Computers in Human Behavior 27 (2011) 2386–2391

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Does online psychological test administration facilitate faking?

Rachel Grieve *, Hidee Tania de Groot

School of Psychology, Australian Catholic University, Australia

ARTICLE INFO

Article history:
Available online 30 August 2011

Keywords:
Online assessment
Internet testing
Faking
Malingering
Psychological testing
Equivalence

ABSTRACT

This study examined for the first time the effect of delivery mode on faking good and faking bad in psychological testing. Participants (N = 223) completed questionnaires either online or in pen-and-paper format in a mixed experimental design. After completing measures of personality (HEXACO-60, Ashton & Lee, 2009) and depression (DASS-21, Lovibond & Lovibond, 1995) under standard instructions, participants then faked the personality measure as if applying for a job, and faked the depression measure as if experiencing severe depression. Equivalence of internet and pen-and paper-administration on faking was then measured between groups. As predicted, participants were able to fake good on the HEXACO-60 and to fake bad on the DASS-21. Also as predicted, there were no significant differences in faked scores as a function of test administration mode. Further, examination of effect sizes confirmed that the influence of test administration mode was small. It was concluded that online and pen-and paper presentation are largely equivalent when an individual is faking responses in psychological testing. Given the advantages of online assessment and the importance of valid psychological testing, future research should investigate whether the current findings can be generalised to other faking and malingering scenarios and to other psychological measures.

1. Introduction

The internet is being increasingly used for psychological research and assessment in a number of contexts, including for vocational (Piotrowski & Armstrong, 2006) and clinical (Hedman et al., 2010) purposes. Much research has examined the equivalence of pen-and-paper and web-based versions of specific psychological measures in a variety of domains (e.g. Coles, Cook, & Blake, 2007; Denniston et al., 2010; Hedman et al., 2010; Lewis, Watson, & White, 2009; Templer & Lange, 2008). However, the equivalence of psychological test presentation via the internet and pen-and-paper has not been examined in regards to the susceptibility of self-reports to faking. This study aimed to explore for the first time the influence that mode of delivery elicits on the fakability of self-report psychological tests.

1.1. Online test delivery

The internet has rapidly become a valuable medium for data collection as it is considered inexpensive, easily accessible, and discrete (Birnbaum, 2004). Other advantages of on-line data collection are that participants can be required to endorse answers to all items (thereby minimising missing data), and that data can be transferred electronically for analysis (thereby reducing data entry error) (Carlbring et al., 2007; Lewis et al., 2009). However, rather than assuming that internet and pen-and-paper administrations of psychological measures are interchangeable, it has been recommended that all psychological measures be evaluated to investigate whether internet and pen-and-paper administrations are comparable (Buchanan, 2002).

To date, a number of tests have been compared, including clinical measures (e.g. Carlbring et al., 2007; Coles et al., 2007; Herrero & Meneses, 2006), personality measures (e.g. Templer & Lange, 2008), ability measures (e.g. Ihme et al., 2009), and health- and risk-related behaviour measures (e.g. Horswill & Coster, 2001; Lewis et al., 2009; McCabe, Couper, Cranford, & Boyd, 2006; Whittier, Seeley, & St. Lawrence, 2004). Overall, findings suggest that the internet is both a feasible and largely comparable method for conducting psychological testing. However, no extant research has examined the role of mode of delivery in the administration of self-report psychometric tests and the potential facilitation of faking behaviour.

1.2. Faking and malingering

Faking or malingering occurs when an individual strategically alters their self representation in a particular test (Grieve & Mahar, 2010). Faking good is characterised by responses that augment an individual's actual state, making them appear psychologically

* Corresponding author. Address: School of Psychology, Australian Catholic University, 1100 Nudgee Rd., Banyo, Queensland 4014, Australia. Tel.: +61 7 3623 7346; fax: +61 7 3623 7277.
E-mail address: rachel.grieve@acu.edu.au (R. Grieve).

0747-5632/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.chb.2011.08.001

superior (for example, in a job application), while faking bad occurs when an individual presents themselves as psychologically worse than they actually are (for example, to be diagnosed with a disorder).

Faking of psychological assessments may have a number of consequences. For example, in vocational contexts, faking will not only influence who gets hired (Mueller-Hanson, Heggestad, & Thornton, 2003), but can also impact the subsequent training and management of employees (Landers, Sackett, & Tuzinski, 2011). In clinical contexts, faking may influence access to therapy or medication (Suhr, Hammers, Dobbins-Buckland, Zimak, & Hughes, 2008).

1.3. The current research

This study aimed to build on the existing research regarding the validity of pen-and-paper and online testing methods by investigating whether administration mode influences an individual's ability to fake a measure. To more fully address this aim, both faking good and faking bad scenarios were employed.

