The educational value of Situational Judgement Tests (SJTs) when used during undergraduate medical training: A systematic review and narrative synthesis

Gurvinder S Sahota[1], Victoria Fisher[1], Bakula Patel[1], Kiranjit JuJ[1], Jaspal S Taggar[1]

[1]University of Nottingham, UK

Dr Gurvinder Sahota: Clinical Associate Professor in Primary Care

Dr Bakula Patel: Clinical Associate Professor in Primary Care

Dr Kiranjit Juj: Research Associate, Primary Care

Dr Victoria Fisher: Research Associate, Primary Care

Dr Jaspal Taggar: Clinical Associate Professor and Head of Undergraduate Primary Care Education

Abstract

Introduction

Situational Judgement Tests (SJTs) are a recognised assessment method for admission into medical school, selection into postgraduate training programs, and postgraduate competency assessment. More recently, however, SJTs have been used during undergraduate medical training (UMT). This systematic review identifies, describes and appraises the evidence for SJTs in UMT to determine educational associations and outcomes.

Methods

MEDLINE, EMBASE, ERIC, PsycINFO, SCOPUS, Web of Science and grey literature were searched for original research studies evaluating SJTs implemented within UMT to 1st November 2022. Studies reporting evaluation outcomes were included. Narrative data syntheses were undertaken. Risk of Bias was appraised using the Quality in Prognosis Studies tool.

Results

24 studies were included. National database-derived SJTs (n=14) assessed against professionalism; postgraduate attainment; construct of medical degree; medical school admissions scores, personality attributes and declaration of disability. In-house derived SJTs (n=10) assessed against professionalism; clinical skills and personality attributes. Most evidence evaluated and reported inverse SJT associations with professionalism and were moderate risk of bias.

Conclusion

SJTs may have utility for developing professional behaviours in medical students. However, further research testing SJT robustness, standard setting methodologies, and prospectively evaluating SJTs against objective outcome measures within the context of UMT is warranted.

200 words

Keywords

Assessment, Undergraduate, Professionalism, Medicine

Introduction

Situational Judgement Tests (SJTs) are a major assessment method for healthcare professionals, being designed to test participant abilities to respond to challenging scenarios in plausible, role-relevant ways. There is a clear logic and evidence-base demonstrating the validity of SJTs in medical education and published 'gold-standard' processes for designing SJTs (Patterson et al. 2016).

Within medical education, the applicability of SJTs can be considered at different stages of a learners' educational journey, notably: selecting students into medical school; teaching and assessing competencies during undergraduate medical training (UMT); selecting undergraduates into postgraduate roles; and subsequent competency testing as medical practitioners.

The majority of SJT literature in medical education focusses on their utility for selection into undergraduate and postgraduate roles and the evaluation of postgraduate competencies. Evidence syntheses (Patterson et al. 2016; Webster et al. 2020) support the use of SJTs in high-stakes selection processes. For assessing postgraduate competencies there is also robust evidence SJTs can test procedural, ethical and legal constructs, demonstrating positive SJT score associations with other assessment metrics such as postgraduate educational supervisor ratings, clinical examination scores and medical simulation testing (Patterson et al. 2009; Lievens and Patterson 2011; Koczwara et al. 2012).

Despite the larger evidence-base for SJTs focussing on their use as selection tools and implementation of SJTs outside of UMT, the generalisability of this evidence into UMT is questionable. Given the rich learning environment and early professional identity formation within medical students, SJTs may have an important role in supporting educational development in UMT.

SJTs have good reliability and psychometric robustness in testing non-academic attributes of medical students during UMT (Husbands et al. 2015; Patterson et al. 2016; Goss et al. 2017). They have been used in formative and summative assessments,

varying in construct from free-text to multiple choice questions and video-based scenarios.

Recent evidence found SJTs may associate with the professionalism of medical students, suggesting other value beyond academic assessment of undergraduates in developing positive student behaviours. Sahota and Taggar (2020) found lower SJT scores were associated with increasing odds of having professionalism concerns identified during UMT. Smith and Tiffin (2018) also found higher SJT scores were associated with greater likelihood of successful foundation doctor programme completion.

However, there has been no systematic appraisal of evidence within UMT and the value of SJTs as an educational tool within this context remains unknown. Understanding such evidence would enable educators to better optimise SJT use and application within undergraduate curricula, whilst also identifying limitations of their use within UMT.

This review aimed to systematically identify, describe and appraise the published evidence of how SJTs have been used in UMT and to determine the educational associations and outcomes from using SJTs within this context. A secondary aim was to identify evidence gaps of using SJTs in UMT, thus enabling development of future research hypotheses.

Methods

Search strategy and selection criteria

This study was conducted in accordance with guidelines and methods for systematic reviews in medical education (Peters et al. 2015; Webster et al. 2020). MEDLINE, EMBASE, ERIC (Education Resources Information Center), PsycINFO, SCOPUS and Web of Science databases were searched from inception until 1st November 2022 (Appendix 1). The reference lists of national guidelines, review articles and included studies were also hand-searched to identify eligible studies. Grey literature, without restrictions, was searched through OpenGrey, EThOS and conference proceedings for 'situational

judgement test'. Authors of SJT studies and experts within this field were also contacted to identify any potentially relevant studies.

