Exploring the validity of the 2013 UKCAT SJT- prediction of undergraduate performance in the first year of medical school: Summary Version of Report

March 2017

Paul A. Tiffin & Lewis W. Paton, Health Sciences, University of York

Executive Summary

Background: Previously it has been shown that the scores from some forms of the UKCAT 2013 SJT predict tutor ratings in medical students. In addition, the cognitive scale scores of the UKCAT tend to predict undergraduate performance on both *theory* and *skills*-based assessments in medical school. In these analyses we model the relationship between such undergraduate academic performance measures and the 2013 SJT scores.

Methods: SJT scores were available on 25, 679 applicants who sat the UKCAT in 2013, of which undergraduate outcomes were available for approximately 1,400 admitted students from eight medical schools. A series of univariable and multivariable regression analyses were conducted to evaluate the relationship between the predictor (SJT scores) and the outcomes of interest.

Results: Two of the five SJT forms showed a statistically significant relationship with *theory* scores and one of the forms significantly predicted *skills*-based undergraduate performance, with relatively low effective sizes. The SJT scores independently predicted *theory* but not *skills* scores after controlling for potential confounding variables. However, scores from form five of the 2013 SJT were independent predictors for both *theory* and *skills* performance. The SJT scores predicted the odds of passing the year at first attempt, though this effect was not significant once A-level performance was controlled for.

Conclusions: The six forms of the 2013 UKCAT SJT have differing relationships with external constructs. A continued focus on further work to create comparable forms of the tests, and confirm these properties, is important. In this case the stronger relationship with *theory* rather than *skills*-based performance suggest that it may be mediating variables (such as conscientiousness) that explain these observed relationships with undergraduate academic performance.

Background

A previous pilot study demonstrated that the 2013 UKCAT Situational Judgement Test (SJT) scores were significantly related to tutor ratings of medical students for *integrity*, *team involvement* and *perspective taking* (Patterson, Edwards et al. 2015, Patterson, Cousans et al. 2017). A subsequent analysis observed that the different forms of the UKCAT 2013 SJT showed variation in their relationship to the validity criteria (Tiffin and Paton 2016). Previously we have demonstrated the predictive validity of the cognitive scales of the UKCAT in relation to undergraduate medical performance on *theory* and *skills* tests (Tiffin, Mwandigha et al. 2016). In this present study we similarly model the relationship between the UKCAT SJT scores and undergraduate performance. It is

hoped that the findings from such analyses will help further understand how the SJT scores are working, guide test development and inform how their role might be optimised within medical selection.

Methods

This study utilised data from medical applicants who took the UKCAT (including the SJT) in 2013. Eight medical schools supplied data on progression at year one in this cohort (Table 1). Predictor scores were standardised as z scores (mean 0, sd 1). Irish and Scottish Higher qualifications were not included. Demographic variables were dichotomised in line with previous methodology (Tiffin, Dowell et al. 2012). Mature entrants were defined as those entering medical school at 21 years of age or older. Analyses were conducted using listwise deletion with no imputation for missing variables. *Theory* and *skills* scores were standardised within medical school. However, we also used this study as an opportunity to demonstrate a method by which such local undergraduate measures can be rescaled to be more comparable nationally (we refer to this as '*peer-competition rescaling*'). Where the intraclass correlation across medical schools was zero, or very close to 0 (i.e. when clustering effects were negligible) single, rather than multilevel, models were used for both linear and logistic regression Multivariable models were built up using a forward stepwise regression approach with a p value of 0.1 as the criteria for entry of a variable. These models were only built up where data were complete for all the variables included, and so models were then tested out on the wider dataset where complete data was only required in the final variables included. Potential interactions between variables in the final models were also evaluated for.

