12-2008

Impact of elaboration on responding to situational judgment test items

Filip LIEVENS

Singapore Management University, filiplievens@smu.edu.sg

Helga PEETERS

Ghent University

DOI: https://doi.org/10.1111/j.1468-2389.2008.00440.x

Impact of Elaboration on Responding to Situational Judgment Test Items

Filip Lievens and Helga Peeters

Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium. filip.lievens@ugent.be

Although faking has been identified as a potential problem in situational judgment tests (SJTs), no studies have investigated proactive approaches for controlling faking in SJTs. Therefore, this study examined the impact of elaboration on responding to SJT items. Elaboration was operationalized as reason-giving. Two hundred and forty-seven master students were assigned to either an honest or a fake condition, and to a non-elaboration or an elaboration condition. Results showed that elaboration decreased the effect of faking for items with high familiarity. Elaboration on familiar items also decreased the percentage of fakers in the top of the distribution. Next, participants in the elaboration condition rated the SJT significantly higher in terms of allowing them to present themselves more realistically and to demonstrate their knowledge, skills, and abilities. Finally, there were no significant differences in participants' satisfaction with the SJT across the elaboration and non-elaboration condition.

1. Introduction

It is generally known that individuals tend to slightly overstate their abilities, skills, and positive characteristics. For example, people typically find themselves more interpersonally adept, trustworthy, and physically attractive than the 'average person.' In the context of personnel selection, self-enhancement represents only part of the picture as candidates also engage in intentional response distortion (Paulhus, 1991). This response distortion (either unconscious self-deception or deliberate impression management) is especially relevant for selection procedures that rely on self-reports such as personality inventories, overt integrity tests, trait-based emotional intelligence measures, biodata inventories, or situational judgment tests (SJTs).

In this study, we focus on response distortion in SJTs. As measurement methods, SJTs confront applicants with verbal descriptions of job-related scenarios and ask them to indicate how they would react by choosing an alternative from a list of response alternatives (McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001; Motowidlo, Dunnette, & Carter, 1990; Weekley, Ployhart, & Holtz, 2006). Although SJTs have significant criterion-related validity (McDaniel, Hartman, Whetzel, & Grubb, 2007), incremental validity over personality and cognitive ability (Clevenger, Pereira, Wiechmann, Schmitt, & Harvey, 2001), positive applicant reactions (Kanning, Grewe, Hollenberg, & Hadouch, 2006), and few adverse impact against minorities, a recent review showed that SJTs are prone to faking good (Hooper, Cullen, & Sackett, 2006). Given these findings, it is important to examine potential proactive strategies for reducing faking on SJTs. To the best of our knowledge, no published studies have explored the effectiveness of such strategies in the context of SJTs. In this study, we investigate whether requiring test-takers to elaborate on their answers is an effective method for controlling response distortion in SJTs.

2. Study background

2.1. Response distortion and SJTs

An individual's conscious distortion of responses to score favorably has also been referred to as faking (e.g.,

Kuncel & Borneman, 2007; Levashina & Campion, 2006; McFarland & Ryan, 2000). Faking has been extensively studied in non-cognitive self-report measures such as personality inventories, biodata inventories, and integrity tests (e.g., Alliger & Dwight, 2000; Becker & Colquitt, 1992; Dalen, Stanton, & Roberts, 2001; Graham, McDaniel, Douglas, & Snell, 2002; Kluger & Collela, 1993; Martin, Bowen, & Hunt, 2002; McFarland, Ryan, & Ellis, 2002; Ones, Viswesvaran, & Reiss, 1996; Ones & Viswesvaran, 1998).