Studies were excluded prior to the year 2000 as SJTs are a relatively new phenomena in medical education which have developed significantly from 2002 onwards.

Inclusion and Exclusion criteria

The following selection criteria were used:

Inclusions:

1. Original research investigating SJTs delivered to students during UMT.
2. Studies including medical students.
3. Studies published/translated in English.
4. Studies of any methodological design.

Exclusions:

1. Studies investigating SJTs administered to students/trainees outside of undergraduate medical programmes
2. Studies prior to 2000.
3. Studies that did not report outcome data for SJT evaluation.

After database/reference searching and duplicate removal, two reviewers (GS/BP) independently screened citations for relevance and reviewed full-text articles using predetermined eligibility criteria. Any disagreements were resolved by consensus with a third reviewer (JT). Study selection was undertaken using Endnote X9.

Data extraction

Data were independently extracted using a pre-specified template (Appendix 2) by two reviewers (KJ/VF). Disagreements were resolved by discussion and consensus with a third reviewer (GS). The data extraction template was developed by reviewing extracted data from a published systematic review of SJTs. Data were extracted for study design, participant characteristics, educational constructs assessed by SJTs, description of SJT assessment(s) and standard setting methods used. Questions to ascertain risk of bias were added and the template was refined to ensure it captured relevant data for educational outcomes, consistent with review hypotheses. Educational associations and outcomes were primarily defined as outcomes related to student performance (knowledge, skills, attitudes and behaviours associated with SJTs with any comparator analysis). We also extracted data for any 'other associations' that did not align to student performance to enable inclusion of all outcome data.

Study quality and risk of bias

Consistent with previous research, data for study quality and risk of bias (RoB) were appraised using the Quality in Prognosis Studies (QUIPS) tool (Hayden et al. 2013; Webster et al. 2020). QUIPS assesses quality in six domains: study participation; study attrition; prognostic factor evaluation; outcome measurement; confounding; and statistical analysis & reporting (Appendix 4).

A scoring system was used to quantify RoB within included studies. QUIPS domains were as: reported (score=2); partially reported (score=1); not reported (score=0). Lower scores represented higher RoB. The number of domains within each study that were rated moderate and/or high risk within QUIPS were used to indicate overall RoB (Appendix 4).

Data synthesis

It was anticipated that quantitative synthesis would not be appropriate for this study. Data were therefore narratively summarised. Studies were stratified into two groups: studies evaluating SJTs derived from nationally recognised databases (e.g. UK Foundation Programme) and studies evaluating SJTs derived in-house for host institutions. These groups were chosen *a priori* to reduce heterogeneity within our findings as SJTs from national databases were likely to have greater robustness in design and are administered to larger groups of students than bespoke SJTs for individual organisations. This would improve translation and applicability of findings into educational practise.

Narrative synthesis of findings was undertaken according to emerging themes within the data extracted, determined through consensus by the study team.
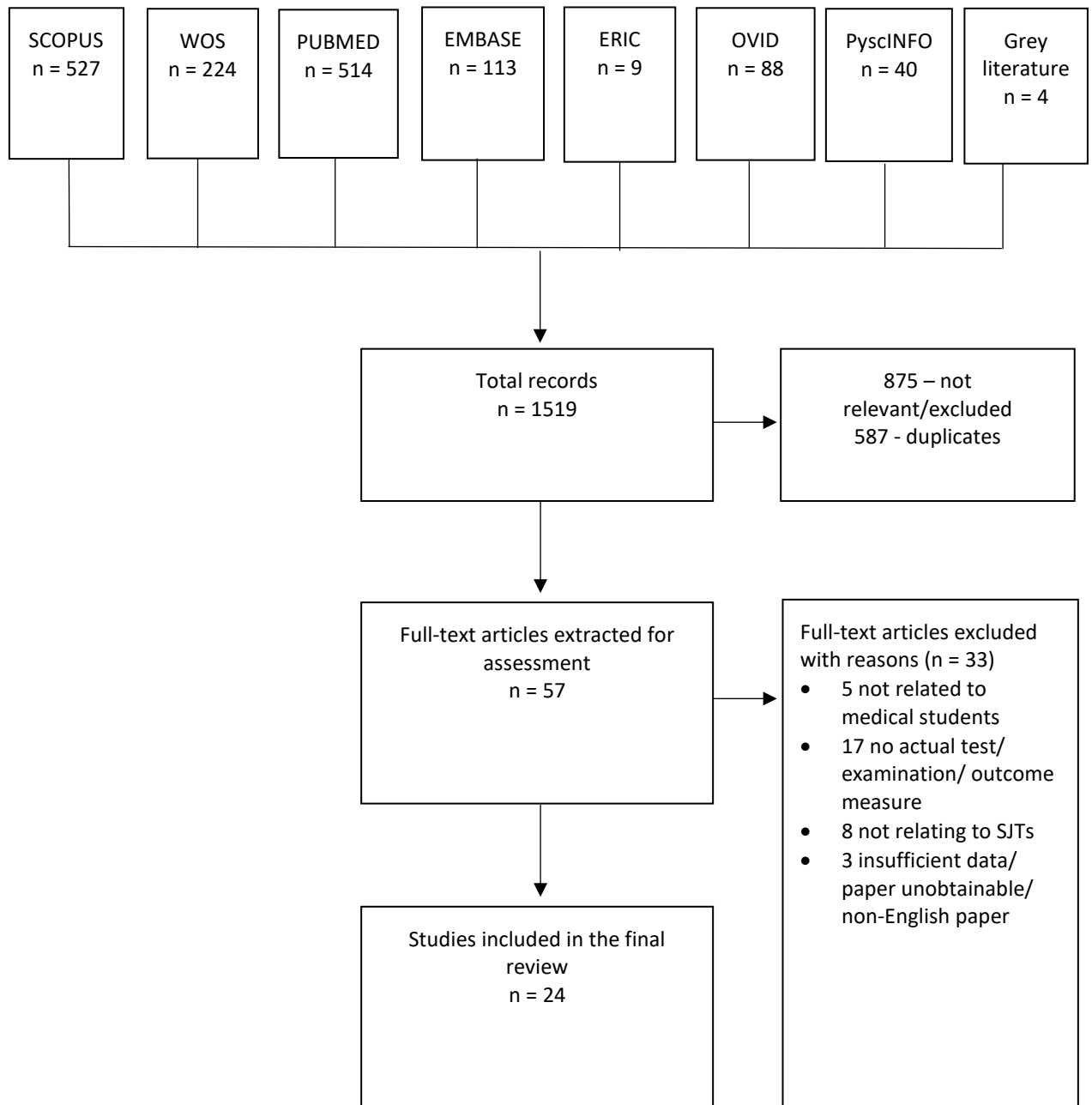
```
┌──────────┐ ┌──────────┐ ┌──────────┐ ┌──────────┐ ┌──────────┐ ┌──────────┐ ┌──────────┐ ┌──────────┐
│ SCOPUS   │ │ WOS      │ │ PUBMED   │ │ EMBASE   │ │ ERIC     │ │ OVID     │ │ PyscINFO │ │ Grey     │
│ n = 527  │ │ n = 224  │ │ n = 514  │ │ n = 113  │ │ n = 9    │ │ n = 88   │ │ n = 40   │ │ literature│
│          │ │          │ │          │ │          │ │          │ │          │ │          │ │ n = 4    │
└──────────┘ └──────────┘ └──────────┘ └──────────┘ └──────────┘ └──────────┘ └──────────┘ └──────────┘
```

