shannon, M., Lewis, K., & Oswaldo, V. (2011). A comparison of multiple mini-interviews and structured interviews in a UK setting. *Medical Teacher, 33*, 397–402.

Patterson, F., Knight, A., Dowell, J., Nicholson, S., Cousans, F., & Cleland, J. (2016). How effective are selection methods in medical education? A systematic review. *Medical Education, 50*, 36–60.

Pau, A., Jeevaratnam, K., Chen, Y. S., Fall, A. A., Khoo, C., & Nadarajah, V. (2013). The multiple mini-interview (MMI) for student selection in health professions training - A systematic review. *Medical Teacher, 35*, 1027–1041.

Perkins, A., Burton, L., Dray, B., & Elcock, K. (2013). Evaluation of a multiple-mini-interview protocol used as a selection tool for entry into an undergraduate nursing programme. *Nurse Education Today, 33*, 465–469.

Rees, E. L., Hawarden, A. W., Dent, G., Hays, R., Bates, J., & Hassell, A. B. (2016). Evidence regarding the utility of multiple mini-interview (MMI) for selection to undergraduate health programs: A BEME systematic review: BEME Guide No. 37. *Medical Teacher*, *38*(5), 443–455.

Rosenfeld, J. M., Reiter, H. I., Trinh, K., & Eva, K. W. (2008). A cost efficiency comparison between the multiple mini-interview and traditional admissions interviews. *Advances in Health Sciences Education, 13*, 43–58.

Thomas, A., Young, M. E., Mazer, B., Lubarsky, S. E., & Razack, S. I. (2017). Reliability and validity of the multiple mini-interview (MMI) for admissions to an occupational therapy professional program. *British Journal of Occupational Therapy, 80*, 558–567.

Ungtrakul, T., Lamlertthon, W., Boonchoo, B., & Auewarakul, C. (2020). Virtual multiple mini-interview during the COVID-19 pandemic. *Medical Education:* https://doi.org/10.1111/medu.14207

Chapter 9
Designing and Implementing a Teacher Selection Program

Abstract Improving teacher selection can positively influence educational outcomes (Jacob, 2016), but implementing new selection methods can be a challenge in the face of time and resource constraints. Despite the potential benefits gained from shifting to more effective and efficient selection methods, making changes to long-standing practices can be disruptive and politically challenging for ITE programs and national education organizations unless there is a clear rationale and a well-thought-out plan in place to support these changes. Our primary goal for this chapter is to provide support and guidance for those considering the development of a comprehensive and evidence-supported selection program. In the research, theory, and practice-driven framework presented in this chapter, we detail the steps involved in testing and refining selection methods that are valid and effective in identifying the best possible teaching candidates, and that are defensible to key stakeholders (such as university administration and government bodies), applicants, and to the public.

9.1 Putting It All Together: Designing an Evidence-Supported Selection Program

Current teacher selection systems—whether for ITE or employment—are frequently developed by well-meaning education professionals who build on past practices, intuition, and folk knowledge about selection: Davies et al. (2016) reported how ITE interviewers relied on "a bit of a conversation" (p. 298) and their instincts to identify applicants who were going to be successful. Unfortunately, research suggests interviewers' instincts are hardly infallible, and in fact are not very useful in predicting eventual success (Dana et al., 2013). The pressures that shape a selection program vary depending on national or regional teaching standards, past practices ('*we've always done it this way*'), and the recruitment landscape, i.e., whether there is a shortage or oversupply of applicants. In education, scant attention has been paid to exploring research to build the various components of the selection program (Klassen & Kim, 2019). We propose that a comprehensive, evidence-driven teacher selection program deserves serious attention before, during, and after the actual selecting of

R. M. Klassen and L. E. Kim, *Teacher Selection: Evidence-Based Practices*,
https://doi.org/10.1007/978-3-030-76188-2_9

teacher candidates takes place, with routine evaluation of the purposes, goals, and outcomes of the whole process.

Selection is a predictive exercise. A selection program, whether for training or employment, is an exercise in prediction that is designed to (a) identify the attributes needed for success, (b) assess these attributes in a way that is valid and fair, and (c) confirm the relations between the assessed attributes and desired outcomes. In education settings, the 'filtering function' of selection is important at key steps in the journey towards becoming a teacher: first at selection into teacher education, then at the point of employment into a teaching position, and then, in some settings, at the point of being awarded a permanent position after a set period of successful teaching practice. Although these filtering activities happen at differing stages, the decisions made are similar in key ways: for entrance into training or into employment, the selectors will ask *What are the key attributes we want to assess? How should we assess them? Can we identify 'latent' attributes that will develop over time?* Assessment of a range of attributes using multiple measures at each point increases the likelihood of successfully predicting future success (Lievens & Sackett, 2017), with important questions needing to be answered about which attributes to target, how to measure them, and how to make selection decisions in a fashion that is evidence-supported, feasible for applicants and staff, and cost-effective. In the next sections, we will look at the steps involved in developing the best possible selection programs.

9.2 Choosing Selection Tools

One of the key steps in building a successful selection program is to choose selection tools that reflect the goals and values of the organization, are feasible to administer, and have evidence supporting their reliability, validity, and fairness. In a review of teacher selection methods, Klassen and Kim (2019) found that many organizations were uncritical about the quality of their selection programs, for four key reasons: (a) they collected little or no data on the effectiveness of their methods, (b) they relied on anecdotal information (e.g., *We chose a very good group of applicants this year!*), (c) they were unaware of selection research in education or in other professional fields, and (d) they lacked the resources and the institutional commitment to make changes to selection systems already in place. An effective teacher selection program is built on research-supported selection methods, includes routine evaluation of its methods, and reflects up-to-date practices from other professional fields.

Multiple predictors and order of administration. Selection practices benefit from multiple predictors because the target outcome—teacher effectiveness—is multidimensional and not easily predicted using a single predictor (Cook, 2009; Hattrup, 2012). In a review of the predictors of teacher effectiveness, Harris et al. (2010) noted that four predictors of effectiveness are commonly studied in research on teachers (and other professions): cognitive ability, personality, experience, and educational background. Rockoff and colleagues (Rockoff et al., 2011) examined

predictors of teacher effectiveness in cohorts of new elementary (i.e., primary) and middle school math teachers in New York City, and found that although few individual predictors significantly predicted student and teacher outcomes, two composite factors—cognitive attributes (subject-area knowledge, reasoning abilities) and non-cognitive attributes (people skills, motivation, commitment)—had a moderately large and statistically significant relation with student outcomes. An effective teacher selection program will include a range of predictors that evaluate applicants' cognitive attributes and non-cognitive attributes using the best possible research-informed methods.

The order of administration of selection methods depends on factors such as available resources (time, space, and money), ensuring that the 'applicant experience' reflects positively on the ITE program (applicants are 'interviewing ITE programs' as much as they are being interviewed), and the goals of the selection program (Tippins, 2012). The cost and time investment of conducting time-intensive interviews is higher than conducting large-group screening, and organizations may administer less expensive methods in a screening capacity to reduce applicant numbers before conducting more intensive (and expensive) methods. Organizations may also consider applicant reactions to tests when determining the order of administration. For teacher education, especially in the current UK situation where ITE programs compete for the best candidates, the selection process is as much about attracting applicants as about selecting them. It is important that the order of methods reflects some consideration of the 'attractiveness' of the methods. Methods with clear face validity and that reflect the target job (e.g., SJTs that are contextualized for teaching settings) may be more attractive to candidates than generic personality tests and so may be administered first. Feasibility of administration may also be an issue in determining order, with online testing possibly more efficient and feasible than bringing all applicants into a central location, especially if applicants are spread around the region or country.

A modular approach. A modular approach to selection processes, espoused by Lievens and Sackett (2017), involves breaking down the selection process into the basic underlying components—the 'building blocks'—of the process. Such an approach offers both scientific and practical utility, because it has the potential to provide a greater understanding of each of the elements used in the selection system (scientific utility), and it allows for flexibility in reviewing and redesigning existing selection procedures (practical utility). Lievens and Sackett propose seven predictor method factors:

1. Stimulus format: *What is the modality by which the test stimuli are presented to applicants?*
2. Contextualization: *To what extent is a detailed context provided to applicants?*
3. Standardization: *How standardized is the material presented to applicants?*
4. Response format: *What is the modality (e.g., face-to-face, remote, computer) used in the selection method?*
5. Evaluation of response consistency: *How standardized are applicant responses (open-ended to computer-scored)?*

6. Information source: *What kind of information is collected (e.g., observation of behaviour, self-reports, external reports)?*
7. Instructions for applicants: *How explicit are the directions for applicants for each of the selection tasks?*

An analysis of the components of the selection process is useful for an organization that is considering either 'tweaking' their selection process, or implementing wholesale changes, due to concerns about selection costs or the effectiveness of current selection methods. In order to implement a modular selection approach, the first step is to describe the selection challenges faced by the organization. The second step is to break down the current selection process into its parts, followed by a research- and theory-led analysis of which of the parts might be improved. Next, modifications of the selection process are implemented based on the previous analysis, and an evaluation of the selection process and its outcomes is undertaken. In parallel with the analysis of the selection process (i.e., the predictors), it may be worthwhile to review and analyze the criteria used to measure success, i.e., the outcomes that are routinely collected (or can be collected) by the organization in order to evaluate the effectiveness of the selection process.