So far, about a dozen studies have addressed the issue of faking in SJTs. Two main research streams can be distinguished. One stream of studies compared responses obtained under different instruction sets (honest and fake-good instructions). Generally, this body of research was conducted in a laboratory setting and aimed to determine the maximal limits of capability to fake on SJTs when instructed to do so (e.g., Juraska & Drasgow, 2001; Nguyen, McDaniel, & Biderman, 2005; Peeters & Lievens, 2005). Another strand of research involved field studies, comparing responses from various groups (e.g., students, applicants, and incumbents). These studies aimed to determine the typical and operational level of faking on SJTs in real-world settings (e.g., Ployhart, Weekley, Holtz, & Kemp, 2003; Reynolds, Winter, & Scott, 1999; Vasilopoulos, Reilly, & Leaman, 2000).

Recently, Hooper *et al*. (2006) reviewed existing research on faking and SJT mean scores. They drew three important conclusions. First, SJTs might be prone to faking good, with effect sizes ranging from $d = .08$ to .89. Second, this large variability could be explained by several moderating variables such as the SJT instructions used, the constructs measured, the transparency of the items, and the study design used. Third, Hooper *et al*. concluded that SJTs were less fakable than personality inventories.

2.2. Overview of approaches to detect or control faking

Although no approaches to detect or control faking have been examined in the SJT domain, several such methods have been explored in the context of faking non-cognitive measures (especially personality inventories, see Kuncel & Borneman, 2007; Mesmer-Magnus & Viswesvaran, 2006; Rothstein & Goffin, 2006, for excellent overviews). A first group of methods aims to detect faking after the test has been completed. These so-called reactive or detection methods often use social desirability or lie scales which gather information about individuals' faking tendency, enabling to make score corrections after the testing. Examples are the Marlowe–Crowne Social Desirability scale (Crowne & Marlowe, 1960) and the Balanced Inventory of Desirable Responding (Paulhus, 1991). Although these scales are easy to administer and score, a major drawback is that they have been found to be susceptible to faking themselves (Ones *et al*., 1996; Pauls & Crost, 2004; Viswesvaran & Ones, 1999). In addition, recent research has shown that correcting applicants' scores had minimal impact on mean criterion performance (Schmitt & Oswald, 2006) and did little to alter the proportion of correct selection decisions (Ellingson, Sackett, & Hough, 1999). Apart from the use of social desirability scales, other detection approaches have been examined. Attempts have been made to detect fakers by inserting bogus items in personality inventories (e.g., Anderson, Warner, & Spencer, 1984), using response latency indices (e.g., Holden, Wood, & Tomashewski, 2001), and using item response theory (e.g., Stark, Chernyshenko, Chan, Lee, & Drasgow, 2001). So far, results have been mixed. Recently, Kuncel and Borneman (2007) developed a new approach that showed somewhat more promise in detecting directed faking. In this method, fakers were detected on the basis of their idiosyncratic item responses.

A second group of methods is more proactive as they aim to prevent faking. One approach has consisted of using warnings that fakers can be identified and will be penalized (Becker & Colquitt, 1992; Dwight & Donovan, 2003; Pace & Borman, 2006; Vasilopoulos, Cucina, Dyomina, Morewitz, & Reilly, 2006). So far, the empirical evidence showed only meager effects [around .25 standard deviations (*SD*s)] for a combination of identification-only and consequences-only warnings on predictor scores and faking scale scores (Dwight & Donovan, 2003). Imposing forced response formats on test-takers has received renewed attention as a second proactive approach. Although a multidimensional forced-choice response format was effective for reducing score inflation at the group level (Bowen, Martin, & Hunt, 2002; Christiansen, Burns, & Montgomery, 2005; Converse, Oswald, Imus, Hedricks, Roy, & Butera, 2008; Jackson, Wroblewski, & Ashton, 2000), it was affected by faking to the same degree as a traditional Likert scale at the individual level of analysis (Heggestad, Morrison, Reeve, & McCloy, 2006).

