



# Employment testing online, offline, and over the phone: Implications for e-assessment



Rachel Grieve\*, Jordana Hayes

University of Tasmania, Australia

## ARTICLE INFO

### Article history:

Received 3 August 2015

Accepted 19 April 2016

Available online 30 May 2016

### Keywords:

Computer-mediated assessment

Internet testing

Telephone testing

Equivalence

Vocational testing

Faking

e-assessment

## ABSTRACT

This research investigated faking across test administration modes in an employment testing scenario. For the first time, phone administration was included. Participants ( $N=91$ ) were randomly allocated to testing mode (telephone, Internet, or pen-and-paper). Participants completed a personality measure under standard instructions and then under instructions to fake as an ideal police applicant. No significant difference in any faked personality domains as a function of administration mode was found. Effect sizes indicated that the influence of administration mode was small. Limitations and future directions are considered. Overall, results indicate that if an individual intends to fake on a self-report test in a vocational assessment scenario, the electronic administration mode in which the test is delivered may be unimportant.

© 2016 Colegio Oficial de Psicólogos de Madrid. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Pruebas de selección on-line, off-line y telefónicas: implicaciones para la evaluación electrónica

## RESUMEN

Este trabajo investiga el falseamiento en los diferentes modos de aplicación de tests en el contexto de las pruebas para conseguir empleo. Por primera vez se incluyó la aplicación telefónica. Se distribuyó a los participantes ( $N=91$ ) aleatoriamente en las modalidades de prueba (telefónica, Internet o papel y lápiz). Los sujetos realizaron una prueba de personalidad con instrucciones estándar y después con instrucciones de que falsearan la prueba como si fuesen aspirantes ideales a la policía. No resultaron diferencias significativas en ninguno de los dominios de personalidad en función del modo de administración. La magnitud del efecto indicaba que la influencia del modo de aplicación era escasa. Se abordan las limitaciones y directrices con vistas al futuro. En general, los resultados indican que si una persona trata de falsear una prueba de autoinforme en el contexto de la evaluación profesional el modo de administración electrónica de la prueba puede carecer de importancia.

© 2016 Colegio Oficial de Psicólogos de Madrid. Publicado por Elsevier España, S.L.U. Este es un artículo Open Access bajo la CC BY-NC-ND licencia (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

### Palabras clave:

Evaluación por ordenador

Pruebas por Internet

Pruebas telefónicas

Equivalencia

Pruebas profesionales

Falseamiento

Evaluación electrónica

The use of self-report psychological tests provides an opportunity for test-takers to provide false, strategic responses (MacNeil & Holden, 2006), in turn threatening the validity of test results (Tett & Simonet, 2011). With alternative forms of psychological test administration, a burgeoning field with particular importance in organisational contexts (Piotrowski & Armstrong, 2006), the aim of this research was to extend examination of electronic test administration mode and faking behaviour. For the first time, this study explored the effect of telephone, internet, and pen-and-paper test

\* Corresponding author. Psychology Division. Faculty of Health. University of Tasmania. Private Bag 30. Hobart, Tasmania, Australia, 7001.

E-mail address: [rachel.grieve@utas.edu.au](mailto:rachel.grieve@utas.edu.au) (R. Grieve).

administration on the faking susceptibility of self-report psychological tests.

### Applicant Faking

When an individual responds dishonestly and in a strategic manner on a psychological test, this is regarded as faking (Grieve & Mahar, 2010). Faking good refers to the act of deliberately altering test scores in order to appear more favourably (McFarland & Ryan, 2000). In high demand contexts such as selection for employment ‘applicant faking’, faking can be especially concerning (Tett & Christiansen, 2007). A job offer may be a reward for faking job-relevant traits well on a personality scale (Tett & Simonet, 2011). Organisations may then be at risk of hiring an applicant who has presented an incongruous personality profile, which may then have negative consequences for the organisation and the employee. It also means that an applicant who is a better fit with the organisation has missed an opportunity to be hired (Tett & Simonet, 2011).

Research into different administration modes has noted that online and telephone administrations provide good alternatives to pen-and-paper testing, and there are a number of benefits associated with these media. For example, the low-cost of online and telephone testing has been cited (Baca, Alverson, Manuel, & Blackwell, 2007; Templer & Lange, 2008), the ability for these media to reach people in rural areas (Baca et al., 2007), increased access to larger samples (Ryan, Wilde, & Crist, 2013), and to broaden the demographic profiles of respondents (Casler, Bickel, & Hackett, 2013; Lewis, Watson, & White, 2009). In the vocational context, these aspects of alternative forms of assessment may be of particular interest given the increasing trend towards teleworking (Mahler, 2012).

However, given the potential consequences resulting from faking, there is a pressing need to be able to detect and measure the behaviour across a number of test administration media. Any findings suggesting that faking differs depending on delivery mode could have critical implications for how tests are administered. For example, if a particular administration mode is susceptible to faking, then it may be prudent for assessors to avoid that mode of delivery. In addition, exploring faking across administration modes may also add to our understanding of faking behaviour.

