Selection of medical students: Review of theoretical models & assessment methodologies

Professor Fiona Patterson
Work Psychology Group Ltd & University of Cambridge, UK

NVMO Conference, 2013
Overview

• **Key concepts** in selection system design

• Research on the **accuracy, strengths & limitations** of various selection methods

• Case material
  1. *Postgraduate selection* (UK General Practice)
  2. *Medical school admissions* (Situational Judgement Tests; SJTs)

• Implications for **theory, policy & practice**
Selection through the ages…

“Work for me, son – I knew your father.”

“Fill out the application form and the job is yours, mate.”

“It isn’t an interview – just be an informal chat, sweetie. Just a formality.”
Selection system design

Role/Job Analysis
- Tasks, roles, responsibilities
- Required level of competence

Create competency model

Create person specification
- Knowledge, skills, abilities, characteristics

Identify selection criteria

Attract pool of applicants

Select in (rank) suitable applicants

Selection

Choose selection methods
- Interviews
- Simulations
- Situational Judgement tests
  - Reliability, validity, utility, fairness

Evaluate candidate reactions

Validate selection decisions

Make selection decisions

Attract pool of applicants

Select out unsuitable applicants

Selection system design

Evaluate candidate reactions

Validate selection decisions
Judging the effectiveness of medical selection?

1. evidence for **reliability** & **validity**
2. ease of **interpretation**
3. **utility** (minimise costs for Schools & maximise value)
4. costs for applicants
5. **practicality**
6. ease of **administration**
7. public interest/stakeholder acceptence?
8. litigation risk
9. ‘widening access’ & **fairness**
10. generates appropriate feedback
11. positive candidate reactions
12. educational impact/value
13. susceptibility to coaching
14. political validity?
The purpose of selection?

Number of Applicants

Minimum level to achieve competence

ABILITY

Posts available (may change from year to year)
• 26,000 applicants for 8,000 medical school places
• 8,000 medical students apply for their first post
• 12,000 speciality applicants
• 24,000 + interviews
• Weeks of offering, rejecting, cascading

1000s Consultant hours

Help!
Selection - Validity of Predictors

High Validity

Predictor Cutoff

- r = .65
- Satisfactory
- Unsatisfactory
- Correct Hires
- Misses
- Correct Rejections
- False Positives
- Reject
- Accept

Predictor Cutoff

- r = .00
- Satisfactory
- Unsatisfactory
- Correct Hires
- Misses
- Correct Rejections
- False Positives
- Reject
- Accept
Key research questions

What attributes are important to be an effective clinician, now and in the future?

What selection methods are available to test these accurately?

Given the costs - beyond some basic assessment – is a lottery the best option?
Selected medical students achieve better than lottery-admitted students during clerkships

Louise C Urlings-Strop,¹ Axel P N Themmen,¹,² Theo Stijnen³ & Ted A W Splinter²

OBJECTIVES A recent controlled study by our group showed that the dropout rate in the first 2 years of study of medical students selected for entry by the assessment of a combination of non-cognitive and cognitive abilities was 2.6 times lower than that of a control group of students admitted by lottery. The aim of the present study was to compare the performance of these two groups in the clinical phase.

METHODS A prospective cohort study was performed to compare the performance of 389 medical students admitted by selection with that of 938 students admitted by weighted lottery between 2001 and 2004. Follow-up of these cohorts lasted 5.5–8.5 years. The main outcome measures were the mean grade obtained on the first five discipline-specific clerkships by all five clerkships than lottery-admitted students (mean ± standard error [SE] 7.95 ± 0.03, 95% confidence interval [CI] 7.90–8.00 versus mean ± SE 7.84 ± 0.02, 95% CI 7.81–7.87; p < 0.001). This difference reflected the fact that selected students achieved a grade of ≥ 8.0 1.5 times more often than lottery-admitted students. An analysis of all mean grades awarded on 10 clerkships revealed the same results. Moreover, the longer follow-up period over the clerkships showed that the relative risk for dropout was twice as low in the selected student group as in the lottery-admitted student group.

