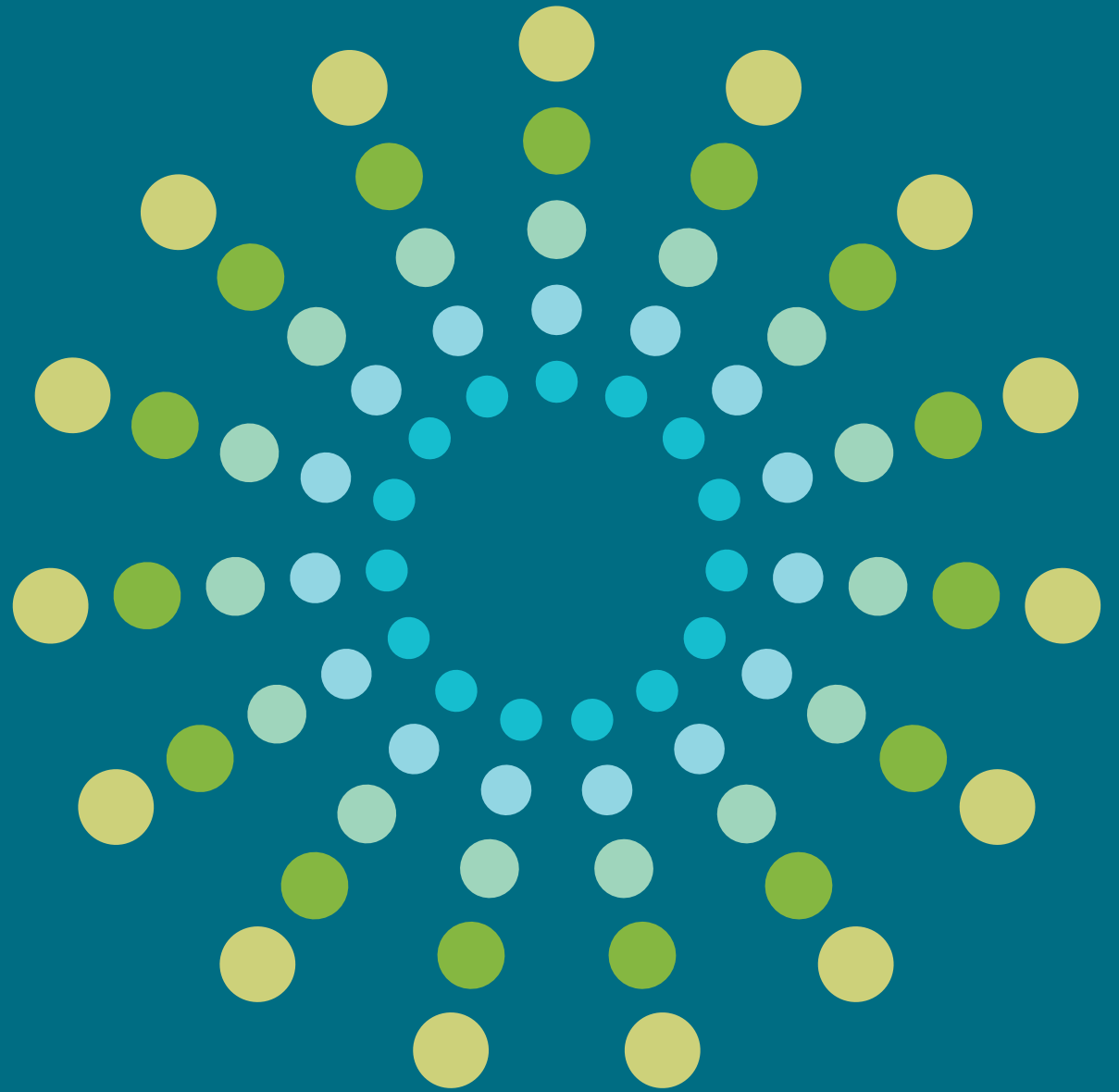**Data Insights Report**

Dr Richard Churches and Jan Lawrance

How to assess the potential to teach

New evidence from a STEM
teacher assessment centre
model in England

Education Development Trust

Future Teaching SCHOLARS

Department for Education

About the authors

Dr Richard Churches is Lead Advisor for Education Reform and Evidence-Based Practice at Education Development Trust and Programme Director for Future Teaching Scholars. He has an MA in Education and a quantitative PhD

in leadership, organisational behaviour and psychology. Richard is also British Psychological Society Level A and B qualified in occupational testing[1] and has completed the OPP Assessment Centre Design Programme.

Jan Lawrance is a Research Analyst at Education Development Trust. She has a BSc and MSc in Psychology and works across a wide range of areas conducting data analyses.
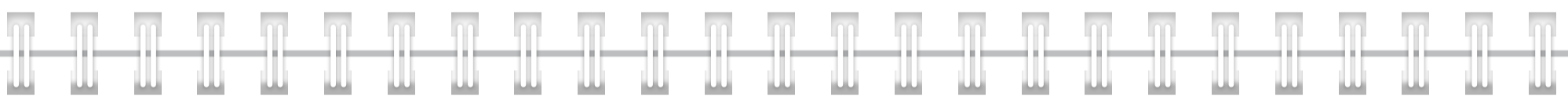
Education Development Trust and our Data Insights approach

Education Development Trust is an international not-for-profit organisation that improves education systems internationally to transform lives, providing millions of people with greater opportunities to learn and thrive. We design and implement improvement programmes for school systems, deliver

expert careers and employability services, and deploy specialists to provide consultancy services internationally. Our work is informed by our continually refreshed body of research and we invest annually in our programme of educational research.

Education Development Trust Data Insights Reports provide high quality quantitative analyses and research insights from our portfolio of live programme delivery around the world. As such they are focused on practical questions related to delivery and the identification of evidence-informed solutions in the field.

[1] BPS Register of Qualifications in Test Use (RQTU) number: 221348.

Foreword by Patrick Brazier

Improving teaching quality is at the heart of what we do at Education Development Trust. We know that effective teaching starts with excellent teachers. Because of this recruiting and retaining high quality teachers has long been a key aim of education policies around the world.

However, in many places, teacher recruitment targets have not been reached, especially in STEM subjects – and many trainee and newly qualified teachers leave the profession within a short period of time. This has real implications for student outcomes, which tend to be worse in schools with high staff turnover. More experienced teachers who have developed their practice tend to be more effective than newly qualified staff. Therefore, teacher retention – especially among new teachers – has to be about more than just ensuring adequate numbers: it also has implications for improving overall teaching quality.

While there are many factors affecting teacher retention rates, one area we can consider, especially where early career

teachers are concerned, is how we recruit and assess candidates for teacher training, and whether there are specific qualities or attributes which indicate how likely they are to remain in the profession. In this paper, we consider the evidence from our own assessment centres used to select candidates for our Future Teaching Scholars programme, and consider how the chosen applicants have fared throughout their teacher training. In doing so, our analysis offers lessons for teacher recruitment as part of teacher workforce reform. This is the first published evaluation of an education assessment centre, and the findings are fascinating and – in some cases – surprising. We hope that as a result, we can begin to lead the way in developing smarter recruitment and assessment strategies, thereby helping to ensure that more outstanding new teachers are recruited – and remain in the classroom for the long haul.

Dr Patrick Brazier,
Chief Executive Officer,
Education Development Trust

> 'Teacher retention – especially among new teachers – has to be about more than just ensuring adequate numbers: it also has implications for improving overall teaching quality.'