Continuous (standardised) variables						
Variable	N	Mean	SD	Min	Max	Missing (%)
SJT score	25,679	0	1	-5.34	3.82	0 (0)
UKCAT total score	25,679	0	1	-4.39	2.83	0 (0)
A-level performance	11,405	0	1	-4.78	1.34	14,274 (55.6%)
Theory	1,383	0	1	-3.94	2.64	[24,296*, 94.61%]
Skills	1,381	0	1	-3.52	2.64	[24,298*, 94.62%]
Binary variables						
Variable	Proportion					Missing (%)
Male	10,853 (42.2%)					0
BME	8,810 (44.7%)					5965 (23.2)
Non-selective school	6073 (53.5%)					14324 (55.8)
Non-professional background	1399 (6.1%)					21403 (11.2)
EASL	7000 (28.35%)					988 (3.85)
Degree	7079 (32.01)					3564 (13.88)
>21 years	7973 (31.05%)					5 (0.02)
End of year outcome (pass)	1,371 (13.06%)					24,308* (94.67)

Table 1. Descriptive statistics for students in the study cohort.

Note: *Missingness will be due to failure to get into medical school, participation status of university etc.

Results

The results from the univariable regression analyses can be seen in Tables 2 to 5. As can be seen from Table 2 there is a significant ability of the SJT scores to predict standardised *theory* performance in year one. The regression coefficient of 0.13 would be interpreted as follows; on

average, for every one standard deviation scored on the situational judgement test above the mean for that applicant cohort the student, on average (assuming medical school entry), would score .13 of a standard deviation above the mean on the year one *theory* exams for his medical school year peer group.

Predictor variable	Regression coefficient	p	Lower 95% CI	Upper 95% CI
SJT score	0.13	<0.01	0.07	0.19
Degree	0.26	<0.01	0.12	0.40
EASL	-0.10	0.14	-0.24	0.03
Non-professional background	-0.29	0.05	-0.58	0.00
Male	-0.03	0.52	-0.14	0.07
BME	-0.25	<0.01	-0.37	-0.13
>21 years	0.29	<0.01	0.16	0.42
Non-selective schooling	0.06	0.39	-0.07	0.19
UKCAT Abstract Reasoning	-0.03	0.31	-0.09	0.03
UKCAT Decision Analysis	0.17	<0.01	0.09	0.25
UKCAT Quantitative Reasoning	0.16	<0.01	0.10	0.23
UKCAT Verbal Reasoning	0.14	<0.01	0.08	0.20
UKCAT Total	0.17	<0.01	0.10	0.25
A-Level performance	0.69	<0.01	0.57	0.82

Table 2. Results from univariable regression analyses, predicting standardised *theory* scores from a number of demographic and educational factors. Note that UKCAT, SJT scores and A-level attainment are all standardised as z scores (mean=0, SD=1) according to the performance of 2013 applicants. Note: EASL- 'English as a second language'.

The ability of the SJT scores to predict performance in *skills* is weaker than that compared to the prediction of *theory* scores. Nevertheless, although the coefficient is small at 0.08, it is statistically significant at the p less than 0.05 level (Table 3).

Predictor variable	Regression coefficient	p	Lower 95% CI	Upper 95% CI
SJT score	0.08	0.01	0.02	0.14
Degree	0.16	0.03	0.02	0.30
EASL	-0.14	0.05	-0.28	0.00
Non-professional background	-0.35	0.02	-0.64	-0.06
Male	-0.18	<0.01	-0.28	-0.07
BME	-0.27	<0.01	-0.39	-0.16
>21 years	0.21	<0.01	0.08	0.34
Non-selective schooling	0.07	0.30	-0.06	0.19
UKCAT Abstract Reasoning	0.00	0.90	-0.05	0.06
UKCAT Decision Analysis	0.15	<0.01	0.07	0.23
UKCAT Quantitative Reasoning	0.08	0.02	0.02	0.15
UKCAT Verbal Reasoning	0.04	0.21	-0.02	0.09
UKCAT Total	0.10	0.01	0.02	0.18
A-Level performance	0.36	<0.01	0.24	0.48

Table 3. Results from univariable regression analyses, predicting standardised *skills* scores from a number of demographic and educational factors.