2.3. Elaboration as a method to control faking

Another approach for controlling distortion in self-report inventories might consist of requiring people to elaborate on the answers provided. Specifically, they might be asked to write down reasons for why they would say or do something. In social psychology, reason-giving has been identified as one of the ways to hold people accountable (Lerner & Tetlock, 1999; Sedikides, Herbst, Hardin, & Dardis, 2002). Various social psychological experiments found that people's

tendency to think about themselves in positive terms and to present themselves favorably can be curtailed if they are asked to give reasons for their views (Halberstadt & Levine, 1999; Lerner & Tetlock, 1999; Sedikides, Horton, & Gregg, 2007; Wilson & LaFleur, 1995). Recently, Sedikides *et al.* (2007) discovered that increasing accountability through reason-giving is effective in toning down self-enhancement because it invokes people to conduct autobiographical memory searches. Such retrospective mental thoughts will bring not only socially desirable behaviors but also socially less desirable behaviors to mind. In turn, this activation and accessibility of a broader and more impartial set of behaviors are expected to lead to more honest responding. Sedikides *et al.* also found that the positive effect of reason-giving is established only when people are required to write down reasons (instead of simply contemplating about them).

If we apply this elaboration logic to SJT items, this might imply that candidates are required to write down reasons for why they chose a given response alternative. Although elaboration in the form of reason-giving has emerged as a promising approach in the social psychology literature, it remains to be seen whether this approach is also successful in controlling faking on SJTs. From a *conceptual* point of view, a key distinction between the research base in social psychology and personnel selection is that the former aims to control only unconscious self-enhancement (self-deception), whereas the latter deals with both unconscious self-deception and deliberate impression management. From an *empirical* point of view, initial evidence shows that elaboration might be a practical means to reduce both self-deception and impression management. Specifically, in two studies Schmitt and colleagues found that elaboration lowered mean biodata scores by .7–.8 *SD* units (Schmitt & Kunce, 2002; Schmitt, Oswald, Kim, Gillespie, Ramsay, & Yoo, 2003). In these studies, elaboration was not operationalized by the reason-giving dimension of accountability but by the verifiability dimension because biodata measures focus on standardized past-oriented questions that are often verifiable (Becker & Colquitt, 1992). Elaboration implied that participants had to give past behavioral incidents to support their answers. In light of these positive results for biodata, we hypothesize that elaboration (reason-giving) will also reduce faking in SJTs (Hypothesis 1).

However, we do not expect that elaboration will reduce faking for all SJT items. Specifically, we believe that elaboration (operationalized as reason-giving) is relevant when test-takers are familiar with the event or situation described in the item (Schmitt & Kunce, 2002). Hereby, 'familiarity' refers to test-takers' prior experience with actions taken as a response to a given situation and their awareness of these actions' consequences. In some selection situations, SJTs typically contain items that are mostly unfamiliar to candidates. An example is the use of SJTs among applicants who are seeking their first full-time job upon graduation or applicants for a managerial job without prior managerial experience. These test-takers have not yet acquired an articulated knowledge of the domain. Other examples are the use of SJTs in scenario-based training (Fritzsche, Stagl, Salas, & Burke, 2006) and educational admission (Lievens, Buyse, & Sackett, 2005; Oswald, Schmitt, Kim, Ramsay, & Gillespie, 2004; Peeters & Lievens, 2005). Conversely, it can be assumed that candidates will be relatively familiar with most SJT situations when SJTs are used in job incumbent samples or for licensing and certification purposes. Ployhart *et al.* (2003) summarized the impact of familiarity on SJT item responding by stating that applicants typically indicate what they think the best answer is, whereas incumbents indicate what they did in the past. We believe that elaboration will be particularly useful for familiar SJT items as compared with unfamiliar SJT items because it is much easier to engage in autobiographical memory searches and provide reasons for one's actions when one has already experienced the consequences of these actions in the past. Thus, we hypothesize that item familiarity will moderate the effects of elaboration on mean SJT scores, with these effects being stronger for familiar items than for unfamiliar ones (Hypothesis 2).