Introduction
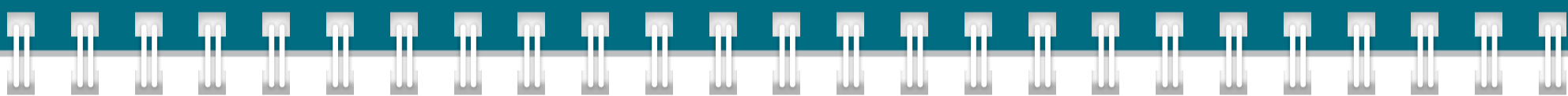
69 million (or maybe more?) – the global need to get teacher recruitment right

Teacher recruitment is a global priority

The United Nations Educational, Scientific and Cultural Organisation (UNESCO) estimates that to achieve its education-related sustainable goal (SDG 4),[2] recruitment of 68.8 million teachers will be required (14.4 million primary teachers; 24.4 million secondary teachers).[3]

This cannot simply be a numbers game

• Quality teachers teach quality lessons – and quality lessons improve learner outcomes.

• In parallel, there is a crisis in teacher retention: almost as fast as we recruit teachers, we lose them early in their career.[4]

14.4 million primary teachers

ENSURE INCLUSIVE AND EQUITABLE QUALITY EDUCATION
Recruitment of
68.8 million teachers
AND PROMOTE LIFELONG OPPORTUNITIES FOR ALL

24.4 million secondary teachers

[2] Sustainable Development Goal 4: "Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all." [3] UNESCO (2016). [4] Carver-Thomas and Darling-Hammond (2019), Mason and Matas (2015), OECD (2005; 2010; 2015).

The reasons behind the recruitment and retention challenge are complex – and the consequences are serious[5]

In England, workload can heavily influence a teacher's decision to leave teaching.

• Early career teachers often make the decision to leave within the first three months of starting to teach.

• There is much that can be done to reduce teachers' workload without affecting pupils' attainment,[6] and in turn, the personal characteristics of teachers seem important predictors of turnover.[7]

• High staff turnover constitutes a major drain on the resources available to develop a modern education system.

In the United States alone, schools lose between $1 billion and $2.2 billion in attrition costs each year as a result of teachers moving or leaving the profession.[8]

$1 - $2.2 billion LOSS

TO SCHOOLS IN THE US AS A RESULT OF TEACHERS MOVING OR LEAVING THE PROFESSION

[5] CooperGibson (2018), Glazer (2018), Guarino et al. (2006). [6] Churches (2020). [7] Borman and Dowling (2008). [8] Alliance for Excellent Education (2014).

Future Teaching Scholars and this data insights analysis

The Future Teaching Scholars programme is a six-year route into teaching.[9]

• During three years of an undergraduate degree, Scholars undertake experiences in schools, alongside online learning, face-to-face training and conferences.
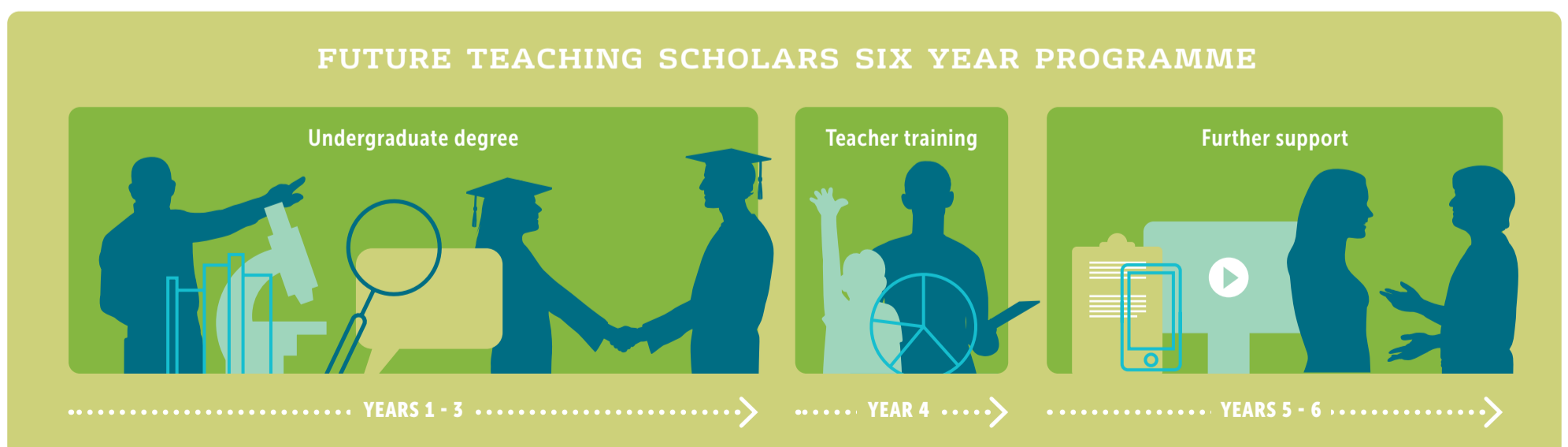
• In the fourth year, Scholars join initial teacher training as an employed unqualified teacher.

• Following attainment of Qualified Teacher Status, they receive two further years of support to help their early career development.

• Scholars also receive a conditional grant of £15,000.[10]

This Data Insights Report evaluates the assessment centre[11] used as part of the Future Teaching Scholars selection process.[12] It contains findings

related to the internal consistency (reliability),[13] construct validity,[14] and predictive validity[15] of the assessment centre.

FUTURE TEACHING SCHOLARS SIX YEAR PROGRAMME

Undergraduate degree	Teacher training	Further support
YEARS 1 - 3	YEAR 4	YEARS 5 - 6

[9] www.futureteachingscholars.com. [10] Scholars are required to pay back their grant if they do not remain on the programme, enter the teaching profession and/or adhere to a stringent code of conduct. [11] Woodruff (2007). [12] The Future Teaching Scholars assessment centre was designed with reference to British Psychological Society best practice guidelines (BPS, 2005) and was reviewed by independent chartered occupational psychologists before implementation (www.shirepro.co.uk). Education Development Trust (2016). [13] Internal consistency is a method for measuring reliability. It can help to judge how well items on a test (or a group of measures) produce comparable results when measuring the same construct. [14] Construct validity is the extent to which a test or tool measures the construct it intends to measure. [15] Predictive validity is the extent to which scores on a test (or scale) predict scores on a criterion (or performance) measure.

Assessment Centres

Since their first use in the middle of the twentieth century,[16] recruiters and employers have seen assessment centres as effective tools for selection, promotion, the diagnosis of strengths and weaknesses and developing managerial competence.[17]

• In the UK, a wide range of organisations make use of assessment centres.[18]

• Assessment centres developed within applied and organisational psychology.[19] Organisations that use them often use this provenance to give their recruitment processes credibility.

However, as with all forms of psychometric measurement, it is vital that the validity and reliability of such processes is continuously evaluated in the new context in which they are being used. This is particularly

important early in the use of an assessment centre, to check that the approach is serving its intended purpose.[20]

The importance of conducting assessment centre research in education

Education Development Trust designed the Future Teaching Scholars programme to include an assessment centre to help ensure value for money.

