Have admissions committees considered all the evidence?

Kent Hecker¹ · Geoff Norman²

Received: 23 December 2016 / Accepted: 26 December 2016 / Published online: 27 March 2017
© Springer Science+Business Media Dordrecht 2016

If you are looking for passionate debates, join your admissions committee. With the possible exception of curriculum committees, you will never engage in more heated, potentially emotionally charged, discussions than those regarding the important characteristics of applicants that we should be selecting for in the health professions. If you read the transcription carefully though, you will notice that the terms “informed-” and “evidence-based-” debates were not included. That’s because too often these discussions (unfortunately) are not informed by research. This absence is particularly notable when dealing with the selection of non-academic attributes, where the debate frequently reduces to vacuous statements like, “I know what it takes to make a great... (insert your favourite health professional)”. The trouble is, if s/he knows, s/he’s not telling. It’s almost impossible to get agreement on what characteristics should be selected for, let alone identifying reliable and valid strategies to measure these attributes.

There are three general domains selected for at the time of admissions: general academic achievement, general aptitude for medical school/medicine, and non-academic attributes/achievement. General academic achievement is used ubiquitously in admissions criteria; this is previous academic ability, whether at the high school or undergraduate level. General aptitude is typically assessed through standardized entrance exams (MCAT, HPAT, UKAT etc.)—more on this and its utility later. Identifying non-academic attributes/achievements is by far the most challenging. As measurement folk and admissions committee members recognize, we are tasked with implementing best practices to objectively measure something that is inherently subjective—then justify the results.

✉ Kent Hecker
kghecker@ucalgary.ca
Geoff Norman
norman@mcmaster.ca

¹ University of Calgary, Calgary, AB, Canada

² McMaster University, Hamilton, ON, Canada

Given these challenges, this special edition, devoted primarily to addressing non-academic attributes, contains thought provoking research and opinion pieces that all admissions committee members should review in order to engage in more informed debate.

As selection criteria evolve to meet societal demands, it has been evident for some time that academic ability is necessary but not sufficient for success/competence in school and beyond. Medical schools have increasingly adopted various methods to assess attributes other than academic ability. While the methods have been adopted—structured interviews like multiple mini-interviews (MMI) and Situational Judgement Tests (SJT), continuing research is required to address three overarching questions: What attributes should we be measuring at the time of admissions? How best should we measure them? And, most importantly, given the importance of admissions, is there enough evidence to support the use of the scores for admissions purposes?

What should we be measuring? In other words, “What non-academic dimensions’ matter?” Whether called non-cognitive abilities, non-academic attributes, personality traits, personal traits or any other permutation, every health professions program is tasked with selecting for non-academic attributes. Unlike academic achievement, which can be blueprinted into content domains like biology, statistics or history, it is exquisitely difficult to determine precisely what each test of non-academic attribute is assessing. It is one thing to say “moral reasoning” or “communication skills”; it is quite another to assemble evidence that this is indeed what any test measures. Indeed tests like the MMI are less tests of a specific characteristic, and more a test format adaptable to assessing many different attributes. And that regrettably leaves the specific attribute in the eye of the beholder—the author of the specific test.

Whatever the attribute, a further complication is social desirability. It is hard to fake knowledge of endocrinology; it is a lot easier to fake a particular characteristic deemed desirable by a personality test, whether as a component of a self-reported test (as in an SJT or personality inventory) or within an interview setting (such as an MMI). These concerns are not hypothetical. Within the personality psychology and selection literature, studies have reported between 22 and 50% of job applicants increased their scores at the time of application. The stated result that, “...applicant faking behavior resulted in significant rank ordering changes that impacted hiring decisions”, strikes an eerily cautionary tone for those using personality inventories as selection criteria to the health professions (Donovan et al. 2014; Griffith et al. 2007).