While many measures have been shown to be equivalent when comparing online and pen-and-paper delivery modes (e.g., Bates & Cox, 2008; Carlbring et al., 2007; Casler et al., 2013; Williams & McCord, 2006), to date, research examining vocational faking as a function of administration is limited, with only one study comparing faking in online and pen-and-paper contexts. Grieve and de Groot (2011) examined faking across Internet and pen-and-paper administration modes. Participants were able to choose whether to complete the measures online or in pen-and-paper format. Participants first completed both measures honestly, completed distractor items, and then completed the HEXACO-60 (Ashton & Lee, 2009) under ‘fake good’ instructions (told to imagine they are applying for their “ideal job”). A between groups analysis found no significant difference in the faked scores across administration modes, and the effect sizes were small. The authors concluded that when an individual is faking, the mode of test administration has minimal influence. However, while providing promising initial insight into the susceptibility of online measures, there are limitations to this study that warrant additional consideration.

Firstly, the ‘fake good’ condition was vague in its use of an “ideal job” (Grieve & de Groot, 2011) as the target profile. It would seem possible, if not probable, that participants would have had varying conceptions of their ideal job, and would have responded differently depending on their job preferences. This would invariably lead to a variety of personality profiles depending on (for

example) whether a participant wants to be a librarian or an advertising executive, as demonstrated by Furnham (1990). Thus, the use of a specific job in the faking good instructions would add greater experimental control. Secondly, participants in Grieve and de Groot’s (2011) study were not randomly assigned to administration conditions, which may have resulted in systematic differences in responses as a function of modality preference. Finally, telephone administration was not considered in Grieve and de Groot’s study. If telephone administration were to yield different results to online and pen-and-paper delivery modes when faking, this would add insight into the use of the telephone for psychological testing and e-assessment more broadly.

### Equivalence of Telephone Administered Measures

Existing research into the equivalence of telephone testing is limited. Knapp and Kirk (2003) explored the equivalence between other administration modes, with the inclusion of an automated touch-tone phone condition. Participants were randomly allocated to either a pen-and-paper group, an Internet group, or a touch-tone phone group and asked to answer highly sensitive questions (for example, ‘Have you ever had phone sex?’) and also rated how honestly they had answered the questionnaire. The results showed no significant difference between groups on any of the questionnaire items and no significant difference in how honestly participants rated their answers.

However, other research comparing telephone, Internet, and mail surveys has found that participants tended to give more extreme positive responses in telephone administration (Dillman et al., 2009). The effect of online and telephone administered surveys on responses regarding alcohol use and alcohol-related victimisation has also been investigated, with the finding that women in the online group answered in a less socially desirable way (Parks, Pardi, & Bradizza, 2006).

So, with limited research on the equivalence of telephone administration, it may be difficult to make inferences about the utility of this delivery mode in psychological testing. Importantly, a specific gap exists in terms of the inclusion of the telephone in faking equivalence research. Thus, it is currently unclear as to how telephonic administration mode might influence faking in job applicants.

### The Current Research

The current study extended investigation of faking across administration modes by using random allocation (thereby mitigating self-selection concerns), by providing a specific job as the target profile (to strengthen the experimental manipulation) and by examining for the first time the influence of telephone mediated administration in addition to online and pen-and-paper testing.

Selection of the specific job to act as the target profile was predicated on including a job role that was broadly known to participants. Mindful of the need to facilitate interpretation of the results within existing job-specific vocational application personality data, the role of police officer was selected. As Detrick, Chibnall, and Call (2010) had investigated faking in police applicants, use of the police officer target profile allowed comparisons to be made in terms of test scores. The applicants in Detrick et al.’s (2010) study self-reported high on emotional stability, agreeableness, extraversion, and conscientiousness, and were able to significantly change their scores to such an extent as to alter their rank ordering. This knowledge allowed predictions to be made about the nature of faking good in the current study. Thus, it was hypothesised that participants would be able to fake good when instructed to complete a personality measure when applying for a job as a police recruit. Specifically, in line with Detrick et al., it was expected that

participants would score significantly higher on conscientiousness, agreeableness, and extraversion and score lower on neuroticism.

However, the main focus of the current research was to examine the role of test administration mode. Consistent with Grieve and de Groot's (2011) approach, the effect of testing modality on faked scores was investigated by comparing faked scores on the personality measure as a function of administration mode. As the current research was exploratory in nature (by including telephone administration for the first time), explicit hypotheses regarding any effects of test administration were not generated. However, in order to effectively examine the effects administration modality, careful consideration of effect sizes was included: this approach was in line with Grieve and de Groot's procedure.

## Method

### Participants

The sample consisted of 91 participants (67 female and 24 male) who completed the questionnaire online (31 participants), over the telephone (30 participants), or in pen-and-paper format (30 participants). Participants were recruited from an Australian university (52.75%) and from the general public (47.25%). The average age of participants was 30 years.