• Well-designed assessment centres[21] are more likely to produce data that can be used to predict future performance than other forms of selection,[22] with a range of notable levels of predictive validity.[23]

• It is not enough to assume the design of an individual assessment centre will achieve the same levels of validity and reliability that previous centres have achieved.[24]

A systematic review of the main education databases and other search systems[25] showed that, despite increased use of assessment centres in education,[26] no studies of education assessment centre validity and reliability have been published.

[16] Bray and Grant (1966). [17] Lievens and Thornton (2005), Thornton and Rupp (2005), Povah and Thornton (2016), Woodruff (2007). [18] Such as IBM, BDO UK LLP, PWC, the RAF, MI5, Network Rail, Ernst and Young and the Civil Service Fast Stream (Bath Spa University, 2019). [19] Andersen et al. (2008), Cascio and Aguinis (2005). [20] British Psychological Society (2005). [21] Cascio and Aguinis (2005). [22] Meriac et al. (2008). [23] For example, .68 (Anderson and Cunningham-Snell, 2000); .53 (Gaughler et al., 1987); .37 (Schmidt and Hunter, 1988). Even at the lowest end of estimate (.37), such findings compare favourably with many medical treatments. In contrast, unstructured interviews have been found to be valid at levels between .11 and .33 (Anderson and Cunningham-Snell, 2000). [24] Smith (1994), Thornton and Gibbons (2009). [25] Australian Education Index (1977 – current), British Education Index (1975 – current), ERIC (1966 – current), Google Scholar, Web of Science. Search terms: education; assessment centre; recruitment, selection, teacher training. [26] Such as Future Teaching Scholars [2016 to present], Fast Track Teaching assessment centre [2001 to 2009] (Churches, Hutchinson and Jones, 2010; Jones, 2008) and the Teach First assessment centre [2002 to present] (McConney, Price and Woods-McConney, 2012).
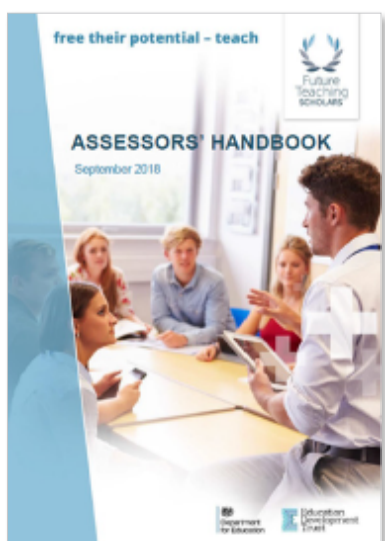
The Future Teaching Scholars Assessment Centre[27]

Future Teaching Scholars is one of a series of Science, Technology, Engineering and Mathematics (STEM) recruitment and retention strategies implemented by the Department for Education in England.

Through the recruitment of people with outstanding subject knowledge, Future Teaching Scholars seeks to:

• Draw on what we know about the importance of subject pedagogical knowledge in teaching (teachers' content knowledge and their ability to understand how students think about the subject and identify common misconceptions).

• Recognise that this is mediated by other important factors (quality of instruction, classroom climate, classroom beliefs, professional behaviour).[28]

[27] See www.futureteachingscholars.com for eligibility requirements. [28] Coe et al. (2014).

Hypothesised importance of simulation within the assessment centre

The classroom roleplay[29] aims to measure a candidate's innate 'mental set'[30] prior to teacher training:

• **'With-it-ness'**
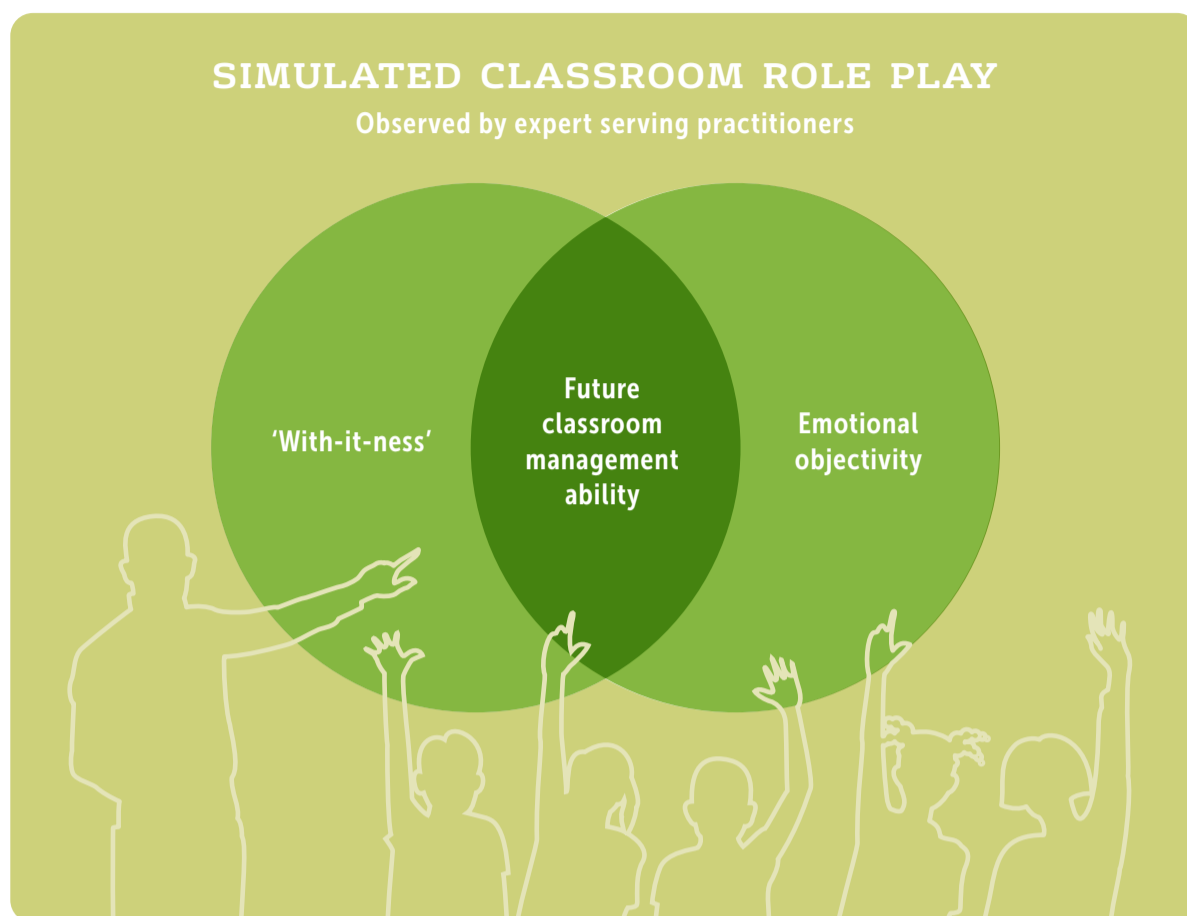The ability to monitor and quickly identify a potential problem and take action

and

• **'Emotional objectivity'**
Staying calm, not getting angry or becoming frustrated.

Together, these areas are known to have a large positive effect on classroom management.[31]

The Future Teaching Scholars assessment centre is unique in that it includes serving subject-specific teachers from outstanding Teaching Schools as part of the assessment centre team.