Choosing measures for high-stakes settings. Many options are available when choosing tools for teacher selection, including well-standardized measures of emotional intelligence or personality that assess general characteristics important for all jobs, and not specifically designed for the teaching environment. One crucial consideration when considering selection measures is to understand the purposes for which a construct or specific measure has been used previously. Take, for example, the example of teacher self-efficacy, with items such as, *I am confident that I can manage disruptive students*. These measures can be strongly predictive of teaching outcomes in a low-stakes research context: Klassen and Tze (2014) found that teacher self-efficacy was a good predictor of evaluated teaching performance ($r = 0.28$). However, self-efficacy, like other constructs that may be valuable in low-stakes research contexts where participants are not heavily invested in the outcomes, may lose their effectiveness when used in a high-stakes selection setting due to social desirability effects.

The question of how context-rich selection measures should be provokes many questions about how personal characteristics are enacted across settings. For example, some SJTs are designed to be used for selection into many and varied professional contexts (e.g., CASPer; Dore et al., 2017). Developers of context-general selection measures argue that the personal characteristics of applicants are not situation-specific, and that personal characteristics should not be confounded with specific job awareness. However, as discussed in Chap. 7, it is the *context-rich* aspect of SJTs that lies at the heart of the methodology and understanding the contexts in which non-cognitive attributes are enacted may be crucial for selection purposes (Chen et al., 2016; Freudenstein et al., 2020). Important non-cognitive attributes are 'triggered' differently according to situational cues for individuals, and it is the interaction of situational awareness and knowledge (*What are schools like? What is appropriate behaviour in classroom settings?*) with individual traits and

attributes (empathy, integrity, adaptability) that is critical to assess during selection (Harris et al., 2016). However, using context-general measures provides for a measure of convenience, with off-the-shelf selection measures not tailored for educational contexts (e.g., personality inventories, emotional intelligence assessments, situational judgment tests) readily available from commercial test developers. The availability of SJTs contextualized for education settings is relatively rare, although research and application in the area is growing (e.g., Klassen et al., 2020).

9.3 Implementing a Selection Program

In Fig. 9.1 we present a five-stage selection program that displays the steps involved in implementation of a two-phase (screening and interview) selection program. Developing a selection program involves consideration of an organization's goals, the recruitment landscape, the kinds of measures that will be included, and decisions about how to implement the whole process. Implementation is a multi-faceted challenge because choices at each stage influence the other stages; for example, if a key

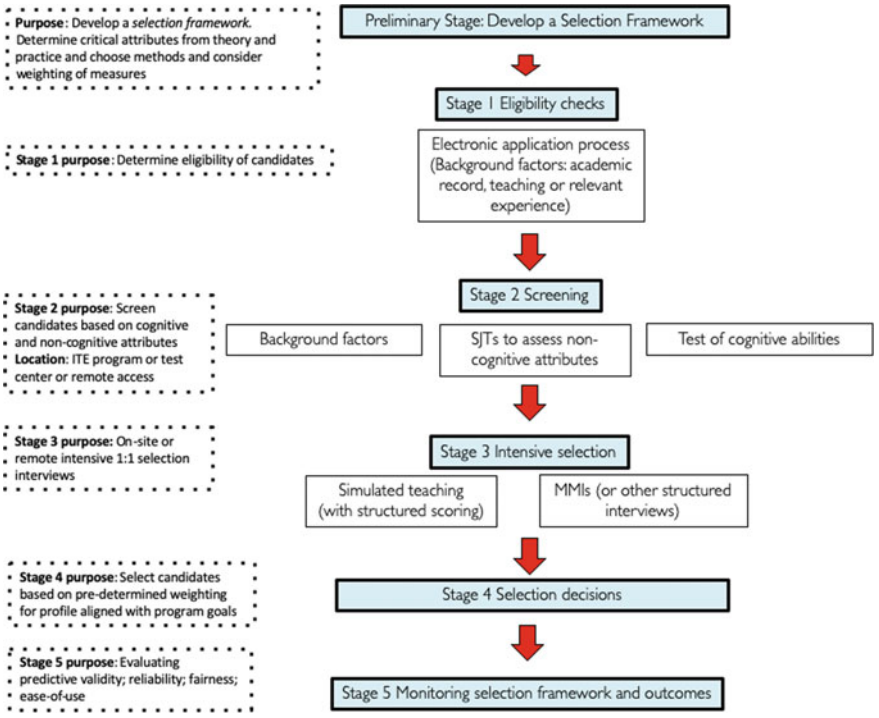


Fig. 9.1 Five-stage (with two selection phases) research and theory-driven selection program (adapted from Klassen & Kim, 2017)

goal is to reduce staff time spent in selection activities, a high threshold for screening in Stage 2 might be implemented to reduce the number of candidates interviewed by staff in Stage 3, allowing greater leeway for the kinds of assessments used for the intensive selection stage. The recruitment landscape and availability of resources also influence each stage. If the recruitment landscape is such that the number of available places is greater than the number of applicants (as is the case in some ITE settings), then a key goal of the selection program might be to build an understanding of applicant strengths and weaknesses for further development. If the applicant-to-places ratio is such that there are far more applicants than places, the key goal may be to develop an efficient (i.e., resource-feasible) and effective selection program that identifies the applicants with the highest potential for success. However, before the stages of the selection program are decided, careful attention needs to be paid to the *selection framework*, which provides guidance for each of the decisions made in implementation.

Preliminary Stage: Design a selection framework. One of the important steps in implementing a selection program is to plan in detail the foundation and goals held by the organization by designing a *selection framework* that reflects organizational goals and values, top priorities, available resources, and the recruitment landscape. Developing a clear, well thought-out selection framework increases the transparency of the process for applicants, for external stakeholders, for organization staff, and for those overseeing the selection program such as organization administrators. A carefully articulated selection framework also improves the chance of support and buy-in from staff involved in selection, especially if they can clearly see the links between target attributes, selection methods, and selection outcomes. For accrediting bodies (for example, Ofsted in the UK), a selection framework can provide evidence that the selection program is defensible in terms of a strong evidence base (i.e., in the types of methods chosen), is linked to national or program standards (through carefully chosen target attributes), and includes a robust monitoring system, where methods and outcomes are routinely evaluated.

Figure 9.2 presents an example selection framework with two models: a one-phase model where all selection data is combined for decision-making) and two-phase (screening and interview phases are compiled separately) models. This crucial first step sets the framework or 'roadmap' for all selection-related decisions and involves a selection or admissions team posing and answering a range of questions:

- What are the key attributes we want to target in our ITE program? What is the evidence for their relations to ITE success and/or teacher effectiveness?
- What selection methods should be considered? What is the research evidence for their effectiveness?
- Are there particular background factors that we value, or that show evidence of predictive validity (e.g., teaching experience, other valued experiences)?
- What are the main goals for each step in our selection program?

  Identifying the top performers (*screen in*)?
  Setting a minimum threshold (*screen out*)?
  Calculating a weighted score to be combined with other scores?

SELECTION FRAMEWORK					
	Screening tools			Interview	
	Academic records	SJT	Background factors	MMI or Structured interview	Simulated teaching
Attribute 1 Cognitive abilities	✓				
Attribute 2 Subject knowledge	✓				✓
Attribute 3 Empathy		✓		✓	
Attribute 4 Integrity		✓		✓	
Attribute 5 Adaptability		✓		✓	
Attribute 6 Social justice		✓		✓	
Attribute 7 Communication		✓		✓	✓
Example weighting for 1-phase selection	20%	20%	10%	40%	10%
Example weighting for 2-phase selection	50%	40%	10%	80%	20%

Fig. 9.2 Sample selection framework

Diversity in selected applicants?
Identifying attributes for future growth?

- How should we combine results from academic records, screening tools, and interview results? What are the weightings we should assign to each category?
- What are the constraints in our selection program in terms of resources (time, space, personnel, finances)? How can we organize our available resources (available space for interviews) and maximize resources (for example, by inviting community members to serve as interviewers)?
- What kind of selection process makes the most sense for us? *Simultaneous* (a one-phase process where all selection information is collected at the same time)? -or- *Sequential* (a two-phase (or more) process where one or more screening phases are implemented)?
- What will our selection process look like to applicants?

  What is the shape of the selection day?
  How will we welcome applicants, 'sell' our program to applicants, and provide convincing evidence that the selection process is reliable and fair, and that our program is excellent?
  How much time will be allocated to each task in the selection process?
  How will we ensure test security so that our selection methods are kept confidential to incoming applicants?

9.4 Stages of a Selection Program

Stage 1: Eligibility checks. The tasks in Stage 1 typically involve administrative activities that assess background factors such as academic record (e.g., meeting a minimum GPA requirement), and assessing eligibility for training program (e.g., appropriate degree, requisite teaching experience). It is expected that few applicants will be filtered out at this stage, but it is essential that eligibility criteria are established and that applicants who continue in the selection process are eligible for the program or job.

Key question: *Who is eligible?*

Stage 2: Screening. A screening stage may be implemented when the number of applicants is high in comparison to the number of spaces that are available. Ideally, screening will include a number of selection methods and will cover cognitive and non-cognitive attributes. Assessment of cognitive attributes (cognitive abilities, subject-area knowledge) is relatively easy, with numerous published tests available. Assessment of non-cognitive attributes is more challenging at the screening stage, especially if the number of applicants to be screened is large. Research suggests (e.g., Klassen et al., 2020; Lievens & Sackett, 2012) that one of the best screening options is an SJT that is designed to evaluate the 'soft skills' associated with teaching success. Other screening possibilities are personality tests, tests of emotional intelligence, letters of reference, or written personal statements, but the evidence supporting each of these assessment methods is not strong in an education context (e.g., Patterson et al., 2016). Including a 'file score' that assigns scores to agreed-upon valued background experiences (e.g., extensive relevant educational experiences) allows ITE programs to tailor the selection process to reflect their own values and goals.