### Design

In order to compare faked and original scores and to investigate the effect of administration mode, a mixed-experimental design was used. Using test instruction as the independent variable (with two levels: standard instructions and instructions to fake) and test score as the dependent variable, a within-groups design was used to compare original scores on the personality scale with faked scores. A between-groups design was then used to investigate the equivalence of faked scores across online, pen-and-paper, and telephone administration modes. The independent variable was administration mode (with three levels: online, pen-and-paper, or telephone) and the dependent variable was faked test score.

**Control measures.** Participants were randomly allocated to online, pen-and-paper, or telephone administration conditions. As is usual in faking research, the faked condition followed the original condition (e.g., Grieve & de Groot, 2011) so that the faking process would not impact original scores. Distractor items were included between the original test items and the items requiring participants to fake in order to minimise memory effects between administrations. This was a 40-item thinking style measure (the Rational-Experiential Inventory; Pacini & Epstein, 1999), and a 33-item measure of trait emotional intelligence (Schutte et al., 1998). All telephone administrations were conducted by one researcher (female), thus every participant in the telephone condition heard the same voice with the same pronunciations, inflections, and tone.

**A priori power analysis.** Power was considered *a priori* in order to ensure the study had sufficient power to find an effect of administration mode on faking. According to Keppel and Wickens (2004), with a minimum of 30 individuals in each group, there was power of .80 to find a medium effect with  $\alpha$  set at .05.

### Materials

**Personality.** Personality factors were measured using the IPIP version of the Five Factor Personality Model (Goldberg, 1999), which includes 50 statements measuring five personality factors: openness, conscientiousness, extraversion, agreeableness, and neuroticism. Participants are asked to read each statement and respond on a five-point scale using the anchors of 1 = *strongly disagree*, and 5 = *strongly agree*. Sample items are 'I have frequent mood

swings' (neuroticism), 'I am the life of the party' (extraversion), 'I believe in the importance of art' (openness), 'I have a good word for everyone' (agreeableness), and 'I am always prepared' (conscientiousness). Previous research noted strong reliability scores for extraversion at .85, agreeableness at .79, conscientiousness at .78, neuroticism at .86, and openness at .77 (Biderman, Nguyen, Cunningham, & Ghorbani, 2011).

**Manipulation check.** To ensure participants had noted the experimental instructions, after participants completed the scales under instruction to fake as an ideal police applicant, they were asked to provide a sentence outlining the strategy they used to respond to the items ('In one sentence, please describe what strategy you used to answer the previous questions').

### Procedure

Approval was gained from the university's Ethics Board. Participants were recruited via word-of-mouth, social media posts, and class announcements. Participants registered their interest in participating in research by writing their details on a sign-up sheet or by emailing the researcher directly. Those allocated to the online group were emailed a link to the online consent form and questionnaire, which was hosted on a secure server. Participants allocated to the telephone group were asked to provide their phone number and the most convenient time they could receive a call from the researcher. When phone contact was made, participants provided verbal consent and had all the instructions and every item read aloud to them by the researcher. Pen-and-paper consent forms and questionnaires were completed in a quiet space, such as an office or computer lab, with the researcher in the room.

To obtain original profile scores, participants first completed the personality measure under standard instructions. Following administration of distractor items, the participants completed the personality measure again, this time under instruction to fake. The faking good instructions read, 'Please take a moment to think about how you may present yourself if you were applying for a job in the police force, and wish to appear as the ideal police applicant. Please imagine you have been given a conditional job offer for the police if you successfully complete this questionnaire. You do not need to respond honestly. Please do the best you can to present yourself as the ideal police applicant.' Participants then completed the manipulation check. After completion of the measures, participants were debriefed and thanked for their time (verbally on the telephone, a hand out in the pen-and-paper group, and on the final page of the online questionnaire).

## Results

### Preliminary Analyses

The sample consisted of 31 participants in the online group, 30 in the pen-and-paper group, and 30 in the telephone group. Some breaches of normality in scores were evident: faked neuroticism, extraversion, agreeableness, conscientiousness were significantly skewed and kurtotic,  $p < .05$ . However, as these were most likely due to floor and ceiling effects when faking and as tests based on the  $F$  distribution are robust (Keppel & Wickens, 2004), the data was analysed in its original form.

**Manipulation check.** Answers to the manipulation check item indicated that every participant had followed the instructions and that the experimental manipulation was successful. Sample responses that indicated a participant had understood and acted on instructions to fake as an ideal police applicant were 'I imagined I was a police officer' and 'I thought about what the police recruiters would want to see in an applicant'. As every participant indicated

**Table 1**  
Descriptive Statistics and Scale Reliabilities across Administration Modes for the Five Factors of the IPIP Personality Scale.