SIMULATED CLASSROOM ROLE PLAY
Observed by expert serving practitioners

'With-it-ness'

Future classroom management ability

Emotional objectivity

[29] For the classroom roleplay, candidates will have prepared a short lesson. They teach this to two assessors (at least one of whom is a serving teacher). The assessors interject in the way that students might, so that the assessors can see the candidate's ability to respond in the moment to potentially frustrating situations (such as low level disruption and a student struggling to understand what is being taught). [30] Marzano, Marzano and Pickering (2003). [31] d = 1.29 (ibid.).
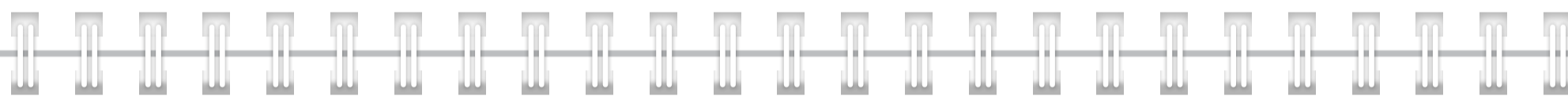
The Future Teaching Scholars recruitment process uses twelve competencies and values to assess candidates:

You as a scientist/mathematician:
• Passion for mathematics or physics
• Knowledge and skill in mathematics or physics
• Initiative and problem-solving ability
• Ability to explain mathematics and physics concepts

You as a teacher:
• Belief in the power of teaching to drive social change
• Passion for working with young people
• Ability to balance competing objectives
• Understanding of the role of a teacher

You as a person:
• Empathy and/or cultural sensitivity
• Perseverance in the face of challenges
• Reliable, responsible and committed
• Reflective and committed to self-improvement

During the day,[32] assessors grade candidates against ten of the competencies and values. Each of these ten competencies and values is evaluated twice within four activities:[33]

Competency-based interview
A 25-minute interview consisting of five standardised questions. Applicants have five minutes to read these questions before the interview begins.

Classroom roleplay
An eight-minute teaching episode assesses applicants' ability to solve problems and to explain concepts in their subject.

Reflection
Applicants reflect on their classroom roleplay performance, evidencing their self-awareness and drive to improve, by talking about their strengths and areas for improvement.

Group problem-solving exercise
A group exercise assesses empathy, cultural sensitivity and problem-solving. Future Teaching Scholars uses a 'collaborative' problem-solving activity rather than a 'competitive' one, and therefore successful candidates are those who demonstrate the greatest ability to work with others effectively.[34]

[32] A film about the assessment centre process was developed to support candidates in preparing for assessment. This can be viewed at: https://www.youtube.com/watch?v=25I9IoSxLB4. [33] See Statistical Appendices: Table 1. [34] The decision whether to use a collaborative or competitive group activity as part of a selection process is a critically important one, which can have far reaching consequences once candidates are in post.

The current global challenges, as described earlier, and the approach chosen on Future Teaching Scholars, beg a number of important questions:

'What is the best way to assess teacher potential (both in terms of cognitive and non-cognitive traits, and retention)?'
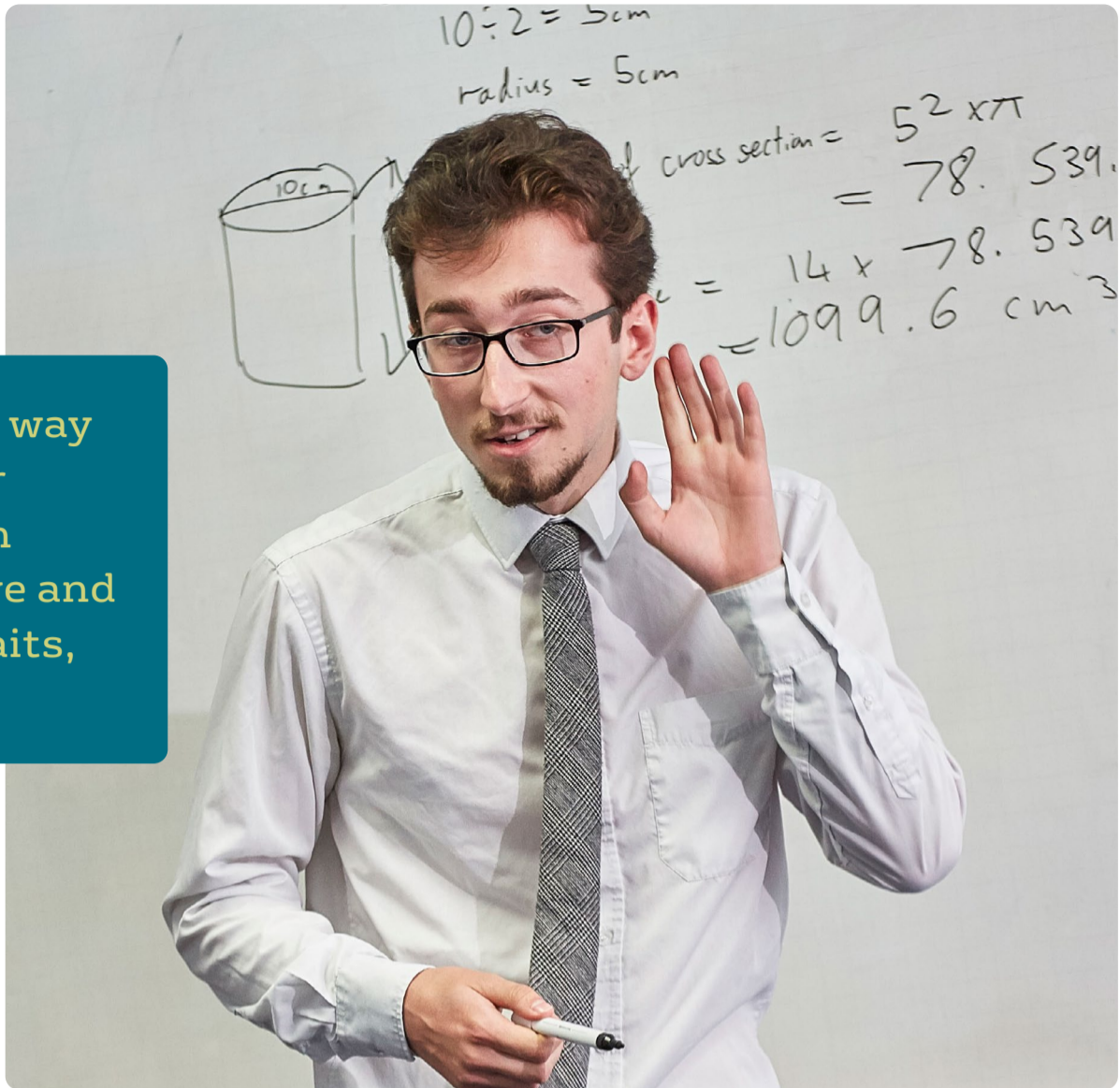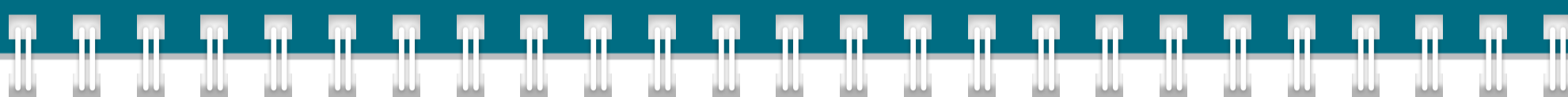
'Can such traits be measured with validity and reliability in an assessment centre?'