Key question: *Who will be invited to interview?*

Stage 3: Intensive selection—the interview. In Stage 3, ITE program staff meet the applicants—typically face-to-face—but increasingly using on-line methods (especially when applicants have to travel long distances or when personal meetings are not feasible). Traditional interviews (i.e., low-to medium-structured format with a high degree of interviewer flexibility) are one of the most commonly used methods for interviewing in ITE (Davies et al., 2016; Hindman & Stronge, 2009), but are unreliable and poor predictors of professional practice (Patterson et al., 2016). Those designing the selection process for teacher education have an ethical responsibility to develop the best possible methods of selection; the use of multiple mini-interviews (MMIs) or other multiple, structured, independent interview approaches is one way to maximize the effectiveness of face-to-face interviews.

Key question: *Which applicants display acceptable levels of key non-cognitive attributes?*

Stage 4: Selection decisions. If a thoughtful selection framework is developed, the decisions about admissions will be relatively straightforward in Stage 4, with weightings of the selection process debated and determined during the development of the framework. However, it may be that important selection elements fall outside of the model developed in the selection framework, and a selection framework might

include additional weighting for 'background factors', the makeup of which will vary according to context. For example, the ITE program might decide to provide additional weighting (in our sample model, 10%) for populations that are under-represented in their usual intake or in the teacher population and provide a weighting for applicants with particular background characteristics. These background factors will vary widely by organization and country, but consideration of increasing the diversity of the intake should be discussed when developing the selection framework and applied at Stage 4 when making selection decisions.

Key question: *Who will we invite into our program?*

Stage 5: Monitoring selection framework and outcomes. A key feature in teacher selection programs is the monitoring stage where outcomes and methods are monitored, and changes considered and implemented. In our experience, teacher selection programs tend to be resistant to change, and once a selection process is in place, may not change in a substantial way for years or decades, even if new research supports the use of different methods. We have noted that selection programs are routinely evaluated for their ease of implementation, and possibly for applicant reactions to the selection activities. However, a regular evaluation of the selection program should also address utility of each of the selection components in (a) predicting success in the training program, with regards to retention, success in practice teaching, academic outcomes (i.e., on any coursework), and in (b) predicting success in beginning teaching, whether that be through external observation records, or, if available, students' classroom achievement using a value-added approach. Furthermore, the impact of the selection program on certain under-represented or protected groups should be assessed and considered after each selection cycle.

Key questions: *What are applicant reactions to our selection process? What do interviewers say about the process? How well does the interview process work for our organization? What is the reliability and validity of our selection methods? Are there adverse impacts on certain sub-groups?*

9.5 Streamlining the Selection Process

In some cases, a more condensed selection process may be desirable. A streamlined selection process might be implemented in cases where the number of applicants is not large compared to the number of available places, and when there is an organizational imperative that all applicants receive the same attention at selection, regardless of their likelihood of being selected. In this case, the preliminary stage of developing a selection framework is still essential to building a strong selection program. After eligibility checks at Stage 1, the selection tasks are administered simultaneously in Stage 2, with weightings of background factors, screening tools, and structured interviews calculated in line with the example in Fig. 9.2 (*Example weighting for 1-phase selection*). Figure 9.3 outlines the steps involved in a condensed selection program. When a single-phase selection process is used, it is important to ensure that the desired attributes to be assessed are identified in the selection framework, and

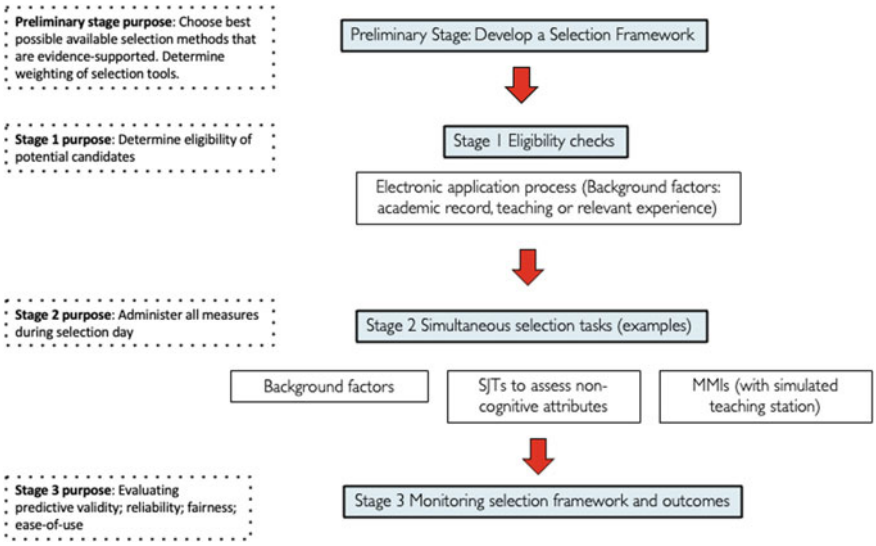


Fig. 9.3 Three-stage (with single selection phase) selection program using ready-made selection tools

that multiple reliable and valid measures are used to assess these attributes. Finally, the duration and mix of activities of the selection day need to be considered from the applicants' perspectives in order to ensure that applicants perceive the selection process as appropriately challenging, but not overwhelming.

9.6 Online Selection Methods

This chapter was written during the COVID-19 pandemic in 2020–2021, when ITE programs struggled with the challenge of bringing in candidates for face-to-face interviews, and furthermore, experienced higher numbers of applicants than in recent years due to the decrease in availability and stability of other employment options. ITE programs struggled to convert their usual face-to-face selection processes into an online or blended delivery system. Some methods are easier than others to implement virtually: online SJTs for screening are now common (e.g., Bardach et al., 2020), but research on other forms of online teacher selection interviewing approaches is rare. Once again, research in fields outside of education can provide some guidance in how to harness the ongoing rise in technology-related solutions related to online selection methods.

Online selection methods have the potential to make the selection program faster, easier to manage, and potentially, more accessible and less stressful for applicants. Woods et al. (2020) proposed five main types of online selection procedures: (a) online applications, where standardized online forms are used to provide personal

details and background information, (b) psychometric testing, including tests of skills, personality, and SJTs, (c) digital interviews using videoconference technologies either conducted in real time or recorded for subsequent scoring by interviewers, (d) gamified assessments, which use gaming elements in non-game contexts (e.g., using a game environment to assess decision-making), and (e) social media analysis, in which applicants' digital footprints in social media use are analyzed to infer job-relevant characteristics. Research on the reliability and validity of online selection methods is emerging, with most of the focus on psychometric tests, including SJTs where applicant reactions have shown a preference for online vs. offline methods of presentation (Woods et al., 2020). In spite of a growing research base, further work is needed on issues of test security, informed consent, adapting for disability, data security, accessibility for all applicants.

Videoconference and digital interviews. Although some selection methods, like SJTs, readily lend themselves to online administration, replacing face-to-face interviews with online substitutes is not so straightforward. Videoconference interviewing typically refers to a 'live' activity where interviewer and interviewee connect and interact through online audio and video modalities. Digital interviewing refers to the practice of asynchronous interviewing, where applicants record themselves while responding to a series of predetermined questions. In contrast to applicant reaction research on psychometric tests which show a preference for online administration, a recent meta-analysis that compared face-to-face with technology-mediated interviews showed significant application preference for face-to-face modality ($d = –0.41$; Blacksmith et al., 2016). When videoconference interviews were compared with digital interviews, applicants rated digital interviews as "creepier and less personal" (Langer et al., 2017, p. 371), and expressed concerns about providing private data to faceless organizations with no opportunity for the personal interactions usually found in interviews. However, applicants' negative perceptions of digital interviews did not result in lower perceptions of organizational attractiveness, suggesting that organizations might continue to explore implementation of digital interviewing as a way to gain a first impression of an applicant.

Multiple mini-interviews can be complex to implement in 'live' settings, with multiple stations, multiple interviewers, and logistical challenges involving applicants and spaces, but online delivery of MMIs has been implemented with success. Lake et al. (2020) compared the use of internet-delivered MMIs (on Skype) and traditional on-site MMIs for selection into a pharmacy training program and found no difference in performance and admission rates between the two formats. Similarly, Tiller et al. (2013) implemented internet-based MMIs for international applicants to medical and dental programmes in an Australian university, and found no differences between internet and face-to-face formats with higher reliability for the internet format. Applicants who participated in the internet MMI were generally positive (76% agreeing the online interview was a good way of selecting candidates; 16% unsure and 9% disagreeing). Most of the interviewers (78%) expressed satisfaction with the interview process and 71% found that the technology was satisfactory. As an additional bonus, conducting online MMIs rather than in person saved the university program approximately AUS $50,000 in costs related to travel and accommodation.

There are challenges associated with online selection activities, including test security, accessibility for applicants, and the cost of the development of online platforms to deliver selection activities, but there are a number of specific benefits as well. Indeed, during the COVID-19 pandemic of 2020–2021, online teacher selection methods became *de rigeur*, with reduced opportunities for conventional face-to-face selection approaches. Continuing work in the area of online teacher selection is clearly needed, and as technology advances, further opportunities for online selection will surely be developed.

9.7 Providing Feedback to Applicants

Two main questions can be posed about the nature of feedback given to candidates: *When* is information provided to applicants? and *How much* information is provided to applicants? In terms of timing, automated selection tests (e.g., online SJTs) can typically provide scores to applicants immediately after the test is completed; however, the scores provided may not be meaningful if other information is included in the selection process, or if cutoffs for admissions are not determined until after all applicants have completed the process.