CONCLUSIONS The selected group received significantly higher mean grades on their first five clerkships, which could not be attributed to
Research evidence on the accuracy, strengths & limitations for various selection methods

• Interviews
• Aptitude tests
• Academic records
• Personal statements, essays & autobiographical submissions
• References
• Personality assessments
• Situational Judgement Tests (SJTs)
• Selection centres
Interviews

- Widely used for many years (even without sound evidence base)
- Format varies widely – ‘traditional’, structured & Multiple Mini Interview

<table>
<thead>
<tr>
<th>Strengths</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Means of assessing non-cognitive skills</td>
<td>Careful design is required to ensure good reliability</td>
</tr>
<tr>
<td>Good approach for some aspects, such as communication skills</td>
<td>Potential for bias (gender, ethnicity, SES)</td>
</tr>
<tr>
<td>Face validity high</td>
<td>Resource intensive (for applicant &amp; school)</td>
</tr>
<tr>
<td>Some evidence they can be ranked effectively</td>
<td>Rarely clear what content is actually assessed within a composite total score</td>
</tr>
<tr>
<td>Belief may help screen out ‘unsuitable’ entrants</td>
<td>Historically little evidence of predictive validity, though changing</td>
</tr>
</tbody>
</table>
Aptitude Tests

• Mixed findings, depending on the type of test used
• UKCAT in widespread use in UK medical schools. Studies of UKCAT have found:

<table>
<thead>
<tr>
<th>Strengths</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moderately predictive of early medical school performance</td>
<td>Only early years studied so far: some mixed outcomes</td>
</tr>
<tr>
<td>If used in strong form, can aid widening participation</td>
<td>Paid for by applicants, so may inhibit some applications</td>
</tr>
<tr>
<td>Norm referenced</td>
<td>No evidence yet of relevance to performance in medicine</td>
</tr>
<tr>
<td>Detailed analysis of test performance</td>
<td></td>
</tr>
<tr>
<td>Small but incremental validity over A levels</td>
<td></td>
</tr>
</tbody>
</table>
**Academic Records**

- Most widely used selection method and also the best studied

<table>
<thead>
<tr>
<th>Strengths</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moderate to good predictor of medical school performance</td>
<td>Less predictive of clinical performance</td>
</tr>
<tr>
<td>Generally administered by other bodies, so low cost to medical schools</td>
<td>In the UK, A Levels are losing discriminating power</td>
</tr>
<tr>
<td>Standardised and well-recognised assessments</td>
<td>Potential socio-economic class bias</td>
</tr>
</tbody>
</table>
Personal Statements, Essays, & Autobiographical Submissions

- Lack of good quality research in relation to this selection method

<table>
<thead>
<tr>
<th>Strengths</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Popular &amp; widespread selection method in medical education</td>
<td>Less reliable than other methods</td>
</tr>
<tr>
<td>Some evidence of predictive validity in relation to medical school performance</td>
<td>Potential data contamination caused by external influences (e.g. length of time to complete application, third party influence, location, etc.)</td>
</tr>
<tr>
<td></td>
<td>Impression management techniques may be used by candidates</td>
</tr>
<tr>
<td></td>
<td>Information inconsistently used by medical schools during the decision-making process</td>
</tr>
</tbody>
</table>
### References & Letters of Recommendation

- Little research examining the use of referee reports in medical selection

<table>
<thead>
<tr>
<th>Strengths</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prevalent selection method in medical education</td>
<td>Poor measures for predicting performance in medical school</td>
</tr>
<tr>
<td></td>
<td>Information contained in reports may result in admission bias</td>
</tr>
<tr>
<td></td>
<td>Information inconsistently used by medical schools during the decision-making process</td>
</tr>
</tbody>
</table>
Personality Assessment

- Overall variable quality of research for this selection method
- Findings are mixed:

<table>
<thead>
<tr>
<th>Strengths</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common assessment method used across multiple industries</td>
<td>No significant link between some personality tools &amp; performance (e.g. MBTI, etc.)</td>
</tr>
<tr>
<td>Moderate to significant links between some personality traits and performance (e.g. PQA, NEO-PI-R, etc.)</td>
<td>Certain traits may be associated with different levels of performance over the course of medical education</td>
</tr>
<tr>
<td>Traits linked to performance:</td>
<td>May lead to a reduction in the diversity of personalities amongst medical students</td>
</tr>
<tr>
<td>- Empathy and motivation → positive</td>
<td></td>
</tr>
<tr>
<td>- ‘Dysfunctional’ personality → negative</td>
<td></td>
</tr>
</tbody>
</table>
Situational Judgement Tests

- Situational Judgement Tests (SJT) are a **measurement method** designed to assess judgement in work-relevant situations:
  - Present challenging situations likely to be encountered at work
  - Candidates make judgements about possible responses
  - Scored against pre-determined key

- SJTs focus on **non-academic/professional attributes**
  (eg. integrity, empathy, resilience, team involvement)
You are reviewing a routine drug chart for a patient with rheumatoid arthritis during an overnight shift. You notice that your consultant has inappropriately prescribed methotrexate 7.5mg daily instead of weekly.

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

A Ask the nurses if the consultant has made any other drug errors recently
B Correct the prescription to 7.5mg weekly
C Leave the prescription unchanged until the consultant ward round the following morning
D Phone the consultant at home to ask about changing the prescription
E Inform the patient of the error
# Situational Judgement Tests (SJTs)

- High quality research, including meta-analyses/systematic reviews

<table>
<thead>
<tr>
<th>Strengths</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>An increasingly popular method of assessment in medical selection</td>
<td>Method of construction &amp; response instructions may affect validity</td>
</tr>
<tr>
<td>Strong predictor of job performance; also predicts performance above cognitive ability &amp; personality tests</td>
<td>Mode of administration may affect candidate reactions (e.g. computer-based vs. video-based)</td>
</tr>
<tr>
<td>Positive candidate reactions</td>
<td>Some item types may be more susceptible to faking, practice &amp; coaching effects than others</td>
</tr>
<tr>
<td>Objective and reliable method of assessment with low-adverse impact to minorities</td>
<td></td>
</tr>
</tbody>
</table>
Selection Centres (SCs)

- Limited, but emerging research examining the utility of SCs in postgraduate selection

<table>
<thead>
<tr>
<th>Strengths</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extensive literature in support of SCs exists outside of medical literature</td>
<td>Potentially more expensive option of assessment; more complex to organise</td>
</tr>
<tr>
<td>Strong evidence on effectiveness available from postgraduate selection research</td>
<td>Requires careful design based on job analysis, which can affect reliability &amp; validity</td>
</tr>
<tr>
<td>Reliable/valid method of assessment</td>
<td></td>
</tr>
<tr>
<td>Uses a multi-trait, multi-method approach</td>
<td></td>
</tr>
<tr>
<td>Selection Method</td>
<td>Cost (school)</td>
</tr>
<tr>
<td>------------------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>Traditional Interviews</td>
<td>Moderate</td>
</tr>
<tr>
<td>Structured Interviews/MMIs</td>
<td>Moderate</td>
</tr>
<tr>
<td>Aptitude testing</td>
<td>Low</td>
</tr>
<tr>
<td>Academic records</td>
<td>None</td>
</tr>
<tr>
<td>Personal statements</td>
<td>Moderate</td>
</tr>
<tr>
<td>References</td>
<td>Moderate</td>
</tr>
<tr>
<td>SJTs</td>
<td>Moderate</td>
</tr>
<tr>
<td>Personality</td>
<td>Low/moderate</td>
</tr>
<tr>
<td>Selection centres</td>
<td>High</td>
</tr>
</tbody>
</table>
5 problems in medical selection research

1. **Over-reliance on reliability estimates** as indicators of quality rather than predictive validity (you could be reliably wrong…)

2. **The ‘criterion problem’** What are we trying to predict? Exam performance? Time to competence? Job performance judged by supervisors, peers, patients?