'If traits can be measured, do they predict teacher effectiveness, student outcomes and retention over the short-, medium- and long-term?'

Data insights approach[35]

Four areas of analysis

ANALYSIS ONE

Internal consistency

Cronbach's alpha evaluated the reliability of the assessment centre scales and competencies.

ANALYSIS TWO

Construct validity

Exploratory factor analysis measured the comparability of the competencies.[36]

ANALYSIS THREE

Predictive validity

Teachers' classroom performance on a Teacher Practice Tool was compared to their scores on the ten competencies that were appraised at the assessment centre.[37]

Future Teaching Scholars who were in their first year of teaching and completing initial teacher training[38] were observed and scored using the Teacher Practice Tool – a 27-item teacher observation inventory for assessing teacher practice and giving feedback.[39, 40]

ANALYSIS FOUR

Comparison of teachers' classroom performance with expected levels for Qualified Teacher Status[41]

We extended the predictive validity analysis to benchmark Future Teaching Scholars' progress towards Qualified Teacher Status.[42] We asked highly experienced serving teacher trainers from the programme's partner Teaching Schools (and School-Centred Initial Teacher Training centres[43]) to estimate where a teacher trainee crossing the thresholds, exemplified by the Teachers' Standards, might score on the Teacher Practice Tool areas.

This enabled a comparison of Scholar classroom practice with these benchmark scores.

[35] Data for participants who attended the assessment centre from 2016 to 2019 was available (N = 329). We conducted the following analyses on this data and on classroom observation data available for the first cohort recruited in 2016, who are now serving teachers (n = 33). [36] In terms of statistical power (a sample size sufficient to detect a significant effect), these results should be considered preliminary. Although we had access to data from 329 assessment centre participants, a wide range of recommendations are made with regard to sample size for effective factor analysis. Gorsuch (1983) recommend five subjects per item, with a minimum of 100 subjects. Guilford (1954) at least 200, while Cattell (1978) recommends three-to-six subjects per item, with minimum of 250. Comrey and Lee (1992) provide the following guidance in relation to the adequacy of sample sizes: 1000 or more = excellent, 500 = good, 300 = good, 200 = fair, 100 = poor. [37] Power analysis, using G*Power (Faux et al, 2007; 2009) indicated that the serving teacher sample size would be sufficient to give 80% power to detect an r = .34 effect as significant with alpha = .05 (two-tailed). [38] n = 33. [39] We used a 7-point scale adaption of the tool with the five written criteria for each item used as a rubric for the allocation of levels 1, 2, 3-4-5, 6 and 7, respectively. [40] G*Power indicated that using a one sample sign test there would be 80% power to detect a d = 0.52 effect size as significant with alpha = .05 (two-tailed). [41] Using a two-tailed one sample sign test. [42] Department for Education (2013). [43] SCITT Qualified Teacher Status is only available in England. As with other initial teacher training, the quality of provision is regularly inspected by Ofsted. To be a Future Teaching Scholars SCITT, the training must be good or outstanding.

Findings
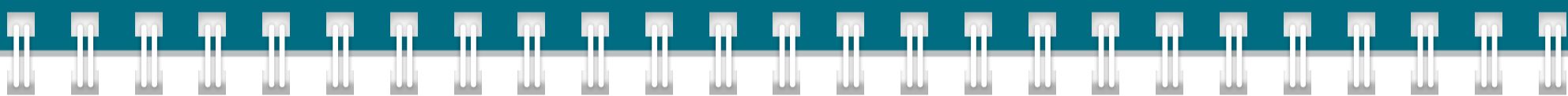
Internal consistency[44] and construct validity

The Future Teaching Scholars assessment centre is internally consistent[45] and has construct validity with respect to two factors (contrasting with the hypothesised three competency areas):

• Behavioural competence and the application of subject knowledge during simulations (Factor A); and

• Candidates' thoughts (expressed during an interview) about joining the teaching profession and working with young people (Factor B).[46]

This is important, as other analyses show that these two clusters of data produce quite different findings on the prediction of later teacher performance in the classroom.

In addition:

• Empathy and cultural awareness were strongly associated with candidates' overall performance across the simulations,[47] alongside initiative and problem solving;[48]

• Both competencies were assessed during the classroom simulation and group problem-solving activities, further pointing to the potential value of such simulations in teacher recruitment.

Further, **candidates who scored highly for empathy during the classroom and group problem-solving simulations were less likely to drop out of the programme** prior to being employed in a school in their fourth year.[49]

[44] We conducted a reliability analysis on the assessment centre scales including all 17 items. [45] α = .88. Alpha decreased if we removed any items, showing that all competences are worthy of retention in future use of the assessment centre. For completeness, we then assessed the averaged repeated competencies (ten items). As before, this showed that the assessment centre had good reliability, with again, a decrease on deletion of any items. By convention values above .70 are considered acceptable, with those greater than .80 or .90 good or excellent, respectively (Nunnally, 1978; Hair et al., 2006). [46] Exploratory factor analysis is not a single fixed approach and requires the researcher to make decisions about the analysis at critical stages in the process. Initially, we conducted a principal axis factoring analysis with oblique (promax) rotation on the averaged repeated indicators dataset, to avoid repeated indicators being over-represented. We used an oblique rotation as some correlation across factors was expected. An unrotated extraction of two factors accounted for 52.9% of the variance, with a correlation of .64. [47] .94. [48] .85. Other competencies making a significant contribution to Factor A included: passion for mathematics or physics (.63); ability to balance competing objectives (.62); ability to explain mathematics or physics concepts (.61). Three competencies made a significant contribution to Factor B: belief in the power of teaching to drive social change (.83); understanding the role of a teacher (.66); passion for working with young people (.53). [49] A. Mann-Whitney test indicated that empathy and cultural awareness was greater for those remaining on the programme (Mean = 3.65) than for those who had withdrawn (Mean = 3.40), U = 2965.5, p =.016. The effect size for this difference was moderately large (d = 0.31).

Predictive validity[50]

To help interpretation, we built an intercorrelation matrix.[51] Twelve correlations were significant, suggesting that the assessment centre competences may be predictive of twelve areas of teacher performance in the classroom. [52, 53]

- Teachers' **ability to explain mathematical or scientific concepts** (assessed during the assessment centre competence-based interview and classroom roleplay) predicted their later demonstration of high expectations, and their ability to structure learning, maximise learning time, respond to students' answers, provide feedback and encourage discussion, as well as their ability to provide a variety of learning tasks that enable students to see, understand and master the content they are learning.[54]

- Assessment centre scores for **initiative and problem-solving ability** (assessed during the classroom roleplay and group activity) also predicted high expectations and maximisation of learning time.[55]

- In addition, teachers' **passion for mathematics and physics** (measured during the competence-based interview and classroom roleplay) also predicted the effective use of learning time.[56]

Surprisingly, we found significant negative correlations between the teachers' assessment centre scores related to passion for working with young people and four Teacher Practice Tool areas (assessed during the competence-based interview).

This suggests that expressing a more fervent desire to work with young people at interview may, in fact, be an indication of a candidate who:

- Later finds it more difficult to provide learning tasks that engage and provide appropriate challenge for all students;

- Is not as effective at delivering a variety of learning tasks that enables students to see, understand and master the content they are learning (both in general and specifically related to their subject area);

- May struggle with the continuous assessment of students.[57]

Looking beyond the significant correlations at the pattern of effects within the intercorrelation matrix:

- **Passion for working with young people scores were negatively correlated will all areas of the Teacher Practice Tool,[58]** reinforcing its lesser importance in the recruitment process, compared to observed behaviours and competence during simulations.