The type of information provided to applicants varies considerably from organization to organization. For small scale selection into ITE, some organizations provide individualized feedback on request, but the task is difficult if the number of applicants is high. Organizations vary considerably on how feedback is provided, both regarding the timing and nature of the information that applicants receive after the selection process is completed. In large organizations, providing individualized feedback to applicants can be administratively challenging unless the process is automated. One possible strategy that can be adopted is to provide ITE applicants with general feedback based on the key attributes targeted in the selection framework. For example, if an MMI addresses multiple domains, a suite of 'Suggestions for further development' can be provided automatically or on request to unsuccessful applicants: *At this interview station you were asked to think about principles of social justice. As a teacher, your classes will be filled with students with different strengths and weaknesses. It is helpful to have open discussions with students regarding their viewpoints and preferences, and to provide opportunities for all students to experience success.* For selection programs that rely on electronic scoring (on laptops, mobile phones, or tablets), more tailored feedback of test performance can be automatically generated by programming the test platform to provide examples of positive indicators and 'areas for development' that the interviewer has noted on the scoring screen. Finally, even greater personalization can be offered if interviewers' constructive feedback is provided to applicants on request, although interviewers will want to be circumspect in the nature of the comments that they include on the scoring matrices.

9.8 Chapter Summary

In this chapter, we examined the steps in designing and implementing an effective teacher selection program. We have seen that multiple predictors are preferred over single predictors when making selection decisions, and that a modular approach to understanding and developing selection methods offers advantages to review and improve a selection program. Should measures of non-cognitive attributes be context-rich, or context-free? We argue that there are advantages and disadvantages to contextualization of the measures, but come down on the side of greater contextualization, since personal characteristics interact in important ways with particular settings. A five-stage, two-phase selection model was proposed, alongside a streamlined, one-phase selection process, anchored by a *selection framework* that provides a crucial foundation and plan for selection-related decisions. We also considered the possibilities of moving to online teacher selection methods and examined some relevant research that generally showed equivalency of face-to-face and online approaches. Overall, the key take-home message from the chapter is the critical importance of carefully considering the *how, what,* and *whys* of teacher selection methods, and how teacher selection programs are, or should be, dynamic and changing through careful planning and review. Teacher selection provides an effective and efficient way to improve the teacher workforce; paying attention to the details of selection programs can provide disproportionate benefits to organizations and education systems. In the next chapter we look to apply the research and methods from teacher selection to the stages before and after selection; that is, to teacher recruitment and development.

References

Bardach, L., Rushby, J. V., Kim, L. E., & Klassen, R. M. (2020). Using video- and text-based situational judgement tests for teacher selection: A quasi-experiment exploring the relations between test format, subgroup differences, and applicant reactions. *European Journal of Work and Organizational Psychology*. https://doi.org/10.1080/1359432X.2020.1736619

Blacksmith, N., Willford, J. C., & Behrend, T. S. (2016). Technology in the employment interview: A meta-analysis and future research agenda. *Personnel Assessment and Decisions, 2*, 12–20.

Chen, L., Fan, J., Zheng, L., & Hack, E. (2016). Clearly defined constructs and specific situations are the currency of SJTs. *Industrial and Organizational Psychology, 9*, 34–38.

Cook, M. (2009). *Personnel selection: Adding value through people* (5th ed.). Wiley-Blackwell.

Dana, J., Dawes, R., & Peterson, N. (2013). Belief in the unstructured interview: The persistence of an illusion. *Judgment and Decision Making, 8*, 512–520.

Davies, P., Connolly, M., Nelson, J., Hulme, M., Kirkman, J., & Greenway, C. (2016). 'Letting the right one in': Provider contexts for recruitment to initial teacher education in the United Kingdom. *Teaching and Teacher Education, 60*, 291–302.

Dore, K. L., Reiter, H. I., Kruger, S., & Norman, G. R. (2017). CASPer, an online pre-interview screen for personal/professional characteristics: Prediction of national licensure scores. *Advances in Health Sciences Education, 22*, 327–336.

Freudenstein, J.P., Schäpers, P., Roemer, L., Mussel, P., & Krumm, S. (2020). Is it all in the eye of the beholder? The importance of situation construal for situational judgment test performance. *Personnel Psychology,* DOI: https://doi.org/10.1111/peps.12385.

Harris, D. N., Rutledge, S. A., Ingle, W. K., & Thompson, C. C. (2010). Mix and match: What principals really look for when hiring teachers. *Education, 5*(2), 228–246.

Harris, A. M., Siedor, L. E., & Listyg, & Carter, N. T. . (2016). In defense of the situation: An interactionist explanation for performance on situational judgment tests. *Industrial and Organizational Psychology, 9*, 23–28.

Hattrup, K. (2012). Using composite predictors in personnel selection. In N. Schmitt (Ed.), *The Oxford handbook of personnel assessment and selection* (pp. 297–319). Oxford University Press.

Hindman, J., & Stronge, J. (2009). The $2 million decision: Teacher selection and principals' interviewing practice. *Educational Research Service Spectrum, 27*, 1–10. Retrieved from http://humanresourcesportfolio.wikispaces.com/

Jacob, B. A. (2016). The power of teacher selection to improve education. *Evidence Speaks Reports, 1 (12).*

Klassen, R. M., & Kim, L. E. (2017). Assessing critical attributes of prospective teachers: Implications for selection into initial teacher education programs. In D. W. Putwain & K. Smart (Eds.), *British Journal of Educational Psychology Monograph Series II: Psychological Aspects of Education* (pp. 5–22). Wiley.

Klassen, R. M., & Kim, L. E. (2019). Selecting teachers and prospective teachers: A meta-analysis. *Educational Research Review, 26*, 32–51.

Klassen, R. M., Kim, L. E., Rushby, J., & Bardach, L. (2020). Can we improve how we screen applicants for initial teacher education? *Teaching and Teacher Education, 87,* 102949.

Klassen, R. M., & Tze, V. M. C. (2014). Teachers' self-efficacy, personality, and teaching effectiveness: A meta-analysis. *Educational Research Review, 12*, 59–76.

Lake, J. D., Brown, T. E., Vu, K., & Tadrous, M. (2020). Comparison of admission rates to a non-traditional Doctor of Pharmacy programme using internet-based multiple-mini interviews versus on-site. *Pharmacy Education, 20*, 282–286.

Langer, M., König, C. J., & Krause, K. (2017). *International Journal of Selection and Assessment, 25*, 371–382.

Lievens, F., & Sackett, P. R. (2012). The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance. *Journal of Applied Psychology, 97*, 460–468.

Lievens, F., & Sackett, P. R. (2017). The effects of predictor method factors on selection outcomes: A modular approach to personnel selection procedures. *Journal of Applied Psychology, 102*, 43–66.

Patterson, F., Knight, A., Dowell, J., Nicholson, S., Cousans, F., & Cleland, J. (2016). How effective are selection methods in medical education? A systematic review. *Medical Education, 50*, 36–60.

Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2011). Can you recognize an effective teacher when you recruit one? *Education, 6*, 43–74.

Tiller, D., O'Mara, D., Rothnie, I., Dunn, S., Lee, L., & Roberts, C. (2013). Internet-based multiple mini-interviews for candidate selection for graduate entry programmes. *Medical Education, 47*, 801–810.

Tippins, N. T. (2012). Implementation issues in employee selection testing. *The Oxford Handbook of Personnel Assessment and Selection* (pp. 881–902). Oxford University Press.

Woods, S. A., Ahmed, S., Nikolaou, I., Costa, A. C., & Anderson, N. R. (2020). Personnel selection in the digital age: A review of validity and applicant reactions and future research challenges. *European Journal of Work and Organizational Psychology, 29*, 64–77.

Chapter 10
Beyond Selection: Applying Lessons from Teacher Selection to Recruiting and Developing Teacher Candidates

Abstract In this chapter of the book, we examine how lessons learned from teacher selection can be applied to recruiting better candidates to consider teaching as a profession, and to developing the candidates who are in training (see Fig. 10.1). We will consider the steps before and after the selection stage, and we will look at how the concepts and methods used in selection might productively be adapted to the purpose of recruiting and developing prospective teachers.

At the beginning of this book, we presented four pathways to improve the quality of the teacher workforce: attraction and recruitment, selection, development during initial teacher education, and development during professional practice. Throughout most of this book we have focused on selection as a pathway to educational improvement, with guidance about how educational systems might improve the way that they identify prospective teachers. However, there is more to improving the teacher workforce than making better selection decisions. How can the lessons learned from teacher selection be applied to what happens *before* and *after* candidates are selected, that is, during recruitment and development?

In this, the penultimate chapter of the book, we examine how these lessons can be applied to *recruiting* better candidates to consider teaching as a profession, and to *developing* the candidates who are in training (see Fig. 10.1). We will consider the steps before and after the selection stage, and in particular, we will look at how the concepts and methods used in selection might productively be adapted to the purpose of recruiting and developing prospective teachers.