	Cronbach's $\alpha$		Online <i>n</i> = 31				Pen-and-paper <i>n</i> = 30				Telephone <i>n</i> = 30			
			Original		Faked		Original		Faked		Original		Faked	
	Original	Faked	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Neuroticism	.86	.65	26.26	7.47	14.87	5.19	23.13	8.19	14.10	4.25	24.73	7.12	14.17	5.19
Extraversion	.83	.73	34.23	7.40	41.06	4.37	33.80	7.06	40.87	5.38	35.27	6.20	41.43	4.25
Openness	.69	.74	37.35	4.76	31.97	5.02	37.87	5.57	31.53	5.75	39.57	5.53	31.63	7.59
Agreeableness	.78	.68	36.97	5.78	39.23	5.93	37.47	4.69	40.33	4.50	38.53	6.32	40.43	4.19
Conscientiousness	.83	.82	35.16	7.48	47.10	5.09	37.80	6.27	48.33	2.54	33.87	6.42	47.60	3.98

Note. *N* = 91, *M* = mean, *SD* = standard deviation.

**Table 2**  
Differences between Scores in the Original Condition as a Function of Administration Mode.

Subscale	<i>F</i>	<i>df</i>	<i>p</i> -value	Partial $\eta^2$
Neuroticism	1.29	2, 88	.28	.030
Extraversion	0.34	2, 88	.70	.007
Openness	1.45	2, 88	.24	.030
Agreeableness	0.61	2, 88	.58	.010
Conscientiousness	2.54	2, 88	.08	.050

they had followed the instructions, all cases were included in the analysis.

**Sample check.** Independent groups *t*-tests were conducted to determine whether sample characteristics might have influenced results. A Bonferroni adjustment was used to control for multiple comparisons, with  $\alpha$  set at .00833. There were no significant differences as a function of gender on any variables (all *p* values were > .05). Therefore, it was deemed appropriate to analyse data from women and men together. The scores of university students and community members were also compared, with no significant differences noted across any of the scales in the honest condition. However, when instructed to fake as an ideal police applicant, students (*M* = 33.77, *SD* = 5.79) reported significantly more openness than community members, (*M* = 29.42, *SD* = 5.74), *t*(89) = 3.56, *p* = .001, *r* = .35. As no other significant differences were noted between students and community members on any of the other faked scales, with all *p* values > .05, and mindful of the fact that random allocation between groups meant that systematic differences as a function of sampling were unlikely, data from students and members of the community were analysed together.

### Descriptive Statistics

Descriptive statistics for the online, pen-and-paper, and telephone administrations for scores on personality factors for both original and faked profiles are presented in Table 1. Original profile scores across the IPIP personality questionnaire were similar to those in previous research, as were reliabilities for the original scores (McCrae & Costa, 2004). These good reliability scores are presented in Table 1. Reliability for faked scores across the personality domains were adequate and ranged from .65 to .82, similar to the reliability range of .73 to .84 previously reported in the literature (McCrae & Costa, 2004).

To examine any pre-existing participant differences in personality between the 3 administration conditions, ANOVAs were conducted on the original personality scores. Results are reported in Table 2. There were no significant differences on any of the personality dimensions. Effect sizes using partial eta squared were calculated and were small, ranging from .007 for extraversion to .05 for conscientiousness. We concluded that there were no systematic differences in personality between the participant groups.

**Table 3**  
Within-group (original vs. faked) *t*-statistic, *p*-values, and Effect Sizes (*r*, Cohen's *d*) for the Five Personality Factors.

Subscale	<i>t</i> -statistic	<i>df</i>	<i>p</i> -value	<i>r</i>	<i>d</i>
Neuroticism	11.69	90	< .001	.61	1.082
Extraversion	-8.80	90	< .001	.46	0.959
Openness	7.22	90	< .001	.37	0.850
Agreeableness	-3.12	90	.002	.10	0.439
Conscientiousness	-15.71	90	< .001	.74	1.188

**Table 4**  
Differences between Scores in the Faked Condition as a Function of Administration Mode.

Subscale	<i>F</i>	<i>df</i>	<i>p</i> -value	Partial $\eta^2$
Neuroticism	.28	2, 88	.75	.005
Extraversion	.11	2, 88	.89	.003
Openness	.04	2, 88	.96	.001
Agreeableness	.56	2, 88	.56	.012
Conscientiousness	.73	2, 88	.49	.016

### Inferential Statistics for Faking as a Police Applicant

Preliminary inspection of means suggested that participants had altered their test scores when asked to fake as if an ideal police applicant. In order to statistically test whether participants were able to fake good as hypothesised, original and faked scores were combined across administration modes and compared using paired-sample *t*-tests. Full details of the *t*-tests, including effect sizes, are presented in Table 3. Results indicated that after instructed to fake as the ideal police applicant, participants reported significantly higher levels of extraversion, agreeableness, and conscientiousness in comparison to their original profiles. Participants also significantly lowered their scores on neuroticism and openness when faking as the ideal police applicant. Effect sizes were interpreted in line with Cohen (1992). The effect of faking instruction on test scores was medium to large for scores of neuroticism, extraversion, openness, and conscientiousness, with between 13.69% and 54.76% of variance in scores explained by test instruction. Although significant, the effect of test instruction on agreeableness was small, with only 1% of variance in test scores explained by instruction to fake as an ideal police applicant.