The observed correlation between the undergraduate outcome measures (*theory* and *skills*) and the SJT scores will be subject to (indirect) range restriction. This is because the outcomes are not observed in those applicants who had not entered the involved medical schools. In such cases, especially where the restriction of range in the selector variable is known, then the Thorndike case III method can be applied attempt to correct for such attenuation (Alexander 1990). The uncorrected correlation between the standardised SJT scores and the standardised *theory* performance scores was **0.1155** (p<.0001). The corrected correlation was **0.1396**. The uncorrected correlation between the standardised SJT scores and the standardised *skills* performance scores was **0.0683** (p<.0001). The corrected correlation was **0.0799**. Imperfect test reliability will also have further attenuated the correlations to some degree.

Tutor ratings (used in the earlier validation pilot study) predict both *theory* and *skills* scores (p≤0.02 in all cases). For *theory* the standardised beta coefficients were 0.18 to 0.27; for *skills* 0.30 to 0.39. It is particularly noteworthy that the relationship with *skills* is stronger than *theory*, as might be expected if the tutors were rating aspects of interpersonal functioning.

As previous analysis showed that the forms of the SJT could not be considered to have equivalent validity in relation to external constructs, an analysis was conducted by form. As can be seen from Table 4 only two of the forms used in the 2013 SJT had a statistically significant relationship with *theory* performance. These were forms four and five. Only the latter, form five, had a significant ability to predict *skills* score (Table 4 also).

	Theory			
Predictor variable	Regression coefficient	p	Lower 95% CI	Upper 95% CI
Form 1	0.09	0.21	-0.05	0.22
Form 2	0.07	0.31	-0.07	0.21
Form 3	0.13	0.16	-0.05	0.31
Form 4	0.17	0.01	0.04	0.31
Form 5	0.25	<0.01	0.09	0.41
Form 6	0.13	0.12	-0.03	0.29
	Skills			
Predictor variable	Regression coefficient	p	Lower 95% CI	Upper 95% CI
Form 1	-0.04	0.54	-0.18	0.10
Form 2	0.03	0.62	-0.10	0.17
Form 3	0.16	0.09	-0.02	0.34
Form 4	0.08	0.23	-0.05	0.22
Form 5	0.23	<0.01	0.08	0.37
Form 6	0.16	0.07	-0.01	0.33

Table 4. Univariable standardised *theory* and *skills* performance prediction by SJT form.

The results of the multivariable linear regressions are shown in Table 5. As can be seen, even after controlling for the effects of potential confounding variables, the SJT scores had a significant ability to predict *theory* performance. However this was not the case for the prediction of *skills*. It can also be seen from Table 5 that there was a significant interaction between BME status and male sex in relation to *theory* prediction. This term was positive in direction and therefore it could be

interpreted that being male and of BME status (combined) tended to offset the disadvantage, in terms of year one *theory* performance, of being either male or BME ethnic status separately.

In order to further understand the mediators for the relationship between SJT scores and undergraduate performance, a multivariable model was constructed with the UKCAT total score as the only other variable. As can be seen from the results in Table 6, the SJT scores remained significant predictors of *theory* performance even after adjustment for cognitive ability, as estimated by the total UKCAT scores, though somewhat diminished in magnitude. A similar picture was seen for the prediction of *skills* performance (Table 7).

	Theory			
Predictor variable	Regression coefficient	p	Lower 95% CI	Upper 95% CI
Male	-0.15	0.05	-0.31	>0.00
BME	-0.32	<0.01	-0.51	-0.14
BME/Male interaction	0.26	0.04	0.01	0.52
SJT score	0.11	0.01	0.03	0.18
A-level performance	0.14	<0.01	0.61	0.87
UKCAT total	-0.11	0.04	-0.22	<0.00
	Skills			
Predictor variable	Regression coefficient	p	Lower 95% CI	Upper 95% CI
SJT score	0.01	0.90	-0.07	0.08
A-level performance	0.37	<0.01	0.24	0.50
Male	-0.25	<0.01	-0.37	-0.12
BME	-0.29	<0.01	-0.43	-0.16

Table 5. Multivariable linear regression results for the prediction of *theory* and *skills* performance. Note a significant interaction term between male sex and BME status for *theory* prediction. No interaction terms were significant at the p<0.05 level for *skills*.