10.1 Recruiting Prospective Teachers

Before teachers are selected for training or employment, they need to be attracted and recruited into pre-professional training. The term ‘attraction’ refers to raising potential applicants’ interest in the profession, especially for those who may not have seriously considered a teaching career. By ‘recruitment’, we mean the organizational practices that encompass influencing people to engage in the formal practices of

R. M. Klassen and L. E. Kim, *Teacher Selection: Evidence-Based Practices*,
https://doi.org/10.1007/978-3-030-76188-2_10

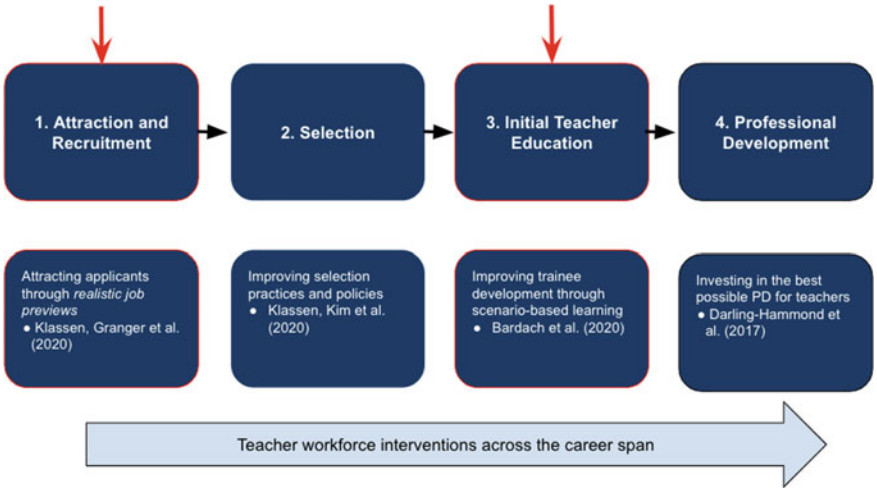


Fig. 10.1 Improving the teaching workforce through recruitment and development

applying for training or employment positions. For ease-of-use, we will use the terms interchangeably in this chapter.

A shortage of high-quality teachers hampers students' ability to learn, diminishes overall educational opportunities, and leads to a drain on economic resources (e.g., through provision of recruitment incentives) that could be better used elsewhere (Garcia & Weiss, 2019). See & Gorard (2019) examined the teacher recruitment landscape in England, and in particular, documented the gap between recruitment targets and the actual number of new entrants to the profession. They found that teacher vacancies tripled between 2011 and 2016, and proposed that teacher shortfalls could be addressed through more coherent policies accounting for supply and demand, revisions to the initial teacher education recruitment process, and a thorough evaluation of the cost and benefits of recruitment incentives. Not all countries face recruitment challenges: Finland, for example, has more applicants than places for ITE programmes, and the Covid-19 crisis and resulting economic uncertainty has increased the number of teaching applicants in the UK (Gibbons, 2020). However, in many developed and developing countries the quality of education systems has been threatened by an inability to recruit sufficient numbers of high-quality applicants.

A two-step process. Teacher recruitment strategies often follow a two-step process, first identifying areas of need (i.e., geographical areas or subject areas) where shortages exist, and second, offering a range of incentives (financial or guaranteed employment) to applicants, sometimes hinging on level of academic attainment (See & Gorard, 2019). However, there are problems associated with this approach: the evidence supporting teachers' general academic attainment and teaching effectiveness is not very strong (Bardach & Klassen, 2020), and recruiting prospective teachers based on cognitive factors alone may not be the best strategy to identify the most promising future teachers. In addition, paying out incentives to recruit people to join

training programs is expensive, and the long-term effectiveness of such recruitment strategies is not very well evidenced (e.g., Podolsky et al., 2019). A report on teacher recruitment produced by the OECD (Organisation for Economic Co-operation and Development) revealed that extrinsic financial incentives do not tend to attract high quality applicants, and in fact, may serve to attract applicants who are more interested in financial pay-off than their fit with the profession (OECD, 2018). There are a number of factors that attract people to a certain profession, including occupational status, work environment, sense of personal contributions, and the financial rewards associated with the profession (Podolsky et al., 2019). Looking at recruitment in other professions and learning lessons from teacher selection research may help inform how prospective teachers might be better recruited into training.

Personal and social utility. The underlying concepts in recruitment strategies usually fall under two categories. First are the appeals emphasizing the *personal utility* of pursuing a teaching career, i.e., by offering grants and bursaries for training, and higher salaries and improved working conditions for employment. The second appeal emphasizes the *social utility* of teaching; that is, through underlining how a teaching career can make a social contribution in terms of improving the lives of children and advancing social change. The 'simple view' of career attraction to teaching is shown in Fig. 10.2, with the two distinct factors contributing to the 'pull' of teaching. However, when making career-based decisions, individuals will weigh multiple factors—not just personal utility and social utility—but also their perceptions of 'fit' based on self-reflection supported by personal experiences and knowledge. This perception of fit influences the way a potential candidate evaluates the personal and social utility of a teaching career, and leads to a consideration of how well they will meet the perceived demands of the job.

Person-vocation fit. The notion of fit between people and their environments is one of the key theories in psychology and forms the foundation of person-vocation fit (PV fit), defined as the congruence between a person's interests and abilities, and the demands of particular jobs (e.g., Darrow & Behrend, 2017). Research shows a strong relation between PV fit and applicants' attitudes before applying for training or employment and also between PV fit and on-the-job behaviors and attitudes, such

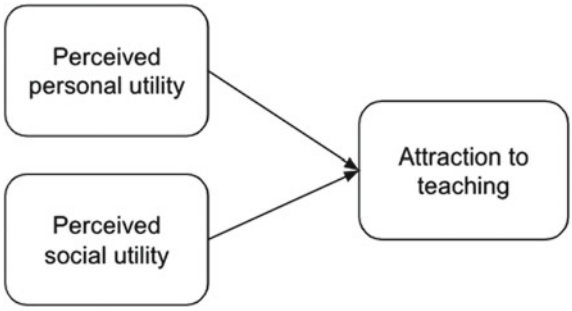


Fig. 10.2 Simple view of attraction to teaching (adapted from Klassen, Bardach, Rushby, & Durksen, 2021)

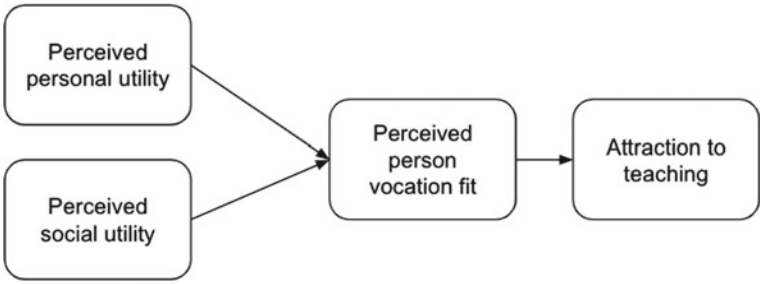


Fig. 10.3 Mediation model of teaching attraction (adapted from Klassen, Bardach, Rushby, & Durksen, 2021)

as job performance and work attitudes (e.g., Vogel & Feldman, 2009). A meta-analysis examining predictors of applicant attraction showed that perceived fit was the strongest predictor of applicant attraction across multiple stages of the recruitment process (Uggerslev et al., 2012). In education, De Cooman et al. (2009) found that selecting teachers who perceived a good fit between their own values and those of their schools were less likely to leave the profession. Findings from a meta-analysis conducted by Chapman et al. (2005) suggested that although characteristics of the job and the organization were key determinants of applicants' recruiting decisions, perceptions of fit were one of the strongest predictors of recruitment decisions. A mediation model of teacher recruitment includes personal and social utility, but these utility factors are mediated by perceptions of fit with the demands of teaching, as shown in Fig. 10.3.

How do education systems recruit applicants? Klassen, Bardach, Rushby, et al. (2021) examined the public-facing recruitment strategies and messages from two influential education organizations in England, Teach First (the largest provider of teacher training in England), and the Department for Education (DfE), which sets policies for education in England. Two sources of data were used for each organization, first, major policy documents outlining recruitment strategies were examined: for Teach First, *Britain at a crossroads* (Sundorph, 2018), and for the DfE, *Teacher recruitment and retention strategy* (Department for Education, 2019). Materials also included the advertising campaigns from each of the organizations, with Teach First's video and print campaigns (*After the outcome,* 2019) and the DfE's *Every lesson shapes a life.* Analysis of the source data included document analysis and an integrated deductive/inductive approach which coded meaning segments from text and video into primary and secondary coding units. For comparison purposes, the recruitment strategies and messages from the national health provider, the National Health Service (NHS), were examined (*Values Based Recruitment Framework,* NHS, 2016).

Findings of the analysis of recruitment messages from Teach First and the DfE (see Fig. 10.4) showed that recruitment messages from the education organizations included some reference to person-vocation fit in the advertising campaigns, but not in the policy documents, which focused on administrative changes and personal utility (DfE strategy document) and social utility and administrative changes (Teach

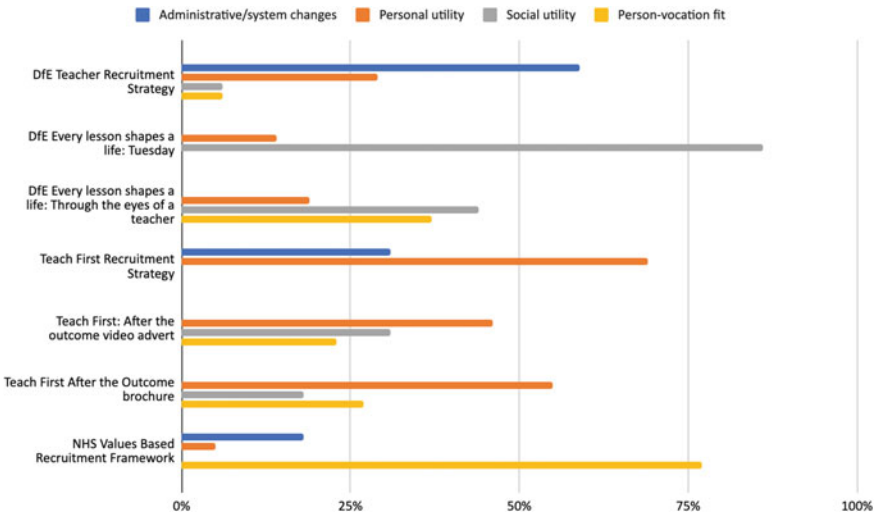


Fig. 10.4 Recruitment messages in strategy documents and advertising (adapted from Klassen, Bardach, Rushby, et al., 2021)

First document). The central focus of the Teach First video advertisements was on personal utility (e.g., "in terms of career progression… it's been a superb choice") and social utility (e.g., "let's give opportunities to kids who wouldn't normally have these opportunities"), with less attention paid to the fit between personal characteristics and teaching. The DfE advertising campaigns, such as *Every lesson shapes a life* (2020) strongly emphasized social utility, through references to shaping students' futures and making a social contribution. In contrast, messages from the NHS were highly skewed towards an emphasis on the importance of the match between the values espoused by the NHS and those values held by applicants: *there needs to be a good fit between an individual's personal values and those of the organisation* (p.52, NHS, 2016). The NHS emphasized that recruitment, selection, and long-term career development should be built on a core set of agreed values that are relevant across the career span.