An ANOVA was then conducted to examine the effect of administration mode on test scores when faking as the ideal police applicant. Results (presented in Table 4) indicated no significant differences between scores on any of the personality domains as a function of administration mode. There were no significant differences (with or without Bonferroni adjustment for multiple comparisons) on any of the personality domains. Effect sizes using partial eta squared were small for the faked personality domains, ranging from .001 for openness to .016 for conscientiousness. The non-significant *p* values and very small effect sizes indicate that participants in the pen-and-paper, online, and telephone groups

all faked in a similar pattern when instructed to respond to the measures as the ideal police applicant.

## Discussion

The aim of this research was, for the first time, to investigate faking good as a function of test administration mode (online, pen-and-paper, and telephone). In regards to faking good, participants were able to alter their scores on a five-factor personality scale when presenting themselves as an ideal police applicant. As hypothesised, participants reported significantly higher levels of conscientiousness, agreeableness, and extraversion, and lower levels of neuroticism when presenting themselves as an ideal police applicant. However, and in contrast to expectations, participants also reported significantly lower levels of openness when faking as an ideal police applicant.

In regards to the effect of administration mode, no hypotheses were developed: this research was exploratory in nature due to the inclusion of a telephone administration mode (in addition to online and pen-and-paper delivery) for the first time. There were no significant differences in scores as a function of administration mode when participants were faking as if they were an ideal police applicant. This finding was in line with the previous faking equivalence research by [Grieve and de Groot \(2011\)](#).

The use of an ideal police applicant as the 'fake good' target profile meant that the faked scores could be examined in the context of a specific personality profile. This strengthened the research design, as simply asking participants to fake for their 'ideal job'. Interestingly, participants created a police applicant profile slightly different to the one reported in existing research ([Detrick et al., 2010](#)). It would seem that the schema participants generated about the ideal police applicant was close, but not quite a complete match, to the actual ideal police applicant profile. Previous research does indicate that when faking, individuals form their own concept or schema of the desirable profile ([Jansen, König, Kleinmann, & Melchers, 2012](#)).

The current study built on previous research ([Grieve & de Groot, 2011](#)) examining the equivalence of electronic assessment methods in a vocational context by including telephone administration and a specific applicant profile. Overall, the results support previous research indicating equivalence between online and pen-and-paper test administration (e.g., [Bates & Cox, 2008](#); [Carlbring et al., 2007](#); [Casler et al., 2013](#); [Williams & McCord, 2006](#)), and between online, pen-and-paper, and telephone administration ([Knapp & Kirk, 2003](#)). When responding as the ideal police applicant, scores did not differ between administration modes on any of the personality scales. Very small effect sizes indicated that across administration modes, participants responded in a similar manner when asked to fake as an ideal police applicant.

These findings support previous research indicating that participants were able to fake good regardless of whether tests were administered online or in pen-and-paper format ([Grieve & de Groot, 2011](#)). These results indicate that the fakability of the personality scale is equivalent across pen-and-paper, telephone, and online administration. Specifically, it would seem that if individuals have an intention to fake, electronically mediated assessment does not facilitate faking.

## Considerations and Directions for Future Research

More broadly, the current study made an important extension to existing research by demonstrating equivalence across administration modes when faking for vocational purposes on another personality measure, the IPIP version of the Big Five ([Goldberg, 1999](#)). It was suggested by [Buchanan \(2002\)](#) that all measures be

assessed for equivalence, and thus the current research provides additional useful insight. Future vocational faking research should include other measures to examine faking across electronic media.

It was not known to what degree memory effects influenced participants' faked scores in the current study, as original and faked measures were completed in the same testing session. Although distractor measures were included between original and faked administrations, the possibility remains that scores were impacted by memory effects. The same limitation was noted in [Grieve and de Groot's \(2011\)](#) study and, as recommended by those researchers, the use of a Solomon Four-group design ([Braver & Braver, 1988](#)) may have been one method to control for any practice effects; however this was beyond the scope of the current study, given its exploratory nature.

As is common in faking research, the current research employed an analogue design to facilitate maximum experimental control and reduce individual differences in faking motivation ([Paulhus, Harms, Bruce, & Lysy, 2003](#)). The difficulty in measuring applicant faking in non-analogue designs has been stated by researchers ([Berry & Sackett, 2009](#)). It may also be difficult to measure faking as it naturally occurs because fakers often do not admit to their behaviour ([Taylor, Freuh, & Asmundson, 2007](#)). Using a manipulation check was a strength of the current research, as it ensured that only those who had deliberately faked were included in the sample.