Predictor variable	Regression coefficient	p	Lower 95% CI	Upper 95% CI
UKCAT total	0.14	<0.001	0.06	0.22
SJT score	0.11	<0.001	0.05	0.17

Table 6. Multivariable linear regression for the prediction of *theory* performance, controlling for cognitive performance, as estimated by the UKCAT total score. N=1383.

Predictor variable	Regression coefficient	p	Lower 95% CI	Upper 95% CI
UKCAT total	0.08	0.04	0.004	0.16
SJT score	0.07	0.04	0.002	0.12

Table 7. Multivariable linear regression for the prediction of *skills* performance, controlling for cognitive performance as estimated by the UKCAT total score. N=1381.

It was noted that, in applicants, the cognitive UKCAT scores tended to correlate from a low (r=0.26, AR) to a moderate degree (r=0.45, VR). However, as can be seen in Table 6, the relationship between SJT scores and *theory* performance remains significant despite controlling for cognitive ability, as estimated by the total UKCAT score, though is slightly diminished in magnitude. Thus, it can be

inferred that classical cognitive ability, as tested by the UKCAT, plays some role in mediating the relationship with *theory* scores, but does not fully explain the relationship.

Regarding the prediction of passing year one at first sitting; the results are shown for the univariable analysis in Table 8. A random intercept across medical schools was used, to allow for any differences across medical schools in the odds of passing year one. As can be seen from Table 8, overall the SJT scores statistically significantly predicted the odds of passing year one at first attempt. The odds ratio in this case was 1.28. This would be interpreted as follows; for every one standard deviation above the mean for applicants scored on the UKCAT SJT, the odds of passing first time (as opposed to another, less desirable academic outcome) will increase by around 28%. The non-linear nature of the odds ratios must be borne in mind, and this estimate would not hold for more extreme SJT scores. This value is comparable to that produced by the cognitive scores of the UKCAT, where the odds ratio was 1.32. However, as might be expected, A-level performance outperforms both these UKCAT scores with an odds ratio of 2.83. The results from a multivariable model are shown in Table 9. According to the stepwise regression, the only variable that remained in the multivariable model with an independent and statistically significant ability to predict pass at year one was A-level performance. If we put this into a model with the SJT scores we can see that there is still a trend for the SJT scores to predict pass at first time but that it is now not statistically significant with a p value of 0.12

Predictor Variable	Odds ratio	p	Lower 95% CI	Upper 95% CI
SJT score	1.28	0.01	1.07	1.54
Degree	1.12	0.62	0.71	1.75
EASL	0.87	0.48	0.58	1.30
Non-professional background	0.57	0.13	0.28	1.19
Male	0.78	0.13	0.57	1.08
BME	0.65	0.02	0.45	0.94
>21 years	1.07	0.73	0.72	1.60
Non-selective schooling	1.23	0.30	0.83	1.83
UKCAT Abstract Reasoning	1.09	0.34	0.91	1.31
UKCAT Decision Analysis	1.30	0.03	1.03	1.64
UKCAT Quantitative Reasoning	1.15	0.20	0.93	1.42
UKCAT Verbal Reasoning	1.12	0.24	0.93	1.34
UKCAT Total	1.32	0.03	1.03	1.68
A-Level performance	2.83	<0.01	1.83	4.37

Table 8. Results from multilevel (random intercept) univariable logistic regressions for the prediction of a pass at first sitting (versus another academic outcome) from the SJT scores. Note: N varies depending on the missing data for each variable (range N=909 to 1371).