3. **Few theoretically driven studies** of predictive & construct validity

4. **Designing efficient/scalable methods** to assess important non-academic attributes accurately

5. **Addressing widening participation** in selection system design
Case Study 1.

**Selecting GPs in the UK**

Validity of selection & developments in theory

N=6,500 applicants per year for approx 3,000 training posts
A new competency model for general practice: implications for selection, training, and careers

**BACKGROUND**

There is increasing emphasis on exploring skills and capabilities of GPs outside the consulting room, relating to leadership, professionalism and engagement in commissioning activities. These skills are needed in addition to designing services for patients, with an increasing shift of patient care from hospitals into the community. The GP job role is broadening from a 'helping model' in doctor-patient consultations, to a wider 'business model' where GPs are required to consider how their work impacts communities and fits within the health service as a whole. However, research evidence is sparse on the skills and attributes that are needed by GPs to be effective in the role as it expands.

The current UK GP selection system is reliable, valid, and generates positive candidate reactions. It uses selection criteria previously validated approach, comprising three phases:

- stakeholder consultation, using a combination of semi-structured interviews ($n = 103$), focus groups ($n = 96$), and over 30 hours of behavioural observations of GPs across a range of practices ($n = 6$);
- validation questionnaire based on phase 1 results, administered to a sample of GP stakeholders ($n = 1082$) to examine the initial competency framework; and
- expert panel ($n = 6$) to review evidence from phases 1 and 2 to confirm the competency framework and identify core themes arising from the results.

This three-phase method was designed to triangulate results consistent with previous job analysis studies.
<table>
<thead>
<tr>
<th>Domain</th>
<th>Example Behavioural Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Empathy and perspective taking</td>
<td>Capacity and motivation to view situations from the patient and colleague perspective, acts in an open and non-judgemental manner.</td>
</tr>
<tr>
<td>2. Communication skills</td>
<td>Tailors language to suit the individual and the situation; provides explanations using non-technical language, builds rapport with others.</td>
</tr>
<tr>
<td>3. Clinical knowledge and expertise</td>
<td>Capacity to apply sound clinical knowledge and awareness to full investigation of problems, reflecting good clinical judgement.</td>
</tr>
<tr>
<td>4. Conceptual thinking and problem-solving</td>
<td>Able to synthesise multiple streams of evidence to make effective judgements, makes decisions confidently and in a timely way.</td>
</tr>
<tr>
<td>5. Organisation and management of resources</td>
<td>Is aware of resources available and manages these appropriately, considers implications of actions and activities on available resources.</td>
</tr>
<tr>
<td>6. Professional integrity</td>
<td>Open and honest with others; willing to admit own mistakes, treats others with respect and knows where personal and professional boundaries lie.</td>
</tr>
<tr>
<td>7. Coping with pressure</td>
<td>Remains calm under pressure, demonstrates self-awareness, understanding of own limitations, manages own emotions, is resilient.</td>
</tr>
<tr>
<td>8. Effective team working</td>
<td>Is open to sharing information, collaborative with other professionals, acknowledges and appreciates others’ expertise, willing to learn from others.</td>
</tr>
<tr>
<td>9. Respect for diversity and the law</td>
<td>Demonstrates awareness and is compliant with nationally or locally agreed policies and works to protocol, guidelines, and legislation.</td>
</tr>
<tr>
<td>10. Learning and development of self and others</td>
<td>Committed to the learning and development of self and others, able to self-manage, actively promotes self-directed learning.</td>
</tr>
<tr>
<td>11. Leading for continuing improvement</td>
<td>Shows leadership skills and organisational awareness within and outside of the practice, is an ambassador for the profession, inspires and empowers others and is positive about the future of general practice.</td>
</tr>
</tbody>
</table>
UK GP Selection

Longlisting

Selection Tests

National

Clinical Problem solving test
CPS
Situational Judgement test
SJT

Regional

Selection Centre (SC)

Selection Centre Using Simulated Consultations

Foundation Competency National panel

Ranking

Matching to Region
What do SJTs offer?

