THE PURPOSE OF SITUATIONAL JUDGEMENT TESTS

Situational Judgement Tests (SJTs) are designed to assess applicants' judgement regarding complex, real-life situations encountered in the workplace. Candidates are presented with a set of hypothetical work-based scenarios and asked to make judgements about possible responses (e.g., how appropriate each of the responses may be in the context of the scenario). These scenarios and lists of responses are typically developed by subject matter experts, who also determine the scoring keys for each scenario.

The nature of SJTs requires applicants to place themselves in situations that they are likely to encounter. This can allow applicants' responses to be greatly representative of how they would act in a real-life scenario (Motowidlo, Hooper, & Jackson, 2006). At high levels of selection, SJTs can consist of scenarios that are highly context and situation specific. This allows for these SJTs to potentially be highly ecologically valid measures of behaviour and knowledge.

SJTs used in the selection of medical professionals are designed to assess applicant attributes important for medical education, training, and practice that are not assessed by knowledge tests. Recent research indicates that SJTs are most effective as tools to screen outlow-scoring applicants (Cousan et al., 2017; Tiffin & Carter, 2015). In conjunction with knowledge-based assessments, SJTs can provide an effective, low-cost method of screening out applicants who are deemed not yet ready to progress to selection interviews for speciality placement positions. Tests are typically presented in a computerised format that is able to be delivered to large-scale cohorts with minimal supervision and assistance required.

SJTs are currently used to select undergraduate and graduate entry medical students, medical residents, and registrars for specialisations such as general practice, obstetrics, and gynaecology across Australia, Canada, and the United Kingdom. The attributes assessed by SJTs may differ depending on what traits are considered most important and desirable for successful applicants. Commonly, assessed traits include leadership skills (Christian, Edwards, & Bradley, 2010; Hay & Henry, 2017), communication skills (Lievens & Sackett, 2007; Lievens, Buyse, & Sackett, 2005), integrity (Hay & Henry, 2017; Husbands, Rodgerson, Dowell, & Patterson, 2015), and empathy (Hay & Henry, 2017). Typically, an SJT will either exclusively assess one trait, or consist of individual sets of scenarios or items to target each attribute of interest. In medical selection, medical professionals and educators will review SJTsprior to test piloting to ensure the scenarios and items appear highly content valid measures ofthe attributes they purport to assess.

SJTs can be used to assess applicants' typical or maximal performance, depending on the response instructions given. Knowledge instructions (i.e., enquiring what should be done in a situation) have been found to correlate more highly with cognitive tests, while behavioural tendency instructions (i.e., enquiring what the applicant would do in the situation) tend to correlate more with personality traits (McDaniels, Hartman, Whetzel, & Grubb, 2007). SJTs with knowledge instructions are by less fakeable, as applicants tend to respond to the best of

their ability regardless of the pressure to fake (Hooper et al., 2006). However, SJTs with behavioural tendency instructions have nonetheless been found less fakeable than personality tests (Hooper et al., 2006).

SJTs are typically preferred by test-takers over cognitive tests (Hay, Tran, Lichtwark, & Hodgson, 2016; Lievens, 2013). Both video and written SJTs have been reported by medical applicants to have significantly higher face validity than many other selection tools such as cognitive tests and assessment of work samples (Koczwara et al., 2012; Lievens & Sackett, 2006). A recent study into a written SJT developed by Hay and colleagues (2016) found most medical applicants could relate to the scenarios in the SJT, and nearly all felt the scenarios were relevant for medical course selection.

SJTs are reported to be highly predictive of performance outcomes relating to the attributes they assess (Lievens & Patterson, 2011; Lievens & Sackett, 2012; O'Connell et al., 2007; Patterson, Baron, Carr, Plint, & Lane, 2009). In general practitioner selection in the UK, an SJT was found to be the best single predictor of supervisor appraisals of job performance for junior doctors in their first year of practice (Lievens & Patterson, 2011). A meta-analysis by McDaniel and colleagues (2007) provided evidence that SJTs provide incremental validity overcognitive ability in selection. In addition, SJTs been found to have good test-retest and split- half reliability (Bruce & Leaner, 1958; Nguyen, Biderman, & McDaniel, 2005; Ployhart, Porr, & Ryan, 2004).

There is some evidence that SJTs are less impacted by socio-economic and cultural bias than other selection tools (Lievens, Patterson, Corstjens, Martin, & Nicholson, 2016; McDaniel et al., 2007; Patterson et al., 2009). Supplementing cognitive tests and other traditional selectiontools with SJTs is suggested to produce sizeable reductions in the racial inequality of entrancescores (Ployhart & Holtz, 2008). However, this does not preclude SJTs from bias, as recent findings from Hay, Lichtwark, Metcalf, and Henry (2017) indicate that applicants who are olderand have English as a second language may show reduced scores and be negatively impacted by test time. As fairness and diversity are key underpinning policy principles for many organisations, taking into account any known potential biases and working to increase equity is key.