Total records n = 1519	875 – not relevant/excluded 587 - duplicates

Full-text articles extracted for assessment n = 57	Full-text articles excluded with reasons (n = 33) • 5 not related to medical students • 17 no actual test/ examination/ outcome measure • 8 not relating to SJTs • 3 insufficient data/ paper unobtainable/ non-English paper

Studies included in the final review n = 24

Figure 1. PRISMA (preferred reporting for systematic reviews and meta-analysis

Results

A total of 1,519 papers were identified through literature searching. After duplicate removal and title and abstract review, 57 studies were retained. 33 further studies did not meet the selection criteria, leaving a final sample of 24 studies for analyses (Figure 1).

Of the 24 included studies, 14 evaluated SJTs derived from nationally recognised data sources and 10 evaluated SJTs derived in-house for host institutions (Figure 2).
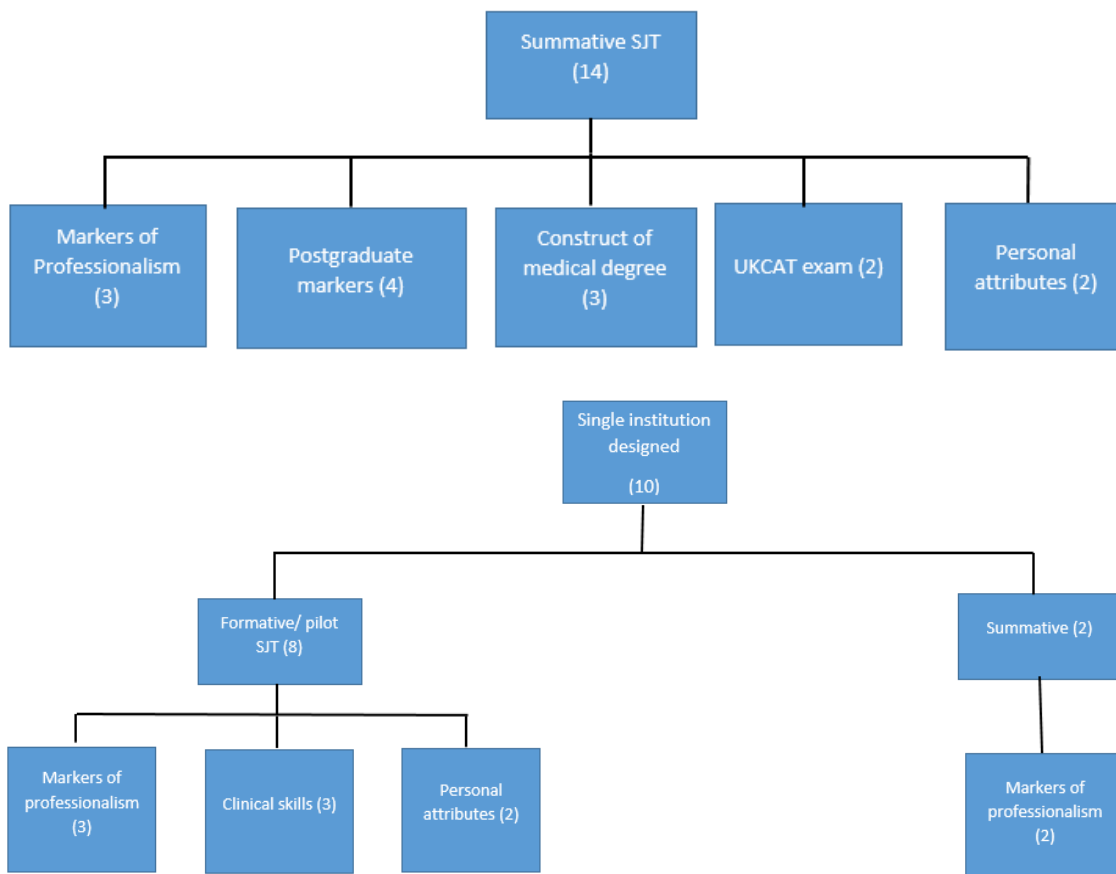