• **Cost-effective** to administer & score

• Meta-analytic studies show **good predictive validity** & substantial **incremental validity** over other measures in selection including IQ tests and personality tests (eg. McDaniel et al, 2001; Lievens, et al, 2005; Patterson et al, 2012)

• **Favourable applicant reactions** & high face validity
Validation studies

Study 1. Supervisor ratings after 1 year into training

Study 2. End-of-training outcomes (licensure exam) after 3 years

Study 3. Structural (theoretical) model to evaluate the incremental validity for each selection method
## Study 1: Correlations between the selection methods & job performance after 1 year

*Patterson et al, 2013 BJGP*

<table>
<thead>
<tr>
<th>N=196</th>
<th>Mean</th>
<th>SD</th>
<th>1.</th>
<th>2.</th>
<th>3.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Selection methods (Predictors)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Clinical problem-solving test</td>
<td>78.88</td>
<td>9.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Situational judgement test</td>
<td>637.87</td>
<td>34.31</td>
<td>.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Selection centre</td>
<td>3.32</td>
<td>0.39</td>
<td>.30</td>
<td>.43</td>
<td></td>
</tr>
<tr>
<td><strong>Outcome variable</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Supervisor ratings</td>
<td>4.63</td>
<td>0.73</td>
<td>.36 (0.54)</td>
<td>.37 (0.56)</td>
<td>.30 (0.50)</td>
</tr>
</tbody>
</table>

*Note.* Correlations between parentheses were corrected for multivariate range restriction. Correlations are significant at $p < .01$
Study 2: Correlations between the selection methods & end-of-training licensure assessments

Patterson et al, 2013, BJGP

<table>
<thead>
<tr>
<th>N=2292</th>
<th>Mean</th>
<th>SD</th>
<th>1.</th>
<th>2.</th>
<th>3.</th>
<th>4.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Selection methods (Predictors)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Clinical problem-solving test</td>
<td>80.08</td>
<td>8.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Situational judgement test</td>
<td>640.13</td>
<td>31.66</td>
<td>.40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Selection centre</td>
<td>3.34</td>
<td>0.36</td>
<td>.24</td>
<td>.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Outcome variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. End of training applied knowledge test</td>
<td>0.26</td>
<td>0.90</td>
<td>.73 (0.85)</td>
<td>.43 (0.69)</td>
<td>.24 (0.41)</td>
<td></td>
</tr>
<tr>
<td>5. End of training clinical skills exam (OSCE)</td>
<td>0.20</td>
<td>0.80</td>
<td>.38 (0.55)</td>
<td>.43 (0.57)</td>
<td>.32 (0.41)</td>
<td>.41</td>
</tr>
</tbody>
</table>

*Note. Correlations between parentheses were corrected for multivariate range restriction. Correlations are significant at p < .01*
Study 3. Structural model showing selection methods & the link to job performance

Lievens & Patterson, 2011, *J Applied Psychology*

- SJT fully mediates the effects of declarative knowledge in predicting job performance
- Selection centre adds significant predictive value over the selection tests
Case Study 2.

**Using SJTs for medical school admissions**

*Gateway 1*

Widening participation & scalability of selection

N= 26,000 per year for 8,000 posts

5 subtests

- Verbal, numerical, abstract reasoning & decision analysis
- SJT – targets empathy, integrity & team involvement

http://www.ukcat.ac.uk/
Test Specification

• An SJT for a novice population (no medical knowledge required)

Content

• Scenarios based in either a healthcare setting or during education/training for a medical/dental career
• Third party perspective

Response Format (rating using a 4 point scale)

• Rate the appropriateness of a response from ‘very appropriate’ to ‘very inappropriate.
• Rate the importance of a response from ‘very important’ to ‘not important at all’
A consultation is taking place between a senior doctor and a patient; a medical student is observing. The senior doctor tells the patient that he requires some blood tests to rule out a terminal disease. The senior doctor is called away urgently, leaving the medical student alone with the patient. The patient tells the student that he is worried he is going to die and asks the student what the blood tests will show.