10.2 Applying Lessons from Teacher Selection to Teacher Recruitment

Although teacher recruitment strategies that highlight social and personal utility can be effective, they can also attract applicants who may have an unrealistic view of teaching, resulting in high attrition rates due to poor fit (e.g., Baur et al., 2014). These traditional 'seduction' techniques can be complemented by recruitment methods that focus on the fit between applicants and a career in education.

Methods used for selection can be adapted for recruitment purposes. Situational judgment tests (SJTs)—typically used for selection purposes—can be repurposed to recruit potential applicants through an intervention called *realistic job previews* (RJPs), built on the tenets of person-vocation fit. RJPs are a recruitment method where potential applicants are presented with authentic workplace scenarios, similar to those presented for selection using SJTs, with the implicit question, *How well do you fit with this job?* Research on RJPs has been conducted for more than 50 years, with results showing that the intervention can result in better integration into a new field, leading to lower attrition and better workplace outcomes (Baur et al., 2014).

Including RJPs in the recruitment process provides three positive benefits: (a) they communicate an honest and believable portrayal of a job, leading to higher levels of applicant trust, (b) they reduce expectations so that new trainees are better prepared for inevitable workplace challenges, and (c) they lead to a self-selection process where applicants might decide not to pursue the profession if the perceived fit is poor. The combination of RJPs with person-vocation fit feedback helps to attract potential candidates who receive positive fit feedback and deters those who receive a message that they may not be well-suited to a particular vocation (Earnest et al., 2011).

RJPs to recruit STEM undergraduates into teaching. In many countries, there is an urgent need to improve the recruitment of STEM (Science, Technology, Engineering, and Mathematics) teachers: in the UK, the shortage of teachers in STEM-related fields has been acute, with shortfalls since at least 2011, and with growing shortages predicted (Foster, 2019). Recent research in the UK has investigated how RJPs might be used to attract undergraduate students in STEM subjects (science, technology, engineering, and mathematics) to consider a teaching career (Klassen, Granger, et al., 2021). In their study, Klassen and colleagues adapted materials from teacher selection tests to conduct a brief online RJP intervention that was delivered to STEM undergraduates, with post-test measures of self-efficacy for teaching, interest in teaching as a career, and match between personal attributes and the attributes required in a teaching career. Participants used their personal devices to view a series of brief classroom dilemmas in animated format. then to rate the appropriateness of three courses of action, and finally to provide a rationale for their response. Real-time feedback was provided on the alignment between their own ratings and those of expert teachers, and a 'fit' message based on their scoring profile was delivered to them (e.g., *Excellent fit – you think like a teacher! Your judgment matches closely with that of experienced teachers*).

Results from the study showed a statistically significant association between RJP scores and interest in a teaching career, but not between RJP scores and self-efficacy or attribute match. The findings held up after including the control variable of prior career intentions, suggesting that the RJP intervention increased interest in teaching as a career for those with and without prior interest in the career. Follow-up individual interviews found that the brief intervention was memorable and (for some participants) effective: *The activity showed that I had similar ideology as a teacher so made me think that maybe I would be suitable; it really helped me think about how teachers think*. Other participants were deterred from considering teaching as a

career: *I realize now that I'm just not patient enough (for teaching)*. Overall, the study showed that methods closely related to methods used for selection, i.e., SJTs, could provide a cost-effective and scalable approach to deliver a recruitment intervention. In addition, the methods adapted from teacher selection could also prove valuable for the development of prospective teachers, through an approach called *scenario-based learning*.

10.3 Developing Preservice Teachers Using a Scenario-Based Learning Approach

Although selection into a teacher education program represents a key starting point in a teacher's career, it is only the beginning of the story. Once an applicant is selected and starts a training program, developing the knowledge, skills, and attributes needed for successful practice becomes the goal. The simulated classroom situations that form the heart of SJTs used for selection (and also the realistic job previews used for recruitment) can be used for development purposes, in a method known as scenario-based learning or *SBL*.

Scenario-based learning. A new approach—SBL—to developing future teachers has recently been examined in the UK. The method of SBL is sometimes referred to as a *developmental* SJT, and it can provide a way to expose preservice teachers to a wide range of classroom contexts and situations in a low-risk learning environment, thus building teaching confidence (self-efficacy) and readiness to enter the classroom. The scenarios taken from SJT methodology provide the ideal vehicle to assess and develop classroom readiness, because the characteristics that make SJTs so valuable in personnel selection—their approximation of real-life scenarios, their adaptability to differing contexts, and their relative ease of administration—make them useful and adaptable to a wide range of classroom contexts. In spite of the apparent utility of SBL for developing training content, there is little empirical research exploring their use in teacher education, and a stronger research base is needed to understand the processes through which they influence learning. There is also a need for further exploration of which elements of scenario-based training (e.g., scenario content and length, scenario medium [video or text], feedback conditions [automated, targeted, supportive, etc.]) positively influence learning processes.

Scenario-based training has been used in some contexts outside of teacher education, for example, in training airline pilots (Fritzsche et al., 2006), and in health-related education, such as in medicine, nursing, and dentistry. Cox et al. (2017) compared the effects of SBL and lecture-based training on procedural and declarative knowledge on volunteers at a humanitarian disaster relief agency. Participants were given pre-tests assessing declarative knowledge about how to provide services after a natural disaster, followed by either lecture-based training or scenario-based training, and post-tests three weeks after the training. Participants in the scenario-based condition had higher procedural knowledge scores three weeks after the training than

those in the lecture-based condition ($d = 0.50$), with pre-test to post-test scores showing significant change only for participants in the scenario-based training condition. In a randomized control trial exploring the effects of scenario-based training on the communication competence of nurses, Hsu et al. (2015) found that nurses' mean communication scores and communication self-efficacy showed greater positive change in an SBL condition compared to a control group. In spite of studies showing the positive effects of SBLs for training, there has been little exploration of how scenario-based training might be useful in developing the competencies of prospective teachers.

Research using SBL for preservice teacher development. Researchers from the Teacher Selection Project recently developed an SBL intervention (Klassen, Bardach, Rushby, Maxwell, et al., 2021) based on teacher selection methods, and specifically, SJTs used to identify promising teaching candidates. They carried out a series of studies using SBL for developing preservice teachers, with an initial study showing that exposing participants to classroom scenarios in isolation was not sufficient to influence participants' self-efficacy and classroom readiness: the added components of self-reflection and feedback from experienced teachers provided a necessary boost to influence the outcome variables (Bardach et al., 2021).

A second study implemented a more intensive intervention: a four-session SBL 'module', with each session consisting of five video or text scenarios (Rushby & Klassen et al., 2021) focusing on the attributes of empathy and communication, emotion regulation, resilience and adaptability, and organization and planning (see Fig. 10.5 for a still shot of a video scenario). Participants were 463 preservice teachers enrolled in early years and primary ITE programs in Australia and the UK and who were preparing for entering the classroom as preservice teachers for a major practicum. Participants completed the four SBL sessions on the device of their choice (i.e., mobile phone, tablet, laptop) over a four-week period. For each scenario, participants:

1. Read or viewed the scenario
2. Rated the appropriateness of three possible response options (from *inappropriate* to *appropriate*
3. Provided a brief reflection for their responses
4. Viewed how experienced teachers rated the scenario responses, and
5. Received tailored feedback on their own responses

At the start of the intervention, and at the end of each session, participants completed brief measures of teaching self-efficacy (e.g., *I am confident that I can manage student behavior),* and emotional and cognitive classroom readiness (e.g., *I feel enthusiastic about teaching,* and, *I think I have the competencies needed to be a good teacher*).

Results from the four-week intervention showed statistically significant increases in SBL performance ($p < 0.05$), with post-hoc comparisons showing a significant difference between session 1 and the subsequent sessions. There was also a significant increase in mean self-efficacy (with a large effect size), emotional classroom readiness (medium effect size), and cognitive classroom readiness (large effect size).