Figure 2. Flowcharts of constructs assessed against nationally derived and single institution derived SJT papers within review

SJTs derived using nationally recognised databases

Characteristics of SJT studies

All 14 studies of SJTs from nationally recognised databases (Table 1; Appendix 3) used questions from the United Kingdom Foundation Programme (UKFP).

All were UK-based and observational, consisting of longitudinal cohort (Devine et al. 2015; MacKenzie et al. 2017; Patterson et al. 2017; Garrud and McManus 2018; Paton et al. 2018; Smith and Tiffin 2018; Curtis and Smith 2020; Ellis et al. 2021; Murphy et al. 2022), prospective correlational (Sawdon and McLachlan 2020) and cross-sectional study designs (MacKenzie et al. 2016; Cousans et al. 2017; McManus et al. 2020; Sam et al. 2021).

Participant characteristics within SJT studies

All participants were final-year medical students. Across studies, 40-58% participants were female, mean ages ranged 18-24 years, and 54.2-74% of participants were Caucasian. Generally, sample sizes within these studies were larger than those evaluating in-house derived SJTs.

Description and construct of SJT assessments

The UKFP is a summative examination that ranks candidates on their overall score. This, combined with other assessment scores, such as the Educational Performance Measure (EPM), determines UK foundation doctor job allocations. The examination is undertaken during the final-year of undergraduate education, currently consisting of 75 questions in a 2.5 hour computer-based written assessment (UKFP 2021). Examinations use a combination of ranking questions (from 1 to 5) and choose best three answers from eight options.

The UKFP SJTs, were designed to test medical student attributes of being junior doctors. Construct domains of testing included: patient focus, professionalism, coping with pressure, effective communication and team working (UKFP 2021).

Standard setting for SJT assessments

Technical reports published by UKFP and Work Psychology Group (private enterprise commissioned to design UKFP examination) state mean and standard deviation scores by year (McLachlan and Illing 2013). There were no published descriptions of standard setting or cut score criteria used within analyses of included studies.

Educational associations and outcomes with SJTs

Included studies were further categorised in five sub-groups according to education outcomes and associations with SJT assessments (Figure 2).

These were studies investigating associations between SJT scores and markers of professionalism (Paton et al. 2018; Sawdon and McLachlan 2020; Sam et al. 2021); postgraduate attainment (Devine et al. 2015; Cousans et al. 2017; Smith and Tiffin 2018; Ellis et al. 2021); construct of the medical degree on SJT outcomes (Patterson et al. 2017; Curtis and Smith 2020; McManus et al. 2020); United Kingdom Clinical Aptitude Test (UKCAT) admissions scores (MacKenzie et al. 2016; Garrud and McManus 2018); and personal attributes (MacKenzie et al. 2017; Murphy et al. 2022).

Markers of professionalism: Three studies evaluated SJTs for associations with markers of professionalism and reported mixed findings. Paton (2018) found increased SJT scores were inversely associated with conduct related issue declarations (OR=0.82). Sawdon and McLachlan (2020) found lower SJT scores were associated with significantly lower Contentiousness Index scores (R=0.373, p<0.001). Conversely, Sam et al. (2021) investigated SJT associations with postgraduate disciplinary action in the first five years after qualification, concluding that there were no significant associations between the two.

Postgraduate attainment: Smith and Tiffin (2018) found greater UKFP SJT scores were associated with a greater likelihood of successful foundation doctor programme completion (OR=1.076, 95% CI 1.036-1.118, p<0.001) and that the EPM was positively

correlated with SJT scores (rho=0.30). In contrast, Devine et al. (2015) and Ellis (2021) found that SJTs did not correlate with postgraduate attainment. Cousans et al. (2017) demonstrated SJT and the EPM scores were positively correlated (r=0.46, p<0.01) and higher SJT scores were associated with higher supervisor ratings. Furthermore, those with lowest SJT scores were five-times more likely to have remedial action than those scoring the highest.

Construct of the medical degree on SJT outcomes: One study found standard entry medical students scored higher than those on gateway courses, however there was no difference in scores after controlling for academic attainment on entry (Curtis and Smith 2020). Another study found students from medical schools that primarily adopted problem-based learning scored lower on SJTs than students studying at non-problem-based learning institutions (McManus et al. 2020). A third study reported that medical school attended accounted for 4% SJT score variation once prior academic attainment had been accounted for (Patterson et al. 2017), suggesting the education institution studied may associate with student performance from SJTs.