Nonetheless, it should be noted that participants in the current study were faking under instruction, not in a real vocational setting. Mindful that faking effect sizes tend to be larger in analogue vocational research than in real life vocational scenarios, a cautious approach to generalising the current findings would be prudent.

It follows that future research may choose to examine whether the current findings can be generalised to settings where the stakes are considerably higher. In addition, situational factors, such as including a warning against faking, can also impact upon intention to fake and faking behaviour ([McFarland & Ryan, 2006](#)). It is possible that responses to warnings, and therefore intention to fake, would differ with administration mode, and this is something that could be investigated in future research.

The use of telephone administration was a unique aspect of the current research. However, there is room to further test this delivery mode. In particular, the use of an automated voice to conduct questionnaires may be something that future researchers could investigate. It may also be costly for an organisation to have an employee conducting these standardised tests over the telephone, so an investigation into a more automated method may have useful practical implications. As noted, a single researcher conducted every telephone questionnaire in order to strengthen experimental control in the current study. However, research indicates that responses from male and female participants may indeed be influenced by the sex of the experimenter ([Fisher, 2007](#)). Still, in a real-life employment application process, it is unclear whether one person would perform every telephone call.

It is also noted that in the current study a significant difference emerged between students and non-students on the 'openness' personality domain when responding as the ideal police applicant. Previous research has indicated that students and non-students fake in a similar pattern ([Grieve & de Groot, 2011](#); [Grieve & Mahar, 2010](#)), but nonetheless, a difference did emerge in the current study. This was the only area where students and non-students diverged, possibly due to a differential understanding of the role of a police officer. Students appeared to perceive the ideal police applicant as more intellectual and imaginative than their non-student peers, perhaps due to different experiences with police officers. In a real vocational environment, however, it should be noted that applicants would have the opportunity to thoroughly research the role and the desired personality attributes. Before completing a

personality test, an applicant would be able to learn about the responsibilities of the position and if, for example, an organisation was seeking an 'out-going, dynamic salesperson', an applicant may infer that a high extraversion score would be desirable.

Despite the possibility of a differential understanding of the target profile between students and non-students, the 'ideal police applicant' was specifically chosen due to police officers being a visible part of the work force with recognisable traits. The use of this target profile sets the current study apart from previous faking research that has instead used the vague "ideal job" as the target profile (Grieve & de Groot, 2011).

A priori power considerations showed that there was sufficient power to find a significant result for medium sized effects (Keppel & Wickens, 2004), but not to find small effects. Having a larger sample would have improved power to find significant small effects. Still, according to Cohen (1992) it remains that the effect sizes were very small, and that a larger sample may have led to the power to find statistically significant effects that were not necessarily meaningful effects.

Finally, the current study instructed participants to fake good but, as acknowledged by Grieve and de Groot's (2011) study, an important distinction needs to be made between ability to fake and tendency to fake. Just because an individual *can* fake when instructed does not mean that they usually *would* fake under standard instructions. While Grieve and Elliott (2013) have recently addressed this issue and found no differences in intentions to fake in online versus face-to-face testing, it remains that the nature of faking intentions in telephone administration is yet to be considered. Further research into the tendency to fake may provide a valuable contribution to the field of faking research, especially as it relates to e-assessment.

## Implications

The equivalence of faking across electronic administration modes has practical implications. In the vocational environment, there are clear benefits to online and telephone test administration. It has been noted that online testing enables testers to reach people in geographically diverse areas, is low-cost, and can be easily used to test a large number of people and to broaden the demographic profiles of respondents (Casler et al., 2013; Lewis et al., 2009; Ryan et al., 2013; Templer & Lange, 2008). Although in some cases the widespread use of telephone assessment may be impractical, the telephone also has benefits such as being able to reach people almost anywhere, being a familiar communication tool, and being inexpensive (Baca et al., 2007). Most recently, the need for ongoing evaluation of the role of technology in selection and assessment (Ryan & Ployhart, 2014) has been noted. If, as results from the current study indicate, online and telephone tests can be faked no more than pen-and-paper tests within a vocational context, then these delivery modes may be used with increased confidence. This may be relevant given trends towards teleworking (e.g., Mahler, 2012), and may be of particular relevance when assessing for employee promotion. Results from the current study combined with the noted benefits of e-assessment may provide encouraging support for online and/or telephone applicant or incumbent testing.

## Summary and Conclusion

Faking behaviour can have vocational implications (Tett & Simonet, 2011), and given the increasing use of non-traditional methods in organisational assessment (Piotrowski & Armstrong, 2006), investigation of faking as a function of test administration mode is indicated. The current study aimed to examine this behaviour across electronic administration modes by including

telephone administration for the first time and by incorporating a refined methodology.