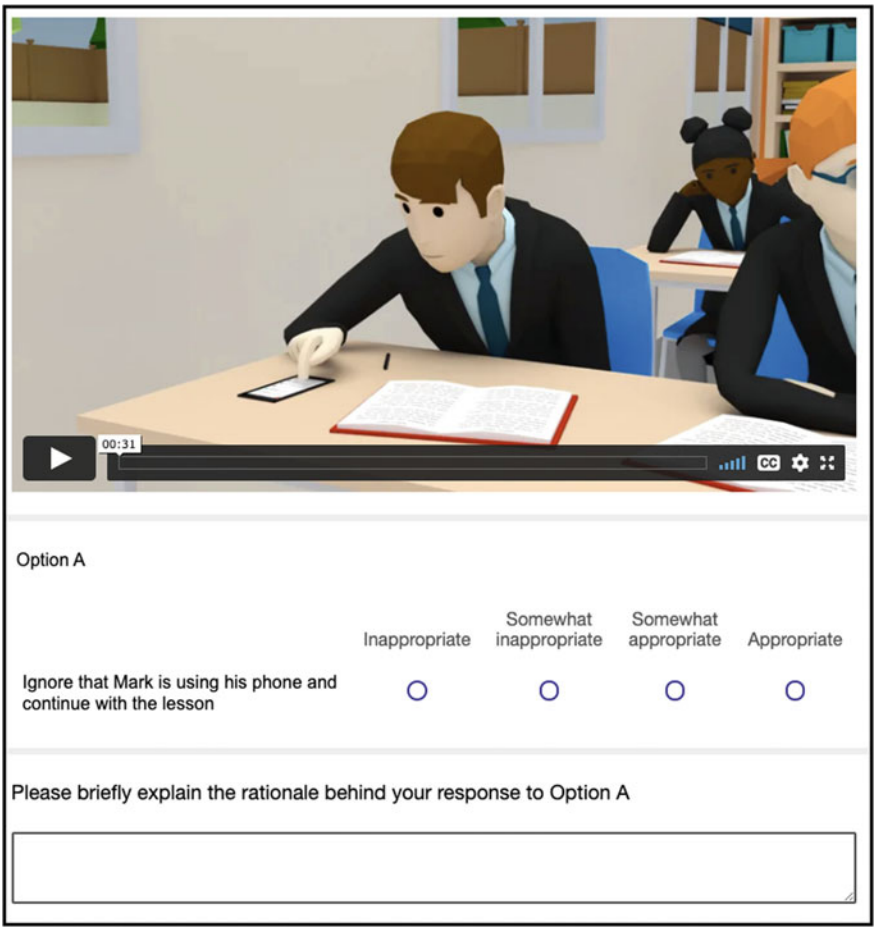


Fig. 10.5 Example item with automated feedback from 2020 SBLP pilot test

Reactions from participants to a series of open-ended questions revealed strong support for the effectiveness of the intervention, with participants highlighting that the authenticity of the scenarios helped them to feel more confident and prepared for their upcoming teaching placements. Participants also noted that the real-time feedback allowed them to make comparisons of their teaching decisions with the decision-making of expert teachers, and that the reflection opportunity encouraged critical thinking about these challenging situations: *It gave me the opportunity to think practically about situations and issues that I have not yet faced… Thinking more in-depth about real situations reduced my overall stress when it comes to thinking about teaching… It makes me feel more confident in my abilities and it helped me to consider how I might want to tackle such problems in the future.* The majority of participants (97.5%) found that the SBL intervention helped them feel

prepared to teach, and 91.3% reported feeling more confident about their teaching. This effective intervention, built on the research and methods from selection research, showed a high degree of effectiveness in boosting preservice teachers' confidence and readiness to enter the classroom as preservice teachers.

Scenario-based learning interventions are built using the same 'engine' as SJTs used for selection: authentic classroom situations that provide prospective teachers with a taste of classroom practice. Including SBL interventions in a teacher education program offers one way to provide novices with the chance to reflect on challenging classroom demands, and to receive tailored feedback from experts. Further work is needed to develop this intervention, but the findings from recent studies are promising across levels (primary and secondary) and a range of ITE programs internationally. For education organizations, the integration of recruitment, selection, and development activities can be built on a coherent framework of shared attributes that underpin teacher effectiveness.

10.4 Chapter Summary

Methods used in teacher selection, and especially the simulated classroom scenarios used as the basis for SJTs, can be used to inform teacher recruitment and development. These authentic slices of classroom practice provide a taste of teaching for potential applicants using a realistic job preview method and can also provide development opportunities for trainees in scenario-based learning applications. Recent research shows that using 'real-world' scenarios with targeted feedback provides a powerful message for recruitment purposes, where a person-vocation fit message can recruit applicants who may not have considered teaching as a career, but also for development purposes, where trainees can experience the classroom, and receive guidance from more experienced professional colleagues. The lessons learned from teacher selection can, indeed, be applied to the stages before and after the selection process and can transform the recruitment and development of high-quality prospective teachers. In the final chapter we propose some of the trends and likely future developments in teacher selection.

References

Bardach, L., & Klassen, R. M. (2020). Smart teachers, successful students? *A Systematic Review of the Literature on Teachers' Cognitive Abilities and Teacher Effectiveness. Educational Research Review*. https://doi.org/doi.org/10.1016/j.edurev.2020.100312

Bardach, L., Klassen, R. M., Durksen, T. L., Rushby, J. V., Bostwick, K. C., & Sheridan, L. (2021). The power of feedback and reflection: Testing an online scenario-based learning intervention for student teachers. *Computers and Education, 169,* 104194.

Baur, J. E., Buckley, M. R., Bagdasarov, Z., & Dharmasiri, A. S. (2014). A historical approach to realistic job previews: An exploration into their origins, evolution, and recommendations for the future. *Journal of Management History, 20*, 200–223.

Chapman, D. S., Uggerslev, K. L., Carroll, S. A., Piasentin, K. A., & Jones, D. A. (2005). Applicant attraction to organizations and job choice: A meta-analytic review of the correlates of recruiting outcomes. *Journal of Applied Psychology, 90*, 928–944.

Cox, C. B., Barron, L. G., Davis, W., & de la Garza, B. (2017). Using situational judgment tests (SJTs) in training: Development and evaluation of a structured, low-fidelity scenario-based training method. *Personnel Review, 46*, 36–45.

Darrow, J. B., & Behrend, T. S. (2017). Person-environment fit is a formative construct. *Journal of Vocational Behavior, 103*, 117–131.

De Cooman, R., De Gieter, S., Pepermans, R., Hermans, S., Du Bois, C., Caers, R., & Jegers, M. (2009). Person–organization fit: Testing socialization and attraction–selection–attrition hypotheses. *Journal of Vocational Behavior, 74*(1), 102–107.

Department for Education (2019). *Teacher recruitment and retention strategy.* Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/786856/DFE_Teacher_Retention_Strategy_Report.pdf

Earnest, D. R., Allen, D. G., & Landis, R. S. (2011). Mechanisms linking realistic job previews with turnover: A meta-analytic path analysis. *Personnel Psychology, 64*, 865–897.

Foster, D. (2019). *Teacher recruitment and retention in England.* House of Commons Briefing Paper 7222.

Fritzsche, B. A., Stagl, K. C., Salas, E., & Burke, C. S. (2006). Enhancing the design, delivery, and evaluation of scenario-based training: Can situational judgment tests contribute? In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement and application* (pp. 301–318). Lawrence Erlbaum.

Garcia, E. & Weiss, E. (2019). The teacher shortage is real, large and growing, and worse than we thought. *Economic Policy Institute.* Retrieved from https://www.epi.org/files/pdf/163651.pdf

Gibbons, A. (2020, September). Lockdown teacher training applications leap by a third. *Times Educational Supplement.* Retrieved from: https://www.tes.com/news/lockdown-teacher-training-applications-leap-third

Hanushek, E. A. (2014). Boosting teacher effectiveness. *What lies ahead for America's children and their schools*, 23–35.

Hsu, L., Chang, W., & Hsieh, S. (2015). The effects of scenario-based simulation course training on nurses' communication competence and self-efficacy: A randomized control trial. *Journal of Professional Nursing, 31*, 37–49.

Klassen, R., Bardach, L., Rushby, J., & Durksen, T. L. (2021). Examining Teacher Recruitment Strategies in England. *Journal of Education for Teaching.* https://doi.org/10.1080/02607476.2021.1876501

Klassen, R., Bardach, L., Rushby, J. V., Maxwell, L., Durksen, T. L., & Sheridan, L. (2021). The development and testing of an online scenario-based learning activity to prepare preservice teachers for teaching placements. *Teaching and Teacher Education.* https://doi.org/10.31234/osf.io/sz2xy.

Klassen, R. M., Granger, H., & Bardach, L. (2021). Attracting prospective STEM teachers using realistic job previews: A mixed methods study. *European Journal of Teacher Education.* https://doi.org/10.1080/02619768.2021.1931110

NHS (2016). *Values Based Recruitment Framework.* Health Education England. Retrieved from https://www.hee.nhs.uk/sites/default/files/documents/VBR_Framework%20March%202016.pdf

OECD. (2018). *Effective teacher policies: Insights from PISA*. OECD Publishing. https://doi.org/10.1787/9789264301603-en

Podolsky, A., Kini, T., Darling-Hammond, L., & Bishop, J. (2019). Strategies for attracting and retaining educators: What does the evidence say? *Education Policy Analysis Archives, 27*, 1–43.

Rushby, J. V., & Klassen, R. M. (2021, August). *A simulated teaching intervention boosts preservice teacher self-efficacy and classroom readiness*. Presented at EARLI 2021, Gothenburg, Sweden. https://docs.google.com/presentation/d/12sGtKv5z1_3ewRy9X4prnqWAnSa8hKguUeE1WWonTmQ/edit#slide=id.p1

See, B. H., & Gorard, S. (2019). Why don't we have enough teachers? A reconsideration of the available evidence. *Research Papers in Education,* DOI: https://doi.org/10.1080/02671522.2019.1568535

Sundorph, E. (2018). *Britain at a crossroads: what will it take to provide the teachers our children need?* Teach First. Retrieved from https://www.teachfirst.org.uk/sites/default/files/2019-08/britain_at_a_crossroads_2.pdf

Uggerslev, K. L., Fassina, N. E., & Kraichy, D. (2012). Recruiting through the stages: A meta-analytic test of predictors of applicant attraction at different stages of the recruiting process. *Personnel Psychology, 65*, 597–660.