UKCAT (pre-medical school) admissions scores: Mackenize et al. (2016) found higher UKCAT SJT scores were associated with higher UKFP SJT scores and female gender, greater affluency and age when sitting SJTs were associates with higher SJT scores. Garrud and McManus (2018) found increased UKCAT scores and female gender were associated with higher UKFP SJT score, with black and minority ethnic students scoring lower on the UKFP SJT than their peers.

Personal attributes: One study investigated personality constructs with SJT scores, finding students with higher aloofness and empathy scored lower in the SJT (MacKenzie et al. 2017). Another study reported SJT scores were lower in those that had a declared disability (Cohens d -0.019), however academic progression was similar for all students regardless of declaration. Patterson et al. (2017) reported no score differences for students with a disability.

SJTs derived in-house for host institutions

Characteristics of SJT studies

Of the 10 included studies (Table 2, Appendix 3), eight were formative or pilot study SJT assessments (Foucault et al. 2015; Kiessling et al. 2016; Frohlich et al. 2017; Goss et al. 2017; Antes et al. 2020; Graupe et al. 2020; Ludwig et al. 2021; Reiser et al. 2021). Two studies were from summative SJT assessments (Tiffin et al. 2011; Sahota and Taggar 2020).

All studies were observational in design, consisting of cohort (Foucault et al. 2015; Kiessling et al. 2016; Frohlich et al. 2017; Graupe et al. 2020; Ludwig et al. 2021) and cross-sectional studies (Tiffin et al. 2011; Goss et al. 2017; Antes et al. 2020; Sahota and Taggar 2020; Reiser et al. 2021). Most were conducted in non-UK settings; five in Germany (Kiessling et al. 2016; Frohlich et al. 2017; Graupe et al. 2020; Ludwig et al. 2021; Reiser et al. 2021), one in the United States (Antes et al. 2020), one in Canada (Foucault et al. 2015) and one in Australia (Goss et al. 2017). Two studies were UK-based (Tiffin et al. 2011; Sahota and Taggar 2020).

Participant characteristics within SJT studies

Participants ranged from first to final year medical students. Generally, sample sizes across studies were smaller than studies of SJTs from nationally recognised databases. Three studies did not report any student demographics (Tiffin et al. 2011; Foucault et al. 2015; Goss et al. 2017). Others consisted of student cohorts comprising 52-74% females and mean ages between 22-30 years. One study reported ethnicity with 61% students being white (Antes et al. 2020).

Description and construct of SJT assessments

SJT question style was broad for studies of in-house designed SJTs. Of these studies, five SJTs comprised ranking style questions (Frohlich et al. 2017; Goss et al. 2017; Graupe et

al. 2020; Sahota and Taggar 2020; Reiser et al. 2021), three used best answer questions (Tiffin et al. 2011; Goss et al. 2017; Antes et al. 2020) and one comprised extended matching questions (Tiffin et al. 2011). One study combined single-best answer and extended matching questions (Kiessling et al. 2016). One study (Ludwig et al. 2021) used multiple choice questions. The question style for one study was not reported (Foucault et al. 2015).

The methods of delivering SJT were video-based, (Frohlich et al. 2017; Graupe et al. 2020; Reiser et al. 2021) computer-based (Foucault et al. 2015; Kiessling et al. 2016; Sahota and Taggar 2020) and written assessments (Tiffin et al. 2011; Goss et al. 2017; Antes et al. 2020).

Standard setting for SJT assessments

One study provided information about standard setting methodology which was a modified-Angoff method (Goss et al. 2017).

Educational associations and outcomes with SJTs

The 10 studies were categorised into three sub-groups according to education outcomes and associations with SJTs (Figure 2). These were: studies investigating associations between SJT scores and makers of professionalism (Tiffin et al. 2011; Foucault et al. 2015; Goss et al. 2017; Antes et al. 2020; Sahota and Taggar 2020); clinical skills; (Kiessling et al. 2016; Ludwig et al. 2021; Reiser et al. 2021) and personal attributes (Frohlich et al. 2017; Graupe et al. 2020).

Markers of professionalism: Five studies assessed professionalism associations (Tiffin et al. 2011; Foucault et al. 2015; Goss et al. 2017; Antes et al. 2020; Sahota and Taggar 2020). Of these, two used SJTs as a mechanism to teach medical students about professional competencies (Foucault et al. 2015; Goss et al. 2017). Goss et al. (2017) found SJTs were a useful tool for teaching professionalism; 82% of students found SJTs relevant and 76% reported SJTs helped progress understanding of how to answer professionalism-related questions. Foucalt (2015) reported 87% of students found

professionalism-related SJTs helpful and realistic. Three studies evaluated SJTs for objective attainment of professionalism attributes. Sahota and Taggar (2020) found lower SJT scores were associated with an increased likelihood of professionalism lapses at medical school; for every point increase in SJT score students were 10% less likely to have multiple professionalism lapses. Antes et al. (2020) found SJT scores correlated with peer ratings of professionalism but not with clerkship ratings or honour and integrity scores. Tiffin et al. (2011) found no relationship between SJT scores and peer ratings of professionalism or with the Conscientiousness Index.