- Similarly, **understanding the role of a teacher (also assessed in the competence-based interview) negatively correlated with 10 out of 11 Teacher Practice Tool areas.[59]**

- Teachers' passion for mathematics and physics as a subject correlated positively with all but one of the Teacher Practice Tool areas.[60]

- **Ability to reflect and show commitment to self-improvement** (assessed during a post-classroom reflection interview), **was positively correlated with all areas**

of the Teacher Practice Tool.[61] A similar pattern can be observed in relation to teachers' abilities to explain maths or physics concepts.[62]

In summary, the predictive validity findings are potentially far-reaching because:

- During recruitment interviews, teacher training programmes often put great store on a candidate's passion for working with children.

- The predictive validity analysis suggests that **recruiters should place more emphasis on the observation of a candidate's actual skills (during a classroom simulation and ability to reflect afterward), rather than the strength of their motivation to work with young people**.

- By extension, this suggests that those who say they are most passionate about working with children may benefit from focused support to develop the ability to deliver in the classroom in a way most likely to yield better pupil outcomes.

[50] Predictive validity is the extent to which a score on a test or scale predicts scores on a criterion (or performance) measure. For example, the predictive validity of a cognitive ability test for job performance can be evaluated using the correlation between test scores and manager performance ratings. In this case, we would consider a cognitive test to have predictive validity if the observed correlations were significant. [51] See Statistical Appendices, Table 2. * = p < .05; ** = p < .01. [52] The small number of significant effects was expected due to the small serving teacher sample size. As with the other analyses, more data will be added over time. [53] Before looking at the ability of the assessment centre to predict teacher performance, we checked the internal consistency of the Teacher Practice Tool results. Alphas were as follows: overall, .97; creating a positive climate, .85; structuring and organising learning, .92; teaching through dialogue, .94; providing well designed learning tasks, .93; assessing learning continuously, .90. [54] r_s = .49**, .43*, .39*, .39* and .40*, respectively. [55] r_s = .37* and .40*, respectively. [56] r_s = .38* [57] r_s = -.43*, -.36*, -.35* and -.42*, respectively. [58] With these effect sizes ranging from r_s = -.10 to -.43. [59] The exception being the demonstration of high expectations (r_s = .11). [60] The exception being continuously assessing students (r_s = -.34). [61] With the range of effects from r_s = .02 to .24. [62] All Teacher Practice Tool areas, r_s = .10 to -.48.

Comparison of Scholars' classroom performance with their peers[63]

By the end of their first term in teaching, Scholars are already performing broadly in line with the performance expected of a teacher at the end of initial teacher training. **This places Scholars, approximately one or two terms ahead of their peers.[64]**

In the following five areas of teacher practice, Scholars were significantly[65] in advance of other teachers completing Qualified Teacher Status at the end of their initial teacher training year:

- Structuring lessons well,[66]

- Asking questions that engage students,[67]

- Asking questions that encourage students to think,[68]

- Responding to students' answers, to provide feedback and encourage discussion,[69]

- Using learning tasks that engage and provide appropriate challenge for all students (all subjects).[70]

[63] See effect size differences in the Statistical Appendices (Table 2). [64] It is difficult to disaggregate whether this enhanced performance is the result of the selection of a particular group of people, or the programme of support and experiences that a Scholar receives during their undergraduate years. However, the predictive validity data suggests that the Future Teaching Scholars assessment centre is able to predict some aspects of classroom performance even with a gap of three years between initial assessment and employment as a teacher. [65] * = p < .05, ** = p < .01, *** = p < .001. [66] r = 0.32***, CI (95%) = 0.15 − 0.46, d = 0.54. [67] r = 0.33***, CI (95%) = 0.12 − 0.49, d = 0.69. [68] r = 0.24*, CI (95%) = 0.01 − 0.43. d = 0.49. [69] r = 0.25, CI (95%) = 0.04 − 0.44, d = 0.52. [70] r = 0.27*, CI (95%) = 0.06 − 0.44, d = 0.55.

Conclusions
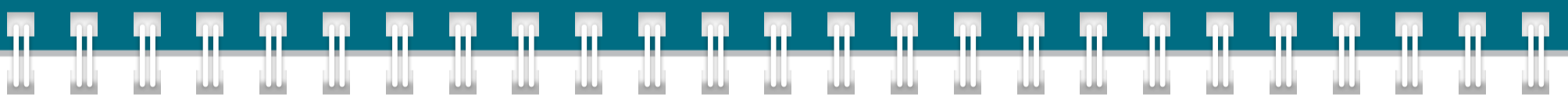
The Future Teaching Scholars Assessment Centre appears to be:

1. Valid and reliable suggesting that assessor training has ensured accurate measurement;

2. Able to make predictions about classroom performance years in advance of a person being employed in a school.

In turn, although limited by the current sample size, the analyses suggest potentially far-reaching implications for teacher recruitment and training programmes:

- A candidate's stated passion for working with children is often a focus for recruitment interviews. This research suggests **recruiters should place more emphasis on the observation of a candidate's actual skills** (during a classroom simulation and ability to reflect afterward).

- By extension, this implies that **those who say they are most passionate about working with children may benefit from focused support to develop the ability to deliver in the classroom in a way most likely to yield better pupil outcomes**.

We will continue to track Future Teaching Scholars until the end of their third year as a serving teacher. At that point, we will be able to make comparisons between their assessment centre scores and classroom performance, retention rates and student examination outcomes.

Towards a blueprint for more effective teacher recruitment and selection
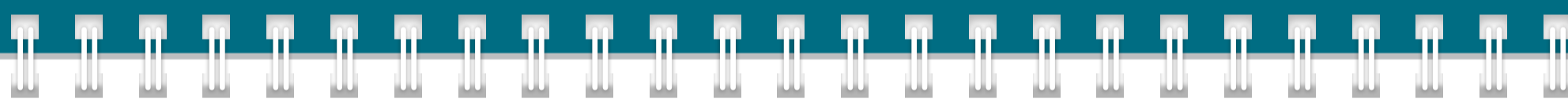
Seven recommendations for improved teacher selection

Assessment centres are time-consuming to design and expensive to run. Therefore, schools and middle-tier[71] organisations are unlikely to fully utilise them. However, the following insights from our research could help schools and other organisations to select teachers more effectively.

ONE

Although interviewing candidates is, of course, important, **selection processes should contain some form of behavioural simulation (such as a short teaching activity).**

TWO

To draw out a candidate's 'mental set', **classroom simulations should include elements where the 'learners' are briefed in advance to provide unexpected challenge.** Such challenges should require the candidate to demonstrate 'with-it-ness' and 'emotional objectivity' (e.g. repeatedly not understanding/low-level disruption). The length of the simulation is probably less important than its quality.

THREE

Where a simulation is used, a **clear scoring rubric** should be written in advance, and the assessors (observers) should be carefully trained to apply it consistently.

FOUR

Serving teachers can effectively role-play learners. However, schools could find drawing on school pupils an effective approach (providing the children are well-trained and able to be consistent across candidates).