Predictor variable	Odds ratio	p	Lower 95% CI	Upper 95% CI
A-level performance	2.79	<0.01	1.81	4.31
SJT score	1.22	0.12	0.95	1.57

Table 9. Multivariable multilevel logistic regression for the prediction of a pass at first sitting (versus another academic outcome) from the SJT scores. N=889.

As previous research has shown that the situation judgement tests tend to yield maximal information on candidates at the lower end of ability (i.e. they are better at discriminating between poorer candidates, rather than better candidates (Tiffin and Carter 2013)) we dichotomised *theory* and *skills* scores into 'poor' and 'non-poor' performers, using a Z score of -1 as the cut-off. We noted that prediction of a low *theory* score (z<-1) from the SJT performance was possible (OR 0.75, 0.63 to 0.88, p<.001) but not poor performance on *skills* (z<-1) (OR 0.93, 0.79 to 1.10, p=0.4). However, the effect on *theory* performance was diminished if A-level achievement was controlled for (adjusted OR 0.82, 0.66 to 1.02, p=.08).

Finally, we also attempted to rescale the undergraduate *theory* scores to make them nationally comparable, to illustrate a potential approach to 'nationalising' local outcomes. This is because using a local measure as an outcome variable for a national predictor may result in an underestimate of the strength of the relationship, even when using multi-level modelling. In this case we did this by looking at the mean and SD of the total UKCAT score (compared to all applicants) for each medical school. We then rescaled the *theory* scores by adding a mean of the UKCAT score for each specific medical school cohort. We then divided the *theory* scores by the SD of the UKCAT total score, specific to that medical school year. Thus, the standardisation which occurred within each medical school year was adjusted for by how well a student's peers performed on the UKCAT. As can be seen from Table 10 this generally increased the strength of the relationship between predictor (SJT) and the outcome (*theory* score). Whilst this is a useful way of weighting such localised scores there are two potential problems with it. Firstly, admissions deans would generally want to know how well a student would do in their particular medical school when entering, based on the selection method measures. Thus national performance is of less importance in this context. Secondly there is a risk of overly inflating the relationship between the UKCAT test scores and the *theory* scores, as the rescaling is informed by the former. Thus this may be seen as something of a statistical tautology. Moreover, where UKCAT scores are used as a selection criterion the scores would be lower (left) censored. Nevertheless, it has been previously shown that performance at national standardised postgraduate exams (such as the MRCP) are highly correlated with performance prior to medical school, such as A-levels (McManus, Elder et al. 2008) (with r values ranging from .7 to .8). Thus, 'peer-competition rescaling' would result in regression and correlation coefficients of the magnitude that are likely to comparable to those seen once the UKCAT is used to predict performance on national, standardised exams, as opposed to localised measures (such as undergraduate performance or the educational performance measure-EPM). Such an approach could theoretically also be applied to the EPM to 'nationalise' them. Here we crudely adjusted local medical school performance z scores for UKCAT performance but it is possible to use a more sophisticated approach where national measures (such as A level and UKCAT scores) that correlate with local undergraduate performance (e.g. EPM) are used as indicators of a latent ability, which is estimated, and the local measure rescaled accordingly.

SJT score	0.22	<0.01	0.10	0.33
Degree	0.41	<0.01	0.14	0.68
EASL	-0.15	0.28	-0.41	0.68
Non-professional background	-0.66	0.02	-1.21	-0.11
Male	-0.02	0.81	-0.23	0.18
BME	-0.49	<0.01	-0.73	-0.26
>21 years	0.49	<0.01	0.24	0.74
Non-selective schooling	0.12	0.35	-0.13	0.37
UKCAT Abstract Reasoning	0.01	0.88	-0.11	0.13
UKCAT Decision Analysis	0.35	<0.01	0.19	0.52
UKCAT Quantitative Reasoning	0.39	<0.01	0.25	0.52
UKCAT Verbal Reasoning	0.30	<0.01	0.18	0.41
UKCAT Total	0.50	<0.01	0.33	0.66
A-Level performance	1.55	<0.01	1.31	1.80

Table 10. Results from univariable regression analyses, prediction standardised and 'nationalised' *theory* scores from a number of demographic and educational factors. In order to 'nationalise' the *theory* scores across medical schools they were rescaled according to the distribution of UKCAT scores for that cohort (*'peer-competition rescaling')*.