How appropriate are each of the following responses by the medical student in this situation?

Q1 Explain to the patient that he is unable to comment on what the tests will show as he is a medical student

Q2 Acknowledge the patient’s concerns and ask whether he would like them to be raised with the senior doctor

Q3 Suggest to the patient that he poses these questions to the senior doctor when he returns

Q4 Tell the patient that he should not worry and that it is unlikely that he will die
UKCAT SJT Evaluation

- Reliability of a 70 item test with similar quality items estimated ($\alpha=.75$ to $.85$)
- Candidate reactions show high face validity (significantly more than the cognitive subsections of UKCAT)
  - Content of SJT relevant for med/dental applicants = 70%
  - Content of the SJT is fair to med/dental applicants = 63%
Group differences & content validity

- **SJT correlates with other subtests** (approx \( r=0.28 \)). Since a large amount of variance is not explained, the SJT is assessing different constructs to the other tests.

- **Gender:** Females outperformed males (0.2 SD)

- **Ethnicity:** White candidates performed better (0.3SD)

- **Occupation & Employment Status:** those in the higher occupational classes (i.e. Managerial/Professional Occupations) do not always score higher than those in lower classes - in some cases those from lowest occupational groups, received the highest mean score.
Implications

- **SJT**s can be designed to measure a variety of non-academic (beyond knowledge) important in clinical practice at various levels in the training pathway.
- Candidates choose answers from a set of predefined responses scored using a pre-determined key. This enables creation of a reliable, standardised, machine-markable test based on responses to a large number of scenarios.
- Strong predictive validity translates into significant gains in utility.
- Early evidence suggests a positive impact on widening access.
- Future research could evaluate additional aptitude domains (e.g., innovation potential).
Adjusting medical school admission: assessing interpersonal skills using situational judgement tests

Filip Lievens

CONTEXT Today’s formal medical school admission systems often include only cognitively oriented tests, although most medical school curricula emphasise both cognitive and non-cognitive factors. Situational judgement tests (SJT) may represent an innovative approach to the formal measurement of interpersonal skills in large groups of candidates in medical school admission processes. This study examined the validity of interpersonal video-based SJTs in relation to a variety of outcome measures.

RESULTS Interpersonal skills assessment carried out using SJTs had significant added value over cognitive tests for predicting interpersonal GPA throughout the curriculum, doctor performance, and performance on an OSCE and in a case-based interview. For the other outcomes, cognitive tests emerged as the better predictors. Females significantly outperformed males on the SJT ($d = -0.26$). The interpersonal SJT was perceived as significantly more job-related than the cognitive tests ($d = 0.55$).

CONCLUSIONS Video-based SJTs as measures of procedural knowledge about interpersonal behaviour show promise as complements to...
Evaluations of situational judgement tests to assess non-academic attributes in selection

Fiona Patterson,1,2 Victoria Ashworth,2 Lara Ziberras,3 Philippa Coan,2 Maire Kerrin2 & Paul O’Neill4

OBJECTIVES This paper presents a systematic review of the emerging international research evidence for the use of situational judgement tests (SJT) for testing important non-academic attributes (such as empathy, integrity and resilience) in selection processes.

METHODS Several databases (e.g. MEDLINE, PsycINFO, Web of Science) were searched to retrieve empirical studies relating to SJTs published between 1990 and 2010. Personal contact with experts in the field was made to identify any unpublished research or work in progress to obtain the most current material. Finally, reference lists were checked to access other relevant journal articles and further research. All research studies were required to meet specific inclusion criteria selected by two independent reviewers. A total of 39 articles that adequately met the inclusion criteria were included in the final review. The research evidence shows that, compared with personality and IQ tests, SJTs have good levels of reliability, predictive validity and incremental validity for testing a range of professional attributes, such as empathy and integrity.