Clinical skills: Three studies assessed SJTs for developing communication skills (Kiessling et al. 2016; Ludwig et al. 2021; Reiser et al. 2021). One study demonstrated a moderate correlation between preadmission communication and SJTs scores (Kiessling et al. 2016). A second study reported SJT scores were significantly correlated with results from a communication skills OSCE (r=0.25, p≤0.01) (Ludwig et al. 2021). A video-based SJT study focussed on communication skills highlighted medical student's judged the test's usability to be high (*M*=3.60, *SD*=0.55), considered the test interesting (*M*=2.97, *SD* = 0.58), and were highly engaged while working on the test, as indicated by their reported effort (*M*=3.54, *SD*=0.47) (Reiser et al. 2021).

Personal attributes: Graupe et al. (2020) assessed relationships between SJT scores and emotional handling skills, finding conflicting evidence of associated emotional handling markers. Another study assessed SJTs with social competencies (Frohlich et al. 2017) finding applicants to medical school scored significantly higher in their study SJT than their current students. Personality traits of agreeableness, conscientiousness, extraversion and openness were correlated with higher SJT scoring and neuroticism was correlated with lower SJT scores.

Risk of Bias

Utilising the QUIPS tool half of the studies were graded moderate RoB. A summary of quality and RoB appraisal is provided in Appendix 3. The common areas for potential bias were identified as i) selection bias, mainly due to samples of convenience being taken or self-selected students engaging with the study; ii) sample sizes, arising from

single institution based SJT studies which often had small sample sizes; iii) confounding, where studies were not controlling for confounders such as academic attainment; iv) SJT reliability, which was either unreported or reported as poor reliability; v) sample population, which was not adequately described in studies.

Discussion

Summary of principal findings

This systematic review identified 24 studies utilising SJTs within UMT. Over half of studies utilised nationally developed SJTs from the UKFP all of which were implemented as summative assessments. The remaining studies utilised bespoke, in-house developed SJTs predominately as formative assessments.

Most research evaluated SJTs for the assessment of professionalism behaviours or attributes in students, suggesting inverse associations of SJT scores with unprofessionalism. Literature also suggests SJT performance may be associated with academic prowess and personal attributes. The evidence for other outcomes, such as developing clinical skills and later performance as post-graduates, was conflicting and heterogeneous.

Strengths and limitations

This is the first systematic review of evidence underpinning the use of SJTs delivered to students during UMT. A strength was use of a comprehensive search strategy, following published guidance for conducting systematic reviews, thus providing confidence all published literature was identified.

Due to the heterogeneity of evidence, quantitative analysis was not undertaken which limits the ability to derive precise outcomes from the review. Heterogeneity was reduced through stratification of studies into two groups according to SJT design. A narrative synthesis was conducted, enabling prominent themes to be described about how SJTs have been used and their associated educational outcomes during UMT.

An appraisal of study quality was undertaken which is a strength of using systematic review methodology improving robustness of the review. However, most studies were graded as moderate RoB, consistent with another published systematic review of SJTs in medical selection (Webster et al. 2020). Some studies also had relatively small sample sizes. Both factors limit the applicability and generalisability of findings into routine undergraduate educational practice.

Discussion of findings in the context of SJT literature

Most healthcare research underpinning SJTs evaluates their use for selecting students into UMT or postgraduate training. SJTs have increasingly been used as educational tools in UMT and most evidence from this review focusses on their associations with professionalism attributes, suggesting lower SJT scores may associate with greater professionalism concerns. Most research within this context is of lower quality and evaluates against perceived markers of professionalism, with few prospective studies investigating objective outcome measures of professionalism, such as recorded unprofessionalism, upheld complaints or fitness to practise proceedings (Paton et al. 2018; Sahota and Taggar 2020; Sam et al. 2021).

Antes et al. (2018), Foucault et al. (2015), and Goss et al. (2017) utilised SJTs as a learning tool for professionalism. Whilst studies highlight students find undertaking SJTs helpful to learning, there was no objective measurement for education attainment, such as knowledge or attitude acquisition.

Goss et al. (2017) used a hybrid model of feedback provision to students with lowest SJT scores. This type of formative individualised feedback, used as an adjunct to SJT delivery, may enhance professional development within UMT and is a consistent theme supported by several other studies (Foucault et al. 2015; Kiessling et al. 2016; Frohlich et al. 2017; Graupe et al. 2020). This is consistent with contemporary thinking about professional identity formation of students and this approach to using SJTs in UMT may enhance student learning.

This review found most SJTs are computer-based written examinations, with question styles either based on ranking responses or choosing single/multiple responses. Video-based simulation were used as part of three German based single institution studies (Frohlich et al. 2017; Graupe et al. 2020; Reiser et al. 2021). These studies found SJT scores were associated with development of positive personality attributes and further evaluation of delivering SJTs using this modality may be warranted. However, benefits of video-based SJTs may be offset by the high initial costs of scenario development (Graupe et al. 2020).

