Faking good: self-enhancement by medical school applicants during assessment of their personal qualities.

David Powis, Miles Bore and Don Munro
School of Psychology, University of Newcastle, NSW 2320, Australia
(correspondence: Professor David Powis, david.powis@newcastle.edu.au)

Currently there is a considerable desire by medical schools in many countries to extend the criteria upon which their students are selected. Hitherto the selection of medical students has been informed mainly by prior academic achievement, moderated sometimes by the assessment of their personal qualities by interview. More recently assessment of applicants' cognitive skills by written tests has been included in the selection repertoire, and there has been serious consideration of the desirability of including also the assessment of non-cognitive or personality measures by written tests (e.g. UK Clinical Aptitude Test, www.ukcat.ac.uk). Future doctors would thus be selected for training on the basis of personal qualities known to be important in performing adequately and competently in their chosen profession, as well as on their qualifications and skills.

In connection with assessment by written tests of non-cognitive qualities a study by Griffin and Wilson[1], on the degree to which the results of such tests can be faked by applicants who respond in a socially desirable direction, is timely. However, it would be tempting to infer from their findings based on one such test, the 120 item IPIP – which measures personality constructs that are in the public domain and are widely known and discussed - that all non-cognitive tests could be faked to an extent that the results would be of little value. As a consequence medical schools might use the findings of Griffin and Wilson to argue against the use of tests of non-cognitive personal qualities in the selection procedure.

We would like to contribute to the discussion some additional data from long term and extensive studies we have conducted with the non-cognitive tests that we have developed for the purpose of assessing the personal qualities of potential health professional students (see www.pqa.net.au).

One of our tests combines the qualities of narcissism, aloofness, confidence in dealing with others and empathy in a single scale measuring detachment from others (narcissism and aloofness) versus involvement (empathy and confidence).[2] Although the components might be expected to be strongly influenced by social desirability considerations, we have developed the wording of (100) items in the test in a way that differs from the items in IPIP scales, in that the socially desirable option is not readily apparent. The indication that this approach has worked has come from comparing data gathered from low stakes test samples (existing students taking the test voluntarily) with that from high stakes test samples (applicants who were aware the results would be used to select them). The figure shows the data gathered.

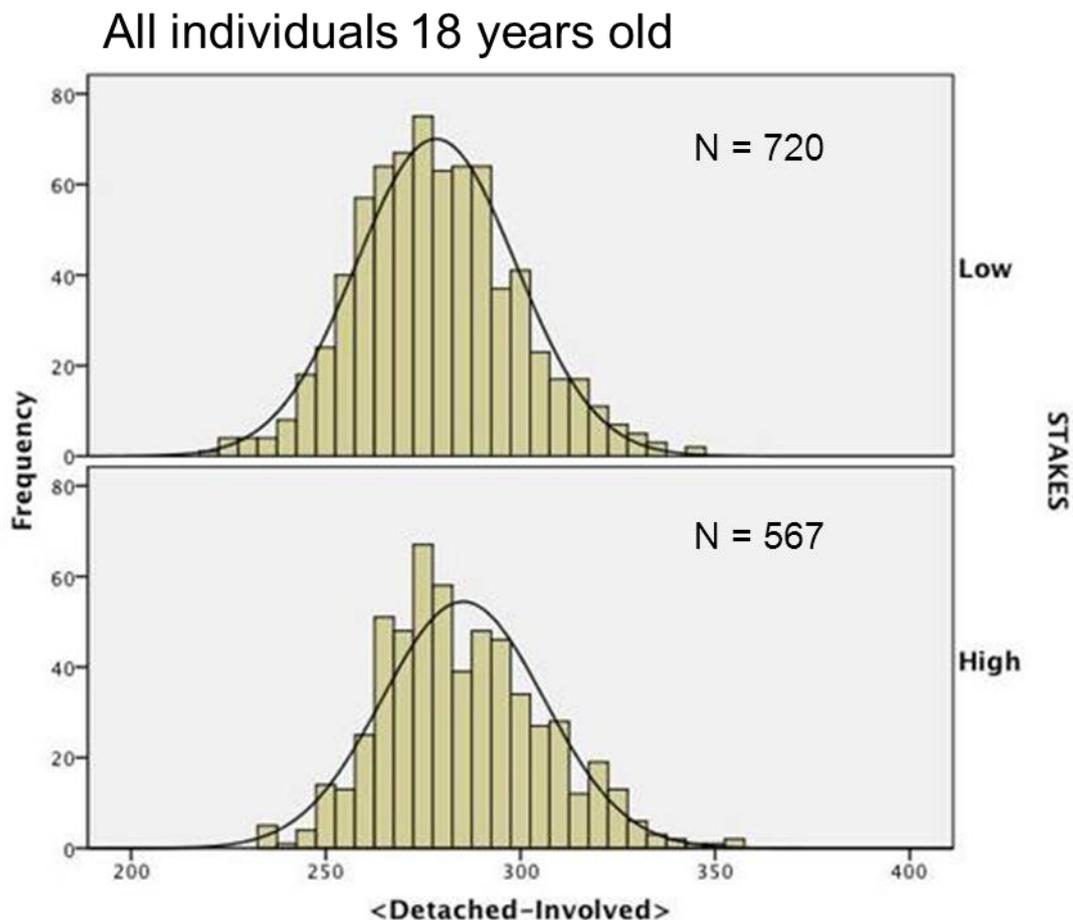## All individuals 18 years old

Figure 1: Comparison of samples of 18 years old individuals who completed the NACE test either under low stakes or high stakes conditions.

An important point to note in the context of the present discussion is that the score distributions of both samples are near normal and comparable. The standard deviations of the two samples are almost equal (20.50, low stakes; 20.77, high stakes). In other words the high stakes distribution is not more leptokurtic, as would be the case if candidates were trying to avoid appearing extreme in any dimension. In addition the high stakes data are not negatively skewed, as would be the case if applicants were more consistently identifying the socially desirable response, i.e., in the more involved direction. Indeed there appears to be a slight skew in the opposite direction, towards 'detachment'.

The mean score is shifted towards the 'involved' end of the distribution (high stakes 285.3; low stakes 278.3), suggesting a social desirability effect in response in high stakes conditions, but the shift, though highly significant statistically (F 35.94; $p < 0.001$) with such large samples, is comparatively small (a difference of 7 points in an observed range of 169 points)

We recognise that faking good will occur to greater or lesser extent in any high stakes testing. However, through careful test design and the use of appropriate norms collected under the same testing conditions it appears to be possible to minimise its effects so that selection based on high versus low scores (or de-selection based on rejection of extreme scores) remains valid.

The additional point worth making is that even if the high stakes distribution were shifted very significantly towards the perceived socially desirable pole then selection decisions could still be made on the basis of the distribution measured under those conditions. The study of Ones *et al.* concluded that "there is little evidence that response distortion among job applicants ruins the psychometric properties, including criterion related validity, of personality measures".[3]

References:

1. Griffin B, Wilson IG. Faking good: self-enhancement in medical school applicants. *Med Educ* 2012; **46:** 485-490.

2. Munro D, Bore M, Powis D. Personality factors in professional ethical behaviour: studies of empathy and narcissism. *Aust J Psychol* 2005; **57:** 49-60.

3. Ones DS, Dilchert S, Viswesvaran C, Judge TA. In support of personality assessment in organizational settings. *Personnel Psychol* 2007; **60:** 995-1027.