Coaching has been found to have little effect on SJT scores (Cullen, Sackett, & Lievens, 2006), as the strategies needed to ascertain the correct responses are more complex than those of other tests. The literature on fakeability is mixed, however one study has found someforms of response instructions for SJTs to not be significantly affected by faking, and that evenother SJT forms may be less affected than other tests that traditionally assess those domains (Hooper et al., 2006).

To summarise, SJTs may represent a low-cost method of assessing desirable characteristics and attributes beyond knowledge, used around the world for medical selection. There is evidence to suggest they have high face validity, predictive validity, and reliability, and that they are potentially less impacted by socioeconomic and cultural bias, coaching, and applicants faking their responses than other traditional methods of assessment.

REFERENCES

Bruce, M. M., & Learner, D. B. (1958). A supervisory practices test. *Personnel Psychology,11*(2), 207-216.

Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology*, *63*(1), 83-117.

Cousans, F., Patterson, F., Edwards, H., Walker, K., McLachlan, J. C., & Good, D. (2017). Evaluating the complementary roles of an SJT and academic assessment for entry into clinical practice. *Advances in Health Sciences Education*, *22*(2), 401-413.

Cullen, M. J., Sackett, P. R., & Lievens, F. (2006). Threats to the operational use of situational judgment tests in the college admission process. *International Journal of Selection and Assessment*, *14*(2), 142-155.

Hay, M. & Henry, S. (2017). *Development and implementation of SJTs for health sciences selection – preliminary findings and current challenges*. Workshop session. Developing and Researching the Economics and Mathematics of Selection (DREAMS): Global Perspectives. Situational Judgement Tests Best Practice Program. Worldwide Universities Network. Sydney, Australia, July 24-25.

Hay, M., Lichtwark, I., Metcalf, J., & Henry, S. (2017). *The influence of first language, gender and age on Situational Judgement Test Scores (SJT) in nursing interns*. Research presentation. AMEE Conference, Innovating in Education. Helsinki, Finland, August 26-30.

Hay, M., Tran, S., Lichtwark, I., Hodgson, W. (2016). *Undergraduate medical course applicants' ratings of the value of a Situational Judgment Test (SJT) as a selection tool*. Oral presentation. AMEE Conference, Innovating in Education. Barcelona, Spain, August 27-29.

Hooper, A. C., Cullen, M. J., & Sackett, P. R. (2006). Operational threats to the use of SJTs: Faking, coaching, and retesting issues. *Situational Judgment Tests: Theory, Measurement, and Application*, 205-232.

Husbands, A., Rodgerson, M. J., Dowell, J., & Patterson, F. (2015). Evaluating the validity of an integrity-based situational judgement test for medical school admissions. *BMC Medical Education*, *15*(1), 144.

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

Lievens, F., Buyse, T., & Sackett, P. R. (2005). Retest effects in operational selection settings: Development and test of a framework. *Personnel Psychology*, *58*(4), 981-1007.

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.

Lievens, F., Patterson, F., Corstjens, J., Martin, S., & Nicholson, S. (2016). Widening access in selection using situational judgement tests: Evidence from the UKCAT. *Medical Education*, *50*(6), 624-636.

Lievens, F., & Sackett, P. R. (2006). Video-based versus written situational judgment tests: A comparison in terms of predictive validity. *Journal of Applied Psychology*, *91*(5), 1181.

Lievens, F., & Sackett, P. R. (2007). Situational judgment tests in high-stakes settings: Issuesand strategies with generating alternate forms. *Journal of Applied Psychology*, *92*(4), 1043.

Lievens, F., & Sackett, P. R. (2012). The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance. *Journal of Applied Psychology*, *97*(2), 460.

McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology*, *60*(1), 63-91.

Motowidlo, S. J., Hooper, A. C., & Jackson, H. L. (2006). A theoretical basis for situational judgement tests. In J. A. Weekley & R. E. Ployhart (Eds.), *SIOP organizational series. Situational judgment tests: Theory, measurement, and application* (pp. 57-81). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Nguyen, N. T., Biderman, M. D., & McDaniel, M. A. (2005). Effects of response instructions on faking a situational judgment test. *International Journal of Selection and Assessment*, *13*(4), 250-260.

O'Connell, M. S., Hartman, N. S., McDaniel, M. A., Grubb, W. L., & Lawrence, A. (2007). Incremental validity of situational judgment tests for task and contextual job performance. *International Journal of Selection and Assessment*, *15*(1), 19-29.

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.

Ployhart, R. E. & Holtz, B. C. (2008). The diversity-validity dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology*, *61*(1), 153-172.

Ployhart, R.E., Porr, W. and Ryan, A.M. (2004). *New development in SJTs: Scoring, coaching and incremental validity*. Paper presented at the 19th annual convention of the Society for Industrial and Organizational Psychology, Chicago, IL, April.

Tiffin, P. & Carter, M. (2015). *Understanding the measurement model of the UKCAT situational judgement test: Summary report*. UK Clinical Aptitude Test 2015.