The results indicated that individuals were able to alter their scores on a personality measure when instructed to fake as the ideal police applicant. The faked personality scores were equivalent across telephone, pen-and-paper, and online administration. Overall, results from this research demonstrate that if an individual intends to fake on a self-report test in a vocational testing scenario, the administration mode in which the test is delivered may be of little consequence. These findings have implications for testing, given the increase in electronically mediated assessment use.

## Conflict of Interest

The authors of this article declare no conflict of interest.

## References

- Ashton, M., & Lee, K. (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment*, 91, 340–345. <http://dx.doi.org/10.1080/00223890902935878>
- Baca, C. T., Alverson, D. C., Manuel, J. K., & Blackwell, G. L. (2007). Telecounseling in rural areas for alcohol problems. *Alcoholism Treatment Quarterly*, 25(4), 31–45. [http://dx.doi.org/10.1300/J020v25n04\\_03](http://dx.doi.org/10.1300/J020v25n04_03)
- Bates, S. C., & Cox, J. M. (2008). The impact of computer versus paper-pencil survey, and individual versus group administration, on self-reports of sensitive behaviours. *Computers in Human Behaviour*, 24, 903–916. <http://dx.doi.org/10.1016/j.chb.2007.02.021>
- Berry, C. M., & Sackett, P. R. (2009). Faking in personnel selection: tradeoffs in performance versus fairness resulting from two cut-score strategies. *Personnel Psychology*, 62, 833–863. <http://dx.doi.org/10.1111/j.1744-6570.2009.01159.x>
- Biderman, M. D., Nguyen, N. T., Cunningham, C. J. L., & Ghorbani, N. (2011). The ubiquity of common method variance: the case of the Big Five. *Journal of Research in Personality*, 45, 417–429. <http://dx.doi.org/10.1016/j.jrp.2011.05.001>
- Braver, M. C. W., & Braver, S. L. (1988). Statistical treatment of the Solomon Four Group Design: A meta-analytic approach. *Psychological Bulletin*, 104, 150–154. <http://dx.doi.org/10.1037/0033-2909.104.1.150>
- Buchanan, T. (2002). Online assessment: Desirable or dangerous? *Professional Psychology: Research and Practice*, 3, 148–154. <http://dx.doi.org/10.1037//0735-7028.33.2.148>
- Carlbirg, P., Brunt, S., Bohman, S., Austin, D., Richards, J., Öst, L.-G., & Andersson, G. (2007). Internet vs. paper and pencil administration of questionnaires commonly used in panic/agoraphobia research. *Computers in Human Behavior*, 23, 1421–1434. <http://dx.doi.org/10.1016/j.chb.2005.05.002>
- Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior*, 29, 2156–2160. <http://dx.doi.org/10.1016/j.chb.2013.05.009>
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159. <http://dx.doi.org/10.1037/0033-2909.112.1.155>
- Detrick, P., Chibnall, J. T., & Call, C. (2010). Demand effects on positive response distortion by police officer applicants on the revised NEO personality inventory. *Journal of Personality Assessment*, 92, 410–415. <http://dx.doi.org/10.1080/00223891.2010.497401>
- Dillman, D. A., Phelps, G., Tortora, R., Swift, K., Kohrell, J., Berck, J., & Messer, B. L. (2009). Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet. *Social Science Research*, 38(1), 1–18. <http://dx.doi.org/10.1016/j.ssresearch.2008.03.007>
- Fisher, T. (2007). Sex of experimenter and social norm effects on reports of sexual behaviour in young men and women. *Archives of Sexual Behaviour*, 31, 89–100. <http://dx.doi.org/10.1007/s10508-006-9094-7>
- Furnham, A. (1990). Faking personality questionnaires: fabricating different profiles for different purposes. *Current Psychology*, 9, 46–55. <http://dx.doi.org/10.1007/BF02686767>
- Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower level facets of several five factor models. In I. Mervielde, I. Deary, F. DeFruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (7) (pp. 7–28). Tilburg, The Netherlands: Tilburg University Press.
- Grieve, R., & de Groot, H. (2011). Does online psychological test administration facilitate faking? *Computers in Human Behavior*, 27, 2386–2391. <http://dx.doi.org/10.1016/j.chb.2011.08.001>
- Grieve, R., & Elliott, J. (2013). Cyberfaking: I can, so I will? Intentions to fake in online psychological testing. *Cyberpsychology, Behavior, and Social Networking*, 166, 364–369. <http://dx.doi.org/10.1089/cyber.2012.0271>
- Grieve, R., & Mahar, D. (2010). The role of fluid and emotional intelligence in malin-gering. *Australian Journal of Psychology*, 62, 103–111. <http://dx.doi.org/10.1080/00049530903032836>
- Jansen, A., König, C. J., Kleinmann, M., & Melchers, K. G. (2012). The interactive effect of impression motivation and cognitive schema on self-presentation in