FIVE

In simulations during initial teacher training programme recruitment, looking for empathy and cultural awareness could be a helpful way to identify candidates who are more likely to remain in training.

SIX

In weighting candidates' scores during a selection process, more emphasis should be placed on a candidate's demonstration of their abilities during engagement with others, than on what a candidate says about themselves and their skills in these areas.

SEVEN

To support teachers' early development in the classroom, results from observations can be fed forward into teacher's development planning (e.g. through the identification of actions to support the development of weaknesses that were observed during selection).

[71] Such as multi-academy trusts, university departments, teacher training centres, local authorities etc.

The key elements of classroom simulation

Observed behaviours

Consistent role playing

Challenge to assess 'with-it-ness' and emotional objectivity

Empathy assessed through what is seen rather than said

User feedback for development planning

Expert trained observers
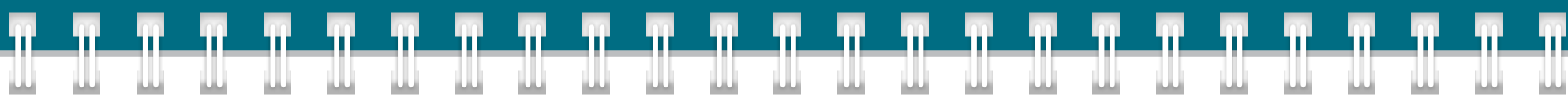
Statistical appendicies

Table 1: Matrix illustrating which selection activities contribute statistical data to which competencies and values

COMPETENCIES	RECRUITMENT ACTIVITY	ASSESSMENT CENTRE				ONLINE APPLICATION	
		Competency-based interview	Classroom roleplay	Reflection (post-classroom roleplay)	Group activity	Paperwork	Essay
You as a scientist/mathematician	Passion for mathematics or physics	✓	✓				
	Knowledge and skill in mathematics or physics					✓	
	Initiative and problem-solving ability		✓		✓		
	Ability to explain mathematics or physics concepts	✓	✓				
You as a teacher	Belief in the power of teaching to drive social change	✓					✓
	Passion for working with young people	✓					
	Ability to balance competing objectives	✓	✓				
	Understanding the role of a teacher	✓					✓
You as a person	Empathy and/or cultural sensitivity		✓		✓		
	Perseverance in the face of challenges	✓	✓				
	Reliable, responsible and committed					✓	
	Reflective and committed to self-improvement	✓		✓			

Table 2: Intercorrelation matrix Teacher Practice Tool elements versus the ten competences and values assessed at the assessment centre[72]

| ASSESSMENT CENTRE COMPETENCIES		YOU AS A SCIENTIST/MATHEMATICIAN			YOU AS A TEACHER				YOU AS A PERSON		
		Passion for maths or physics	Initiative and problem-solving ability	Ability to explain maths or physics concepts	Belief in the power of teaching to drive social change	Passion for working with young people	Ability to balance competing objectives	Understanding the role of a teacher	Empathy and/or cultural sensitivity	Perseverance in the face of challenges	Reflective and committed to self-improvement
Creating a positive climate	The teacher demonstrates high expectations	.339	**.374***	**.489****	.0004	-.101	.322	.105	.138	.218	.244
	The teacher treats all students fairly	.039	.291	.268	-.031	-.209	.029	-.094	.014	-.037	.027
	The teacher recognises pupils with Special Educational Needs and provides them with relevant support	.010	.083	.104	-.092	-.343	-.118	-.105	-.093	.064	.030
Structuring and organising lessons	Lessons are well-structured	.333	**.404***	**.425***	.049	-.102	.151	-.041	.160	.329	.134
	Learning time is maximised	**.383***	.342	**.390***	.035	-.101	.169	-.080	.235	.279	.126
Teaching with dialogue	The teacher's explanations and instructions are clear	.126	.215	.329	.074	-.199	.137	-.076	-.039	.096	.118
	The teacher asks questions in ways that engage students	.038	.212	.138	.0003	-.194	-.112	-.048	-.043	.197	.084
	The teacher asks questions that encourage students to think	.099	.063	.224	.150	-.303	.056	-.218	-.086	.078	.136
	The teacher responds to students' answers to provide feedback and encourage discussion	.206	.161	**.389***	.239	-.219	.109	-.162	.001	.020	.164
Providing well-designed learning tasks	Learning tasks engage and provide appropriate challenge for all students (all subjects)	.047	.146	.264	-.031	**-.426***	-.030	-.233	.041	.040	.024
	The teacher provides a variety of learning tasks that enable students to see, understand and master the content they are learning (all subjects)	.129	.147	.218	.109	**-.364***	-.043	-.231	.086	.014	.083
	The teacher provides a variety of learning tasks that enable students to see, understand and master the content they are learning (subject specific)	.313	.203	**.398***	.040	**-.348***	.095	-.193	.017	.165	.066
Assessing learning continuously	The teacher continuously assesses students	-.007	.009	.234	.104	**-.415***	.013	-.335	-.155	-.081	.023

[72] * = p <.05, ** = p <.01

Table 3: Future Teacher Scholars' classroom performance versus benchmark for a teacher passing Qualified Teacher Status

		EFFECT SIZE (r) [73]	CI (95%)	P-VALUE [74]	[d]
Creating a positive climate	The teacher demonstrates high expectations	0.228	-0.004 – 0.427	.054	0.404
	The teacher treats all students fairly	0.240	-0.012 – 0.452	.060	0.422
	The teacher recognises pupils with Special Educational Needs and provides them with relevant support	0.183	-0.05 – 0.395	.104	-.092
Structuring and organising lessons	Lessons are well-structured	0.318	0.147 – 0.464	.002	0.535
	Learning time is maximised	0.167	-0.07 – 0.379	.166	0.339
Teaching with dialogue	The teacher's explanations and instructions are clear	0.115	-0.165 – 0.370	.432	0.231
	The teacher asks questions in ways that engage students	0.327	0.124 – 0.493	.002	0.691
	The teacher asks questions that encourage students to think	0.236	0.012 – 0.429	.040	0.487
	The teacher responds to students' answers to provide feedback and encourage discussion	0.253	0.040 – 0.435	.020	0.523
	Learning tasks engage and provide appropriate challenge for all students (all subjects)	0.265	0.057 – 0.442	.014	0.549
Providing well-designed learning tasks	The teacher provides a variety of learning tasks that enable students to see, understand and master the content they are learning (all subjects)	0.206	-0.028 – 0.409	.084	0.420
	The teacher provides a variety of learning tasks that enable students to see, understand and master the content they are learning (mathematics)	0.123	-0.104 – 0.332	.292	0.247
	The teacher provides a variety of learning tasks that enable students to see, understand and master the content they are learning (physics)	0.022	-0.026 – 0.069	.376	0.044
Assessing learning continuously	The teacher continuously assesses students	0.203	-0.019 – 0.398	.072	0.415

[73] We have reported a non-parametric equivalent of Cohen's d because of the distribution (Rosenthal, 1994). [74] Two-tailed one sample sign test.

References
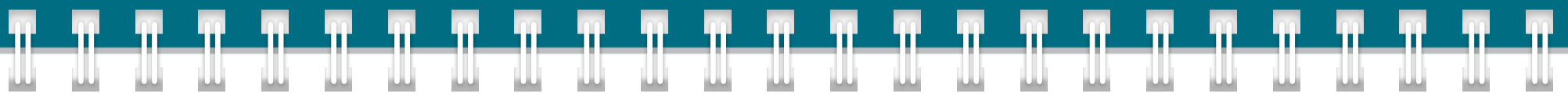
Anderson, N., Salgado, J., Schinkel, S. and Cunningham-Snell, N. (2008) Personnel selection. In N.Chmiel (Ed.), pp.257–280, *Introduction to work and organizational psychology*. Oxford: Blackwell.