Discussion

The present findings were in keeping with those recently reported, in that the different forms of the 2013 UKCAT SJT demonstrate different relationships with external 'validity' criteria (Tiffin and Paton 2016). Previously, in terms of tutor ratings, we noted that form 2, and to an extent form 1, of the SJT, were those with the closest relationship between scores and supervisor ratings. However, in terms of prediction of undergraduate performance we found that it was scores from form 5, and to an extent form 4 of the SJT that were most predictive of undergraduate performance. Indeed, it could be considered that it was form 5 that was doing most of the 'heavy lifting' and that it was the relationship between these scores that was mainly accounting for the predictive validity of the SJT overall. It was interesting to note that the relationship between *theory* scores and the SJT scores was stronger than that with *skills*. This is somewhat counterintuitive, as skills-based assessments would generally be considered to be more sensitive to the "social cognition" performance that is presumably evaluated via SJTs. This observation could be explained in a number of ways. Firstly SJT performance may be related to certain constructs such as conscientiousness, which is required for performance in undergraduate work. SJTs will also tap into conscientiousness with those candidates who sought out practice examples, tending to improve their performance. Previously we have shown that undergraduate medical exams that tap into semantic knowledge may be more sensitive to conscientiousness than items that ask around professionalism (Tiffin, Finn et al. 2011). In the light of this, given the likely content of *theory* based exams in the first year of medical school, this finding is not so surprising. Moreover, it is unclear what is actually being measured by exams declared as "skills" by medical schools. This lack of a reliable outcome could also affect the ability to show an association within the predictors. Moreover, there is a moderate correlation between verbal reasoning ability and SJT performance and this is more likely to affect the ability of a student to perform on the *theory* rather than *skills* based evaluation. Overall, SJT performance had some ability

to predict performance at both *theory* and *skills*, but it was only in the case of *theory* that this effect was independent of other, potentially confounding factors. It was also noted that overall SJT performance had the ability to predict the odds of passing the first year of medical school at first attempt, although this effect was no longer statistically significant once the effects of advanced qualifications, in the form of A-level performance, were controlled for.

Potential strengths and limitations

This study utilised a relatively large dataset. Thus, even when analysing the outcomes according to each of the six forms of the situation judgement test there was probably adequate power to detect effects of 'educational significance'. In this case we consider the major limitation of the study the lack of standardisation of the outcome measures (i.e. *theory* and *skills* scores). However, this study also provide an opportunity to evaluate a novel method for "nationalising" local undergraduate performance scores, by rescaling them according to the performance on the UKCAT of entrants that medical school that year. This tended to increase the magnitude of the coefficients for the predictors, unsurprisingly. This principle could be extended and applied, in the future, to the EPM in order to produce similar 'peer-competition rescaled' values.