1.3.1. Faking good

Previous research has shown that individuals are readily able to fake good in vocational contexts (for example, as if applying for a job) by maximising positive, job-relevant personality aspects and minimising negative personality aspects (Mahar et al., 2006). Therefore, an initial hypothesis was that participants would be able to alter their original personality profiles to a more positive faked profile when asked to complete a personality measure as if they were applying for a job. Specifically, it was anticipated that the faked profiles would score significantly higher than the original profiles on desirable employee characteristics (honesty/humility, extraversion, agreeableness, conscientiousness, and openness), and significantly lower on undesirable employee characteristics (emotionality).

The second hypothesis addressed the main research question. Given that most research into the equivalence of online and pen-and-paper personality testing has found that both modes of administration elicit similar test results (e.g. Templer & Lange, 2008), it was hypothesised that faked profile scores would be equivalent regardless of which mode of administration was used. While it is acknowledged that this is in fact testing the null hypothesis, and that it is difficult to ascertain whether a hypothesis of no difference is true (Nickerson, 2000), a hypothesis of this nature was required by the research question. It follows that, in order to test the second hypothesis, a close examination of effect size, rather than statistically significant differences alone, was indicated.

1.3.2. Faking bad

It has also been shown that individuals are able to fake bad in clinical contexts (for example, as if they have depression, see Grieve & Mahar, 2010). Therefore, it was hypothesised that when participants were asked to complete a depression measure as if they had depression, they would be able to alter their original scores on that measure to faked scores suggesting a provisional diagnosis of depression.

In order to address the main research question, scores were compared between groups of participants who faked the depression measure either online or using pen-and-paper. Again, as previous research has largely supported the equivalence of the two modes of administration for clinical measures (e.g. Carlbring et al., 2007), it was hypothesised that there would be no significant differences in faked depression scores as a function of administration method. Once more, as this prediction was testing the null hypothesis, close examination of the effect size was also undertaken.

2. Methods

2.1. Participants

The sample consisted of 223 participants (54 men, 169 women) who completed the questionnaire on the internet (63%) or on paper (37%). Participants were recruited from the student body of an Australian university (41.5%), and the general public (51.8%). 6.7% of participants did not report whether or not they were students. Participants were invited to participate via in-class announcements, word of mouth, and using the social networking website *Facebook*. Participation was voluntary and no remuneration was offered. However, participants enrolled in undergraduate psychology courses were eligible for course credit.

2.2. Design

A mixed experimental design was used. A within groups comparison was made between participants' original scores and faked scores, on the personality and depression measures, where the independent variable was test instruction (with two levels: standard instruction and faking instruction), and the dependent variable was test score. A series of distractor measures was included between the initial completion of the personality and depression items and the faked personality and depression items to minimise practice effects. The administration of faking good and faking bad instruction was counterbalanced to control for order effects.

A between groups design was then used to examine the equivalence of pen-and-paper versus the internet testing on faked scores. The independent variable was test administration with two levels, internet and pen-and-paper. The dependent variable was the faked test score.

2.3. Materials

2.3.1. Personality

Personality was measured using the HEXACO-60 (Ashton & Lee, 2009), a 60-item Likert-style self-report report measure that captures personality in six constructs: Honesty/humility, Emotionality, Extraversion, Agreeableness, Conscientiousness, and Openness to experience. The anchor points were 1 = *strongly disagree*, and 5 = *strongly agree*. Sample items are "*I wouldn't use flattery to get a raise or promotion at work, even if I thought it would succeed*" (Honesty/humility), "*I sometimes can't help worrying about little things*" (Emotionality), "*I prefer jobs that involve active social interaction to those working alone*" (Extraversion), "*Most people tend to get angry more quickly than I do*" (Agreeableness), "*People often call me a perfectionist*" (Conscientiousness), and "I *like people who have unconventional views*" (Openness). Cronbach's alphas for these scales were α = .73–.80 in Ashton and Lee's sample, and .72–79 in the current sample, suggesting they have adequate internal reliability.

2.3.2. Depression

The short version of the Depression, Anxiety, Stress Scale (DASS-21; Lovibond & Lovibond, 1995) was used. This 21-item self-report asks participants to indicate how items applied to them over the past week, with the anchors 0 = *Did not apply to me at all*; 1 = *Applied to me to some degree, or some of the time*; 2 = *Applied to me to a considerable degree, or a good part of the time*; and 3 = *Applied to me very much, or most of the time*. For the purpose of this study, only the seven items related to depression were utilised. A sample item is "*I couldn't seem to experience any positive feeling at all*". Lovibond and Lovibond found reliabilities for the DASS-21 to be good, with a Cronbach's alpha of .91, for the depression