Alliance for Excellent Education (2014) On the path to equity: improving the effectiveness of beginning teachers [online paper]. Washington: Alliance for Excellent Education.

Bath Spa University (2019) Guild to assessment centres [online paper]. Bath: Bath Spa University Careers.

Bray, D.W. and Grant, D.L. (1966) The assessment centre in the measurement of potential for business management. *Psychological Monographs*, 80, 1–2.

British Psychological Society (2005) Design, implementation and evaluation of assessment and development centres. Best practice guidelines [online paper]. Leicester: British Psychological Society.

Borman, G.D. and Dowling, N.M. (2008) Teacher attrition and retention: a meta-analytic and narrative review of the research. *Review of Educational Research*, 78, 367e409.

Carver-Thomas, D. and Darling-Hammond, L. (2019) The trouble with teacher turnover: how teacher attrition affects students and schools. *Education Policy Analysis Archives*, 27(36), 2–32.

Cascio, W.F. and Aguinis, H. (2005) *Applied psychology in human resource management*. New Jersey: Prentice Hall.

Cattell, R.B. (1978) *The scientific use of factor analysis*. New York: Plenum.

Churches, R. (2020) Supporting teachers through the school workload reduction toolkit [online paper]. London: Department for Education.

Churches, R., Hutchinson, G. and Jones, J. (2009) Fast Track teaching: beginning the experiment in accelerated leadership development. *School Leadership and Management*, 29(3), 277–293.

Coe, R., Aloisi, C., Higgins, S. and Elliot Major, L. (2014) What makes great teaching? [online paper] London: Sutton Trust.

Comrey, A.L. and Lee, H.B. (1992) *A first course in factor analysis*. New Jersey: Erlbaum.

CooperGibbson Resaerch (2018) Factors affecting teacher retention: qualitative investigation, research report [online paper]. London: Department for Education.

Costa, P.T. and McCrae, R.R. (1992) *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) manual*. Florida: Psychological Assessment Resources.

Department for Education (2013) Teachers' standards guidance for school leaders, school staff and governing bodies [online paper]. London: Department for Education.

Dunn, R. and Dunn, K. (1992) *Teaching secondary students through their individual learning styles*. Massachusetts: Allyn and Bacon.

Education Endowment Foundation (2017) Sutton Trust–Education Endowment Foundation Teaching and Learning Toolkit [online resource]. London: Education Endowment Foundation.

Education Development Trust (2016) *Assessment centre: assessor handbook*. Reading: Education Development Trust/Department for Education.

Faul, F., Erdfelder, E., Lang, A.G. and Buchner, A. (2007) G*Power 3: a flexible statistical power analysis program for the social, behavioural and biomedical sciences. *Behaviour Research Methods*, 39, 175 – 191.

Faul, F., Erdfelder, E., Buchner, A. and Lang, A.G. (2009) Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behaviour Research Methods*, 41, 1149 – 1160.

Gaugler, B.B., Rosenthal, D.B., Thornton, G.C. and Bentson, C. (1987) Meta-analysis of assessment centre validity. *Journal of Applied Psychology*, 72, 493–511.

Glazer, J. (2018) Learning from those who no longer teach: viewing teacher attrition through a resistance lens. *Teaching and Teacher Education*, 74, 62–71.

Gorsuch, R.L. (1983). *Factor Analysis*. New Jersey: Erlbaum.

Guarino, C.M., Santibanez, L. and Daley, G. (2006) Teacher recruitment and retention: a review of the recent empirical literature. *Review of Educational Research*, 76, 173–208.

Hair, J., Black, W., Babin, B., Anderson, R. and Tatham, R. (2006) *Multivariate data analysis*. New Jersey: Pearson Prentice Hall.

Jones, P. (2008) *Review and evaluation of the Fast Track Teaching programme*. Nottingham: National College for School Leadership.

Lievens, F. and Thornton, G. C., III (2005) Assessment centres: recent developments in practice and research. In A.Evers, N.Anderson, and O.Voskuijl (Eds.), pp.243–264, *The Blackwell handbook of personnel selection*. Massachusetts: Blackwell.

Povah, N. and Thornton III, G.C. (Eds.) (2016) *Assessment centres and global talent management*. Oxford: Routledge

Marzano, R.J., Marzano, J.S. and Pickering, D.J. (2003) *Classroom management that works: research-based strategies for every teacher*. Virginia: Association for Supervision and Curriculum Development.

Mason, S. and Matas, C.P. (2015) Teacher attrition and retention research in Australia: towards a new theoretical framework. *Australian Journal of Teacher Education*, 40(11, 3), 45–66.

McConney, A., Price, A. and Woods-McConney, A. (2012) *Fast track teacher education: A review of the research literature on Teach For All schemes*. Perth: Murdoch University, Centre for Learning, Change and Development.

Meriac, J.P., Hoffman, B.J. Woehr, D.J. and Fleisher, M.S. (2008) Further evidence for the validity of assessment centre dimensions: a meta-analysis of the incremental criterion related validity of dimension ratings. *Journal of Applied Psychology*, 93(5), 1042–1052.

Nunnally, J.C. (1978) *Psychometric theory*. McGraw-Hill, New York.

Organisation for Economic Cooperation and Development (2005) Teachers matter: attracting, developing and retaining effective teachers [online paper]. Paris: OECD.

Organisation for Economic Cooperation and Development (2010) Social capital, human capital and health. What is the evidence? [online paper] Paris: OECD.

Organisation for Economic Cooperation and Development (2015) TALIS - The OECD teaching and learning international survey [online paper]. Paris: OECD.

Rosenthal, R. (1994) Parametric measures of effect size. In H.Cooper and L.V.Hedges (Eds.), *The handbook of research synthesis*, pp. 231–244, New York: Russell Sage Foundation.

Schmidt, F.L. and 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.

Smith, M. (1994) A theory of the validity of predictors in selection. *Journal of Occupational and Organizational Psychology*, 67(1), 13–31

Thornton, G.C., III, and Rupp, D.E. (2005) *Assessment centres in human resource management: strategies for prediction, diagnosis, and development*. New Jersey: Lawrence Erlbaum.

Thornton, G.C. and Gibbons, A.M. (2009) Validity of assessment centres for personnel selection. *Human Resource Management Review*, 19, 169–187.

Woodruffe, C. (2007) *Development and assessment centres: identifying and developing competence*. London: Human Assets.

United Nations Educational, Scientific and Cultural Organization (2016) The world needs almost 69 million new teachers to reach the 2030 Education goals, UNESCO Institute for Statistics (UIS), Fact Sheet No 33 [online paper]. Paris: UNESCO

Education Development Trust's large scale delivery is underpinned by evidence and research. As a registered charity, we reinvest a percentage of our surplus into a publicly available programme of educational research. We also invest in an ongoing Research and Development cycle to rigorously review the impact of our own programmes, generate new evidence on what works, and share insights to support broader debate and policy.

The evidence and insights shared in this handbook are part of this ongoing commitment to Research and Development.