Greater SJT scores appear to be associated with higher student cognitive attainment. SJT scores tend to be negatively skewed, with restricted variability at higher ranges of marks (Curtis and Smith 2020). Several studies comment on the positive relationship and predictive ability of examinations preceding SJT implementation during UMT and test scores (Devine et al. 2015; Kiessling et al. 2016; MacKenzie et al. 2016; MacKenzie et al. 2017; Garrud and McManus 2018; McManus et al. 2020). The educational impact of this is likely to be modest as Mackenzie et al. (2016) also highlighted an increase in UKCAT score of 333 led to only 1 extra mark in the UKFP SJT. Our review supports these concepts and SJTs could be part of the 'academic backbone' of assessments (McManus et al. (2020).   SJTs may add extra value by testing something over and above cognitive tests alone, which is a growing view in recent SJT literature (Smith and Tiffin 2018; Webster et al. 2020).

Nearly all studies included in the review explicitly mentioned SJTs being designed with the use of curriculum or incident review, utilising subject matter experts for question design as recommended in Patterson's Gold Standards (Patterson et al. 2016). However, half of the studies of in-house derived SJTs made no explicit reference to piloting questions prior to the main study SJT which limits quality of SJT design and implementation.

Only one study stated a clear standard setting-mechanism (Goss et al. 2017), citing a modified Angoff approach. No clear rationale was provided for using this methodology. Standard setting has been described as "a process that allows human judgements to be synthesised in a rational and defensible way" (Champlain 2014). It's possible that

utilising subject matter experts to develop and design SJTs mitigates the need for formal standard setting processes and may explain why it wasn't reported in studies. However, the importance of cut scores within research was recently highlighted in a study investigating associations between SJT scores and professionalism concerns. Sahota and Taggar (2020) varied cut-scores from one to two standard deviations below the mean SJT score, demonstrating increased odds of students being identified with professionalism concerns (from 4 to 11-fold higher than those with higher SJT scores).

It was surprising, given the extent of literature on SJTs, that a limited number of studies were in the context of UMT. These studies were also mostly derived from UKFP-based SJTs, with the remainder as part of formative assessments. There was little evidence supporting the use of SJTs for educational attainment beyond developing professionalism in learners. Studies were generally of lower quality, opening debate about the transferability and utility of SJT evidence into UMT, whilst highlighting the need for more research in this area.

Conclusions

Evidence for the use of SJTs in UMT as an educational tool suggests value of SJTs for developing or assessing professionalism in students. However, higher-quality research that prospectively evaluates SJTs against objective measures of professional performance or behaviours is warranted. There is also a paucity of research evaluating the educational value of SJTs in UMT beyond developing behaviours and attributes, and research determining optimal standard setting methods for SJTs is also warranted.

Practice points

1. Most evidence for SJTs in UMT focuses on score associations with professionalism, and students with lower SJT scores are more likely to have unprofessional behaviours.
2. There is no consensus on standard setting methods for SJTs in UMT.
3. Further research that prospectively investigates the relationship between SJTs and objective occurrences of unprofessional behaviours would strengthen the evidence-base.

References

Antes AL, Dineen KK, Bakanas E, Zahrli T, Keune JD, Schuelke MJ, DuBois JM. 2020. Professional decision-making in medicine: Development of a new measure and preliminary evidence of validity [Article]. PLoS ONE. 15(2).

Cousans F, Patterson F, Edwards H, Walker K, McLachlan JC, Good D. 2017. Evaluating the complementary roles of an SJT and academic assessment for entry into clinical practice. Adv Health Sci Educ Theory Pract. 22(2):401-413. eng.

Curtis S, Smith D. 2020. A comparison of undergraduate outcomes for students from gateway courses and standard entry medicine courses. BMC medical education. 20(1):4. English.

Devine OP, Harborne AC, McManus IC. 2015. Assessment at UK medical schools varies substantially in volume, type and intensity and correlates with postgraduate attainment [Article]. BMC Medical Education. 15(1).

Ellis R, Scrimgeour DSG, Brennan PA, Lee AJ, Cleland J. 2021. Does performance at medical school predict success at the Intercollegiate Membership of the Royal College of Surgeons (MRCS) examination? A retrospective cohort study. BMJ Open. 11(8):e046615. eng.

Foucault A, Dubé S, Fernandez N, Gagnon R, Charlin B. 2015. Learning medical professionalism with the online concordance-of-judgment learning tool (CJLT): A pilot study. Med Teach. 37(10):955-960. eng.

Frohlich M, Kahmann J, Kadmon M. 2017. Development and psychometric examination of a German video-based situational judgment test for social competencies in medical school applicants. International Journal of Selection and Assessment. 25(1):94-110.

Garrud P, McManus IC. 2018. Impact of accelerated, graduate-entry medicine courses: a comparison of profile, success, and specialty destination between graduate entrants to accelerated or standard medicine courses in UK. BMC medical education. 18(1):250. English.

Goss BD, Ryan AT, Waring J, Judd T, Chiavaroli NG, O'Brien RC, Trumble SC, McColl GJ. 2017. Beyond Selection: The Use of Situational Judgement Tests in the Teaching and Assessment of Professionalism. Academic medicine : journal of the Association of American Medical Colleges. 92(6):780-784. English.

Graupe T, Fischer MR, Strijbos JW, Kiessling C. 2020. Development and piloting of a Situational Judgement Test for emotion-handling skills using the Verona Coding Definitions of Emotional Sequences (VR-CoDES). Patient Education and Counseling. English.

Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C. 2013. Assessing bias in studies of prognostic factors. Annals of internal medicine. 158(4):280-286. eng.

Husbands A, Rodgerson MJ, Dowell J, Patterson F. 2015. Evaluating the validity of an integrity-based situational judgement test for medical school admissions. Bmc Medical Education. 15.

Kiessling C, Bauer J, Gartmeier M, Iblher P, Karsten G, Kiesewetter J, Moeller GE, Wiesbeck A, Zupanic M, Fischer MR. 2016. Development and validation of a computer-based situational judgement test to assess medical students' communication skills in the field of shared decision making. Patient Education and Counseling. 99(11):1858-1864. English.

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, Patterson F. 2011. The Validity and Incremental Validity of Knowledge Tests, Low-Fidelity Simulations, and High-Fidelity Simulations for Predicting Job Performance in Advanced-Level High-Stakes Selection. Journal of Applied Psychology. 96(5):927-940.

Ludwig S, Behling L, Schmidt U, Fischbeck S. 2021. Development and testing of a summative video-based e-examination in relation to an OSCE for measuring communication-related factual and procedural knowledge of medical students. GMS J Med Educ. 38(3):Doc70. eng.

MacKenzie R, Dowell J, Ayansina D, Cleland J, MacKenzie RK, Cleland JA. 2017. Do personality traits assessed on medical school admission predict exit performance? A UK-wide longitudinal cohort study. Advances in Health Sciences Education. 22(2):365-385.

MacKenzie RK, Cleland JA, Ayansina D, Nicholson S. 2016. Does the UKCAT predict performance on exit from medical school? A national cohort study. BMJ open. 6(10):e011313. English.

McLachlan J, Illing J. 2013. Health Education England report: An independent review of events relating to the use of situational judgement tests in the selection for foundation process 2013. Available from: https://www.medschools.ac.uk/media/1932/hee-independent-review-of-the-sjt.pdf; [accessed 18/01/18].

McManus IC, Harborne AC, Horsfall HL, Joseph T, Smith DT, Marshall-Andon T, Samuels R, Kearsley JW, Abbas N, Baig H et al. 2020. Exploring UK medical school differences: the MedDifs study of selection, teaching, student and F1 perceptions, postgraduate outcomes and fitness to practise. Bmc Medicine. 18(1).

Murphy MJ, Dowell JS, Smith DT. 2022. Factors associated with declaration of disability in medical students and junior doctors, and the association of declared disability with academic performance: observational study using data from the UK Medical Education Database, 2002–2018 (UKMED54). BMJ Open. 12(4). English.

Paton LW, Tiffin PA, Smith D, Dowell JS, Mwandigha LM. 2018. Predictors of fitness to practise declarations in UK medical undergraduates [Article]. BMC Medical Education. 18(1).

Patterson F, Ashworth V, Baron H, Thomas S, Smith D. 2017. The contribution of "Value-add" methodology to understanding medical school outcomes. UKMED: Work Psychology Group, General medical Council, University of Bristol.

Patterson F, Baron H, Carr V, Plint S, Lane P. 2009. Evaluation of three short-listing methodologies for selection into postgraduate training in general practice. Medical education. 43(1):50-57.

Patterson F, Zibarras L, Ashworth V. 2016. Situational judgement tests in medical education and training: Research, theory and practice: AMEE Guide No. 100. Medical Teacher. 38(1):3-17.

Peters MD, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. 2015. Guidance for conducting systematic scoping reviews. International journal of evidence-based healthcare. 13(3):141-146. eng.

Reiser S, Schacht L, Thomm E, Bauer J, Figalist C, Janssen L, Schick K, Berberat PO, Gartmeier M, Dorfler E. 2021. A video-based situational judgement test of medical students'

communication competence in patient encounters: Development and first evaluation. Patient Education and Counseling. English.

Sahota GS, Taggar JS. 2020. The association between Situational Judgement Test (SJT) scores and professionalism concerns in undergraduate medical education. Medical teacher.(42:8):937-943. English.

Sam AH, Bala L, Westacott RJ, Brown C. 2021. Is Academic Attainment or Situational Judgment Test Performance in Medical School Associated With the Likelihood of Disciplinary Action? A National Retrospective Cohort Study. Acad Med. 96(10):1467-1475. eng.

Sawdon M, McLachlan JC. 2020. '10% of your medical students will cause 90% of your problems': a prospective correlational study. BMJ Open. 10(11):e038472. eng.

Smith DT, Tiffin PA. 2018. Evaluating the validity of the selection measures used for the UK's foundation medical training programme: A national cohort study. BMJ Open. 8(7):e021918. English.

Tiffin PA, Finn GM, McLachlan JC. 2011. Evaluating professionalism in medical undergraduates using selected response questions: findings from an item response modelling study. BMC medical education. 11:43. English.

UKFP FP. 2021. Situational Judgement Test (SJT) FAQs. [accessed 2021 15 December]. https://foundationprogramme.nhs.uk/faqs/situational-judgement-test-sjt-faqs/.

Webster ES, Paton LW, Crampton PES, Tiffin PA. 2020. Situational judgement test validity for selection: A systematic review and meta-analysis. Medical education. 54(10):888-902. eng.