Vogel, R. M., & Feldman, D. C. (2009). Integrating the levels of person-environment fit: The roles of vocational fit and group fit. *Journal of Vocational Behavior, 75*, 68–81.

Chapter 11
The Future of Teacher Selection

Abstract Throughout this book we have argued that implementing research-based teacher selection methods is a 'quick win' to improve education systems, but that selection methods for teacher education and employment have not kept pace with those used in other disciplines. In this final chapter, we share a few concluding thoughts about our prediction of the future direction of teacher selection research and practice. We suggest six themes that will characterize teacher selection work: (a) developing attribute-based approaches to teacher selection, (b) addressing the diversity-validity dilemma, (c) building modular approaches to selection, (d) testing longitudinal validity of selection methods, (e) designing new methods based on technological advances, and (f) improving the recruitment and selection of school leaders.

Throughout this book we have argued that implementing research-based teacher selection methods is a 'quick win' to improve education systems, but that selection methods for teacher education and employment have not kept pace with those used in other disciplines. We began with a look at calls for the reform of teacher selection methods from the 1920s, with education 'influencers' such as John Dewey wondering whether 'the sciences' could help solve the problem of differential teacher quality. Now, a century after multiple calls for building the teacher workforce through better selection, we have shown in this book that (a) effective teacher selection is a crucial building block for strong education systems, and (b) teacher selection can be improved through implementing proven methods that have a sound theoretical foundation, a robust research base, and that have a track record of successful application. In this final chapter, we share a few concluding thoughts about our prediction of the future direction of teacher selection research and practice.

Changing the status quo. Although it seems clear that developing and implementing better selection methods is an effective and efficient way to improve educational outcomes, we occasionally see a certain resistance to changing selection programs, with a reluctance to engage with research on the effectiveness of current and potential selection approaches. While it is easier to maintain the status quo with selection practices, making changes to recruitment and selection structures can pay off with better candidates in teacher education programs, a higher quality teacher

R. M. Klassen and L. E. Kim, *Teacher Selection: Evidence-Based Practices*,
https://doi.org/10.1007/978-3-030-76188-2_11

workforce, and the enhanced reputation of the profession as a whole. Structural changes to selection programs occur in three phases—initiation, implementation, and institutionalization—and it is entirely predictable that resistance to proposed changes will occur during the initiation and implementation phases. Murphy (2016) proposed that during the initiation phase, managing resistance to change involves presenting a clear message about the urgent reasons and rationale for proposed changes so that a 'sense of purpose' is fostered among those implementing the changes. During the implementation phase, a detailed support plan can help address the needs of those charged with implementing new systems, with two-way communication essential to help manage the inevitable challenges. Implementing changes in large organizations is always challenging, but the reasons for improvements to teacher selection are persuasive. Improvements to teacher selection not only have the potential to strengthen the teacher workforce but can also build economic and social health at the country level with relatively modest investment (Hanushek, 2014).

The future of teacher selection. In this book we have discussed a number of research-based approaches to improving teacher selection practices, but we expect that further research will bring increased opportunities to develop the ways we identify the best possible teachers. We suggest that these six themes will characterize selection work in the next decade.

1. *Developing attribute-based approaches to teacher selection.* We have seen a significant shift towards attribute-based (or *values-based*) approaches to selection in professional fields outside of education (e.g., in the NHS and College of Policing in the UK). We suggest that educational organizations will follow. The key attributes, once identified and endorsed, can form the bedrock of recruitment, selection, and development strategies. Building a selection framework (Chap. 9) encourages thoughtful consideration of key attributes and the methods used to assess them. Of particular interest will be cross-cultural work that explores the universality and the cultural specificity of the key attributes deemed essential for successful teaching practice. For education organizations and systems, an integrated recruitment → selection → development framework benefits greatly from an attribute-based approach (see Chap. 10).
2. *Addressing the diversity-validity dilemma.* Some countries experience a lack of diversity (in gender, ethnic, social class, sexuality) in the teaching workforce (Hodge & Marsh, 2015) and some selection methods show 'adverse impact' or bias towards certain groups (see Chap. 4). Little attention has been paid to reducing group differences in teacher selection, but most selection methods (interviews, cognitive measures, SJTs, MMIs) show some form of group differences. Acknowledging (and measuring) these differences can help make selection fairer with more representative pools of successful applicants. For example, video SJTs may reduce gender differences compared to text SJTs, but may not have an effect on ethnic group differences (Bardach et al., 2021). Developing the best possible selection methods includes a focus not just on 'raw' predictive validity, but on the nuances of group differences in performance using particular selection tools.

3. *Building modular approaches to selection.* A program of selection will benefit from breaking down the selection process into the basic underlying components, as suggested by Lievens and Sackett (2017), and as discussed in Chap. 9. Attention to the building blocks of selection (e.g., stimulus format, contextualization, response formats, variety of information sources) allows for better decision-making when considering changes to a selection program. Of key importance is developing a *selection framework* that incorporates multiple predictor variables and multiple outcome variables for validation.
4. *Testing longitudinal validity of selection methods.* In some fields, especially medical education, testing the validity of selection methods is routine. In teacher education, testing the validity of selection methods is rare, and data collected at the point of selection is rarely examined in terms of important future outcomes (Klassen & Kim, 2019). In order to maximize the effectiveness of selection methods, further research is needed to measure important teacher outcomes (e.g., attrition rates, teaching observation ratings, student value-added achievement) from the point of selection through training and well into professional practice.
5. *Designing new methods based on technological advances.* The restrictions associated with the COVID-19 pandemic resulted in ITE programs implementing remote selection processes in many settings. Technological advances in selection will likely go far beyond online delivery. One likely area of advance is *gamified assessments* (serious games) that incorporate gaming elements in non-game contexts (e.g., Woods et al., 2020). For example, SJTs can be 'gamified' by placing candidates in a virtual school classroom using virtual reality technologies, and 'branched' SJTs can provide a more interactive and tailored selection process (e.g., Reddock et al., 2020). One such immersive virtual teaching environment is SimLab, developed in the US (https://kognito.com) and recently piloted for teacher education at Murdoch University in Australia (Ledger & Fischetti, 2020), with results showing increased self-efficacy of preservice teachers. Recent advances in virtual reality and simulated teaching environments will spur on new ways to select prospective teachers (e.g., Ke & Xu, 2020).
6. *Improving the recruitment and selection of educational leaders.* There is a shortage of high-quality school leaders world-wide and current selection processes can be arbitrary and opaque (Yang et al., 2021). Lessons learned from teacher recruitment and selection (e.g., building from a foundation of key attributes, using evidence-informed methods, measuring the validity of methods) can improve the approaches used to attract and select prospective school leaders. Further research focused on building and testing new methods to identify the best possible educational leaders has the potential to improve educational systems worldwide.

Final words. Selecting the next generation of teachers presents a genuine opportunity to improve education systems and positively influence social and educational outcomes for children and young people. The overarching goal of a well-functioning teacher selection program is to use the best possible methods to recruit and select

outstanding prospective teachers. In this book we have argued that selecting future teachers is an important challenge that has been too-long neglected, and that modest efforts focused on improving selection methods will pay off with a stronger teacher workforce, a higher quality education system, and a more productive and prosperous society.

References

Bardach, L., Rushby, J. V., Kim, L. E., & Klassen, R. M. (2021). Using video-based situational judgment tests for teacher selection: A quasi-experiment exploring the relations between test format, subgroup differences, and applicant reactions. *European Journal of Work and Organizational Psychology, 30*, 251–264.

Hodge, K., & Marsh, S. (2015, November 19). Teaching profession fails to reflect multi-cultural student population. *The Guardian.* Retrieved from: https://www.theguardian.com/teacher-network

Ke, F., & Xu, X. (2020). Virtual reality simulation-based learning of teaching with alternative perspectives taking. *British Journal of Educational Technology, 51*, 2544–2557.

Klassen, R. M., & Kim, L. E. (2019). Selecting teachers and prospective teachers: A meta-analysis. *Educational Research Review, 26*, 32–51.

Ledger, S., & Fischetti, J. (2020). Micro-teaching 2.0: Technology as the classroom. *Australasian Journal of Educational Technology, 36*, 37–54.

Lievens, F., & Sackett, P. R. (2017). The effects of predictor method factors on selection outcomes: A modular approach to personnel selection procedures. *Journal of Applied Psychology, 102*, 43–66.

Murphy, M. (2016). The tug of war between change and resistance. *Educational Leadership, 73*, 66–70.

Reddock, C. M., Auer, E. M., & Landers, R. N. (2020). A theory of branched situational judgment tests and their applicant reactions. *Journal of Managerial Psychology, 35*(4), 255–270.

Woods, S. A., Ahmed, S., Nikolaou, I., Costa, A. C., & Anderson, N. R. (2020). Personnel selection in the digital age: A review of validity and applicant reactions, and future research challenges. *European Journal of Work and Organizational Psychology, 29*, 64–77.

Yang, M., Lee, S. W., & Goff, P. T. (2021). Labor dynamics of school principals in rural contexts. *AERA Open, 7*, 1–18.

Index

R. M. Klassen and L. E. Kim, *Teacher Selection: Evidence-Based Practices*,
https://doi.org/10.1007/978-3-030-76188-2