Main implications for policy and further development of the SJTs

As with the previous report on SJT scoring, this study has highlighted the challenge and importance of achieving *true equating* between different forms of the SJT. Moreover, the results highlight the importance of performing analyses by different forms of the test in order to elicit the properties of the SJT in relation to external constructs (e.g. validity criterion). How the SJT scores should relate to such external criteria is not fully decided. Ideally the test scores should predict future performance in clinical practice (or high fidelity simulation). However, such outcome data may not be available for several years, given the length of medical training. Nevertheless, emerging data relating to the use of SJTs later on in medical selection suggest some moderate relationship with outcomes such as clinical performance in post-graduate exams. It is not clear whether this will also hold true for the SJTs for undergraduate selection as the test formats and content will be somewhat different. Moreover it may be more difficult to test an understanding of professionalism in applicants to medical school where this construct may not had time to develop and form fully through exposure to clinical practice. In particular it is interesting to note that forms 4 and 5 of the SJT tended to predict *theory* performance, whereas these were very weakly related to tutor ratings in the pilot validity study. Thus different forms of the SJT may be picking up different attributes in candidates. When considering the implications for test development it may be more desirable to build content around form 2, which related more strongly to tutor ratings, than those forms which had scores at related more to academic performance. Certainly further research should focus on what the optimum validity criteria should be, against which SJT scores are evaluated in this context. Indeed, it may be desirable that the scores have relatively little relationship with undergraduate performance, especially in the preclinical years of training. Rather, the scores should predict performance in more clinically orientated settings and the degree those that relate to interpersonal functioning (for example, indicated by professionalism issues, such as fitness to practice referrals).

As noted in our previous, recent report, there are several options for addressing the formidable challenge of achieving a true equating between the different forms of the SJT. Certainly identifying factors which relate more to supervisor ratings and building and sharing content relating to these

items within and between forms is more likely to result in test forms that behave in a similar way in relation to external criteria. If a robust enough validity criteria could be found, in theory, machine learning approaches could be applied to side-step the equating issue (i.e. a machine would observe a pattern of SJT response and classify a candidate according to a predictive algorithm).

Summary and conclusions

These analyses have demonstrated that the SJT scores have some ability to predict undergraduate performance in terms of *theory* and *skills* based assessments. This relationship is statistically significant only for forms 4 and 5 of the 2013 UKCAT, and only for form 5 in relation to *skills* assessments. Test development should focus on achieving test forms that behave in a more similar way in relation to external criteria (i.e. roughly equivalent validity) and further building test content, based on emerging knowledge, that relates external criteria of interest.

References

Alexander, R. A. (1990). "Correction Formulas for Correlations Restricted by Selection on an Unmeasured Variable." Journal of Educational Measurement **27**: 187-189.

McManus, I., A. Elder, A. De Champlain, J. Dacre, J. Mollon and L. Chis (2008). "Graduates of different UK medical schools show substantial differences in performance on MRCP(UK) Part 1, Part 2 and PACES examinations." BMC Med **6**: 5.

Patterson, F., F. Cousans, H. Edwards, A. Rosselli, S. Nicholson and B. Wright (2017). "The Predictive Validity of a Text-Based Situational Judgment Test in Undergraduate Medical and Dental School Admissions." Acad Med.

Patterson, F., H. Edwards, A. Rosselli and F. Cousins (2015). UKCAT SJT Predictive Validity Study Summary Report. Derby, UK., Work Psychology Group.

Tiffin, P. A. and M. Carter (2013). Understanding the measurement model of the UKCAT Situational Judgment Test: Summary Report, The UKCAT Board.

Tiffin, P. A., J. S. Dowell and J. C. McLachlan (2012). "Widening access to UK medical education for under-represented socioeconomic groups: modelling the impact of the UKCAT in the 2009 cohort." BMJ **344**: e1805.

Tiffin, P. A., G. M. Finn and J. C. McLachlan (2011). "Evaluating professionalism in medical undergraduates using selected response questions: findings from an item response modelling study." BMC Medical Education **11**: 43.

Tiffin, P. A., L. M. Mwandigha, L. W. Paton, H. Hesselgreaves, J. C. McLachlan, G. M. Finn and A. S. Kasim (2016). "Predictive validity of the UKCAT for medical school undergraduate performance: a national prospective cohort study." BMC Medicine **14**(1): 140.

Tiffin, P. A. and L. Paton (2016). Report to the UKCAT board: enhancing the scoring of the situation judgement test component of the UKCAT; reanalysis of the validity pilot study data. University of York, UK Clinical Aptitude Test Board.