- a personality inventory. *Journal of Applied Social Psychology*, 42, 1932–1957. <http://dx.doi.org/10.1111/j.1559-1816.2012.00925.x>. ISSN: 0021-9029
- Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook*. London: Pearson Prentice Hall.
- Knapp, H., & Kirk, S. A. (2003). Using pencil and paper, internet and touch-tone phones for self-administered surveys: does methodology matter? *Computers in Human Behaviour*, 19, 117–134. [http://dx.doi.org/10.1016/S0747-5632\(02\)00008-0](http://dx.doi.org/10.1016/S0747-5632(02)00008-0)
- Lewis, L., Watson, B., & White, K. M. (2009). Internet versus paper-and-pencil survey methods in psychological experiments: equivalence testing of participant responses to health-related messages. *Australian Journal of Psychology*, 61, 107–116. <http://dx.doi.org/10.1080/00049530802105865>
- MacNeil, B. M., & Holden, R. R. (2006). Psychopathy and the detection of faking on self-report inventories of personality. *Personality and Individual Differences*, 41, 641–651. <http://dx.doi.org/10.1016/j.paid.2006.03.004>
- Mahler, J. (2012). The telework divide: Managerial and personnel challenges of telework. *Review of Public Personnel Administration*, 32, 407–418.
- McCrae, R. R., & Costa, P. T. (2004). A contemplated revision of the NEO Five-Factor Inventory. *Personality and Individual Differences*, 36, 587–596. [http://dx.doi.org/10.1016/S0191-8869\(03\)00118-1](http://dx.doi.org/10.1016/S0191-8869(03)00118-1)
- McFarland, L. A., & Ryan, A. M. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology*, 85, 812–821. <http://dx.doi.org/10.1037/0021-9010.85.5.812>
- McFarland, L. A., & Ryan, A. M. (2006). Toward an integrated model of applicant faking. *Journal of Applied Social Psychology*, 36, 979–1016. <http://dx.doi.org/10.1111/j.0021-9029.2006.00052.x>
- Pacini, R., & Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of Personality and Social Psychology*, 76, 972–987. <http://dx.doi.org/10.1037/0022-3514.76.6.972>
- Parks, K. A., Pardi, A. M., & Bradizza, C. M. (2006). Collecting data on alcohol use and alcohol-related victimization: A comparison of telephone and web-based survey methods. *Journal of Studies on Alcohol*, 67, 318–323. Retrieved from <http://www.jsad.com>
- Paulhus, D. L., Harms, P. D., Bruce, M. N., & Lysy, D. C. (2003). The over-claiming technique: measuring self-enhancement independent of ability. *Journal of Personality and Social Psychology*, 84, 890–904. <http://dx.doi.org/10.1037/0022-3514.84.4.890>
- Piotrowski, C., & Armstrong, T. (2006). Current recruitment and selection practices: A national survey of Fortune 1000 firms. *North American Journal of Psychology*, 8, 489–496.
- Ryan, A. M., & Ployhart, R. E. (2014). A century of selection. *Annual Review of Psychology*, 65, 693–717. <http://dx.doi.org/10.1146/annurev-psych-010213-115134>
- Ryan, R. S., Wilde, M., & Crist, S. (2013). Compared to a small, supervised lab experiment, a large, unsupervised web-based experiment on a previously unknown effect has benefits that outweigh its potential costs. *Computers in Human Behavior*, 29, 1295–1301.
- Schutte, N. S., Malouff, J. M., Hall, L. E., Haggerty, D. J., Cooper, J. T., Golden, C. J., & Dornheim, L. (1998). Development and validation of a measure of emotional intelligence. *Personality and Individual Differences*, 25, 167–177.
- Taylor, S., Freuh, B. C., & Asmundson, G. J. G. (2007). Detection and management of malingering in people presenting for treatment of posttraumatic stress disorder: methods, obstacles, and recommendations. *Journal of Anxiety Disorders*, 21, 22–41. <http://dx.doi.org/10.1016/j.janxdis.2006.03.016>
- Templer, K. J., & Lange, S. R. (2008). Internet testing: equivalence between proctored lab and unproctored field conditions. *Computers in Human Behavior*, 24, 1216–1228. <http://dx.doi.org/10.1016/j.chb.2007.04.006>
- Tett, R. P., & Christiansen, N. D. (2007). Personality tests at the crossroads: A response to Morgeson, Campton, Dipboye Hollenbeck, Murphy, and Schmitt (2007). *Personnel Psychology*, 60, 967–993. <http://dx.doi.org/10.1111/j.1744-6570.2007.00098.x>
- Tett, R. P., & Simonet, D. V. (2011). Faking in personality assessment: a "multisaturation" perspective on faking as performance. *Human Performance*, 24, 302–321. <http://dx.doi.org/10.1080/08959285.2011.597472>
- Williams, J. E., & McCord, D. M. (2006). Equivalence of standard and computerized versions of the Raven Progressive Matrices Test. *Computers in Human Behavior*, 22, 791–800. <http://dx.doi.org/10.1016/j.chb.2004.03.005>