CONCLUSIONS SJTs can be designed to test a broad range of non-academic constructs depending on the selection context. As a relatively low-fidelity assessment, SJTs are a cost-efficient methodology compared with high-fidelity assessments of non-academic attributes, such as those used in objective structured clinical examinations. In general, SJTs are found to demonstrate less adverse impact than IQ tests and are positively received by candidates. Further research is required to explore theoretical
Situational judgement tests represent a measurement method and can be designed to minimise coaching effects

Fiona Patterson,1,2 Victoria Ashworth,2 Maire Kerrin2 & Paul O’Neill3

Editor – Coaching is controversial in all selection contexts and we thank Rostom et al.1 for drawing attention to this because research in this area is lacking. Our review2 emphasises that situational judgement tests (SJT) are a measurement method – in which there is no single approach to design – and that each SJT should be evaluated individually regarding issues relating to coaching, validity, fairness, and so on. We highlight three key areas to inform discussion.

1 Situational judgement tests can be designed to be less susceptible, for example, the UK Foundation Programme4 SJT uses a more complex response format that employs two types of response (‘Rank all five possible responses in order’ and ‘Choose the three best responses from a list of eight’) that are significantly more cognitively loaded than those used in the response format in the Belgian test (‘Pick the best response from a choice of four’). Using complex, cognitively loaded formats is likely to ‘make SJTs less prone to coaching effects’.3

2 Research must examine whether SJT coaching produces genuine or artificial effects. Formal education and training in important domains (such as communication) could and, indeed, should be beneficial to

3
Designing Selection Systems for Medicine: The importance of balancing predictive and political validity in high-stakes selection contexts

Fiona Patterson*, Filip Lievens**, Máire Kerrin***, Lara Zibarras**** and Bernd Carette**

*Department of Psychology, University of Cambridge, Downing Street, Cambridge, CB2 3EB, UK. fcp27@cam.ac.uk
**University of Ghent, Ghent, Belgium
***Work Psychology Group Ltd, Derbyshire, UK
****City University, London, UK

Selection into medical education and training is a high-stakes process. A key unanswered issue is the effectiveness of measuring noncognitive predictors via both low-fidelity and high-fidelity selection approaches in this high-stakes context. We review studies investigating the effectiveness of multiple selection instruments in terms of predictive validity, incremental validity, and applicant reactions in both entry-level and advanced-level medical selection. Our results show that the situational judgment test (SJT) is the best single predictor of performance, operationalized in multiple ways. In addition, the low-fidelity SJT has incremental predictive power over cognitively oriented tests, and high-fidelity assessment center (AC) exercises add incremental validity over the low-fidelity (and less costly) selection methods. Concerning applicant reactions, results show that overall, the selection system is perceived as fair and valid.
Summary & conclusions

1. **Over-reliance on reliability estimates rather than validity**
   Focus less on ‘how much validity’ does a selection method add & more on ‘valid for what’?

2. **The ‘criterion problem’ – what are we trying to predict?**
   Use multiple **outcome measures** & pay closer attention to criterion constructs

3. **Few theoretically driven studies**
   **Structural modelling** can significantly enhance our understanding of how/why selection systems work & helps in addressing causality

4. **Designing efficient/scalable methods**
   Strong **predictive validity** translates into significant gains in **utility**. Computer-delivery is efficient, but **content updates** can be expensive

5. **Address widening participation**
   SJTs appear to be **less influenced by socio-economic status**
Thank you

f.patterson@workpsychologygroup.com

Some useful links

http://www.workpsychologygroup.com/
http://www.gprecruitment.org.uk/
http://www.isfp.org.uk/
http://www.ukcat.ac.uk/about-the-test/behavioural-test/
http://www.gmc-uk.org/about/research/14400.asp