How best should we measure these attributes? Well, a close to perfect admissions process to assess non-academic attributes (barring observation of each and every applicant in their natural setting) would include the use of an expert interviewer who asks the same questions in the same environment using the same scoring system for all the non-academic attributes of interest. Is this feasible? Of course not. So, developments in interviewing methods and offsite non-interview based methods are meant to provide glimpses of applicant’s ability to reason and humanitarian orientation. These methods (or any method, for that matter), if not developed and implemented using best practices, could be fraught with issues. Although the literature may suggest scores from an MMI or SJT have good reliability and validity in other settings, a successful implementation requires continuing effort to blueprint the process, to develop scoring keys, to train the raters/interviewers and then to combine scores to determine an optimal total score. An example of this is the article by De Leng et al. in this issue (2016) on scoring method of a SJT (De Leng et al. 2016) where, depending on scoring method, the reliability coefficient (internal consistency) varied between .33 and .73. The interpretation and usefulness then of these scores take on very different meanings.

How then can we address the thorny issue of what, precisely, is being assessed by a particular instrument? We must move beyond the exploration of admissions data in terms of the standard dimensions of reliability and validity to theory testing, and truly begin to examine the congruence between the test scores and underlying theoretical models. While bivariate correlation matrices and regression models are important, what is required is the development of *a priori* models outlining relationships between academic, aptitude, and non-academic attributes and future performance indicators. One example of the difficulty with theory—free measurement is the repeated call for the use of personality tests (Powis 2015). While psychologists have converged on five dimensions (Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism) no one has specified what personality traits are important for what kinds of professional (Kulasegaram et al. 2010). Perhaps just as well, since the only stable characteristic that emerges is that conscientiousness predicts grades. But then we can get grades from the registrar, so the question arises regarding the utility of measuring conscientiousness.

Hypothetical models have been suggested and reported (Collin et al. 2009; Ferguson et al. 2003). Collin et al. explored the relationships between “academic ability” (grade point average), “aptitude for medicine” (MCAT scores) and “competence in medicine” (USMLE Step 1–3) using combined data from the Association of American Medical Colleges and the National Board of Medical Examiners in the United States of America (total n was 250,792 of which a random sample of 10% was used $n = 20,714$). This latent variable path analysis showed that “aptitude for medicine” was found to be a significant predictor for “competence in medicine” with a path coefficient of .55 accounting for 30% of the variance. This finding supports recent reports stating that general mental abilities significantly predict for future performance, not only in medical education but other professional realms (Kreiter and Otaki 2015).

Of course, a critic will immediately point out that this analysis perpetrates the traditional “grades predict grades” models, and says nothing about the non-cognitive domains. But these models can be expanded to include non-academic attribute scores, and begin to close the gap between cognitive and non-cognitive variables used in selection, just as some research is now investigating the relation between emotions and cognitive learning (McConnell and Eva 2012; LeBlanc et al. 2015). While some may state there are “limitations” of psychometric theory and measurement models (Dore et al. 2016), it is an important piece of the validity puzzle to analyze and report upon.

This leads to questions regarding the weight (emphasis) that schools put on these scores for the admission process, is it a bar that is reached that moves them into the next level? Is it worth 75% of their admissions score? How are these strategies justified and where is the rationale? While little evidence is provided on how schools calculate their admission scores, we in theory could work backwards to test the effect of various weightings or cutscores on predicting subsequent success.

Or can we? Given the non-academic attributes of interest, where there appears to a further gap in research is the exploration of whether scores from the various selection methods assess the same or different attributes. In other words, if we think we are selecting for a number of non-academic attributes or traits why then, in virtually all admissions selection research reports, only the total score (total score of the MMI, total score of the SJT, total score of the CASPer) is included in the validation study? Where is the evidence to support claims that multiple attributes are being assessed? Shouldn’t there be convergent and discriminant evidence demonstrating the utility of the attributes being measured? Indeed Oliver et al. (Oliver et al. 2014) tried to “draw a line in the sand” and suggest that if we are measuring multiple attributes using an MMI, then we should attempt to show some

evidence of differential prediction from the various sub-components. While some evidence of convergent and discriminant validity was provided, the correlation between communication and problem evaluation was large (.87). This is perhaps not surprising, if you are comfortable with a topic you should be better at communicating your thoughts and ideas. However, an alternative explanation is that we really cannot assess much more than communication skills in MMIs with stations where applicants are meant to verbally respond to various scenarios.

As Kreiter indicates in this issue (2016), there is still much (thoughtful) research to be done in medical education regarding the utility of non-academic attributes scores from various selection methods (Kreiter 2016). Is there enough evidence to support the use of non-academic attribute scores in the selection process? If, so, how much should they be worth in the process? That is likely a decision that rests with each individual institution, so is up to the admissions committee to deliberate upon and justify to their respective faculty and stakeholders. Does the time and cost of developing and using these methods justify the results? That is a matter of more research and definitely more “informed” and “evidence-based” debate.

References

- Collin, V. T., Violato, C., & Hecker, K. (2009). Aptitude, achievement and competence in medicine: A latent variable path model. *Advances in Health Sciences Education Theory Practice*, 14(3), 355–366. doi:[10.1007/s10459-008-9121-7](https://doi.org/10.1007/s10459-008-9121-7).
- De Leng, W. E., Stegers-Jager, K. M., Husbands, A., Dowell, J. S., Born, M. P., & Themmen, A. P. N. (2016). Scoring method of a Situational Judgment Test: Influence on internal consistency reliability, adverse impact and correlation with personality? *Advances in Health Sciences Education*. doi:[10.1007/s10459-016-9720-7](https://doi.org/10.1007/s10459-016-9720-7). (this issue).
- Donovan, J. J., Dwight, S. A., & Schneider, D. (2014). The impact of applicant faking on selection measures, hiring decisions, and employee performance. *Journal of Business and Psychology*, 29(3), 479–493. doi:[10.1007/s10869-013-9318-5](https://doi.org/10.1007/s10869-013-9318-5).
- Dore, K. L., Roberts, C., & Wright, S. (2016). Widening perspectives: reframing the way we research selection. *Advances in Health Sciences Education*. doi:[10.1007/s10459-016-9730-5](https://doi.org/10.1007/s10459-016-9730-5). (this issue).
- Ferguson, E., James, D., O’Hehir, F., & Sanders, A. (2003). Pilot study of the roles of personality, references, and personal statements in relation to performance over the five years of a medical degree. *British Medical Journal*, 326, 429–432.
- Griffith, R. L., Chmielowski, T., & Yoshita, Y. (2007). Do applicants fake? An examination of the frequency of applicant faking behavior. *Personnel Review*, 36(3), 341–355. doi:[10.1108/00483480710731310](https://doi.org/10.1108/00483480710731310).
- Kreiter, C. D. (2016). A research agenda for establishing the validity of non-academic assessments of medical school applicants. *Advances in Health Sciences Education*, 21(5), 1081–1085. doi:[10.1007/s10459-016-9672-y](https://doi.org/10.1007/s10459-016-9672-y). (this issue).
- Kreiter, C., & Otaki, J. (2015). Constructing a more comprehensive validity argument for medical school admission testing: Predicting long-term outcomes. *Teaching and Learning in Medicine*, 27(2), 197–200. doi:[10.1080/10401334.2015.1025016](https://doi.org/10.1080/10401334.2015.1025016).
- Kulasegaram, K., Reiter, H. I., Wiesner, W., Hackett, R. D., & Norman, G. R. (2010). Non-association between Neo-5 personality tests and the multiple mini interview. *Advances in Health Sciences Education*, 15(3), 415–423.
- LeBlanc, V. R., McConnell, M. M., & Monteiro, S. D. (2015). Predictable chaos: A review of the effects of emotions on attention, memory and decision making. *Advances in Health Sciences Education*, 20(1), 265–282.
- McConnell, M. M., & Eva, K. W. (2012). The role of emotion in the learning and transfer of clinical skills and knowledge. *Academic Medicine*, 87(10), 1316–1322.
- Oliver, T., Hecker, K., Hausdorf, P. A., & Conlon, P. (2014). Validating MMI scores: Are we measuring multiple attributes? *Advances in Health Sciences Education*, 19(3), 379–392. doi:[10.1007/s10459-013-9480-6](https://doi.org/10.1007/s10459-013-9480-6).
- Powis, D. (2015). Selecting medical students. *Medical Teacher*, 37, 252–260.