Apart from positing hypotheses regarding the effects of elaboration on mean SJT scores it is also important to consider the effects of elaboration from the perspective of the people who elaborate. As noted above, the basic premise behind elaboration is that it enables people to remember past behaviors more accurately and to self-evaluate more realistically (Schmitt & Kunce, 2002). This reasoning is in line with self-presentation theory (Hogan, 1991; Johnson, 1981), which posits that 'the best strategy for designing a valid scale is not make lying or misrepresentation difficult, but to make self-presentation as easy as possible' (Johnson, 1981, p. 767). Consistent with these assumptions, we hypothesize that participants will perceive that an SJT with elaboration permits them to self-evaluate significantly more realistically as compared with an SJT without elaboration (Hypothesis 3).

Besides more realistic self-evaluation the requirement to elaborate on one's answers might also provide participants with the opportunity to justify their responses. Although no studies in personnel selection have examined participants' perception of this potential benefit of elaboration, there exists empirical research in educational psychology that speaks to this issue. In the educational domain, there exists a rich tradition in multiple-choice testing of providing students with the opportunity to justify the item option chosen. Research

has consistently shown that in high-stakes exam situations students appreciate this opportunity to elaborate on their answers (e.g., Dodd & Leal, 1988; McKeachie, Pollie, & Speisman, 1955; Nield & Wintre, 1986). Conceptually, these positive results of the opportunity to elaborate can be framed in justice theory (Gilliland, 1993) because providing people with the opportunity to perform has been identified as a key procedural justice dimension (Arvey & Sackett, 1993; Bauer, Truxillo, Sanchez, Craig, Ferrara, & Campion, 2001; Gilliland, 1993). Opportunity to perform refers to 'the perception that one had an adequate opportunity to demonstrate one's knowledge, skills, and abilities (KSAs) in the testing situation' (Schleicher, Venkataramani, Morgeson, & Campion, 2006, p. 560). On the basis of these empirical and conceptual arguments, we hypothesize that an SJT with an elaboration format might lead to significantly higher perceptions of opportunity to perform as compared with an SJT with a non-elaboration format (Hypothesis 4).

From the perspective of the people who elaborate, however, elaboration is also not without potential drawbacks. In fact, one might question whether people will like to elaborate on a test when they know that their elaborations will not be evaluated. One might even ask whether it is ethical to require applicants to complete elaborations when there is no intention of using them. Similar ethical questions have been posited about the ethics of providing participants with warnings as a means for reducing faking (Rothstein & Goffin, 2006). Again, prior research about the use of elaborations (answer justifications in high-stakes multiple-choice testing) in educational psychology is instructive here. For example, Wittmaier (1976) discovered that answer justification lead even to negative reactions (i.e., more frustration and less satisfaction with the exam) when students were told that their elaborations would not be accounted for in the grading. As elaborations on the SJT will also not be taken into account in computing the SJT score in this study, we hypothesize that participants will be significantly less satisfied with an SJT with an elaboration format as compared with an SJT with a non-elaboration format (Hypothesis 5).

2.4. Present study

Taken together, this study contributes to the literature on faking in SJTs by examining whether elaboration (reason-giving) might be a useful proactive strategy for reducing faking on SJT scores. We scrutinize the effects of elaboration from a 'hard' psychometric perspective as well as from a 'soft' candidate perspective. First, we examine the effects of elaboration on mean SJT scores and assess item familiarity as a possible moderator. Second, we investigate participants' reactions to elaboration as means for reducing faking.

3. Method

3.1. Sample and procedure

A total of 249 third-year students from a large Belgian university volunteered to participate in this study. The average age was 22.3 years ($SD = 1.63$) and 70% was female. The sample mainly contained students majoring in Law and Criminology (40%), Economics (31%), and Political and Social Sciences (20%).

Participants were recruited by an invitation email for a session on psychological testing and assessment. At the start of the session, it was explained that the advantage of taking part in this session was that they could increase their experience with taking a variety of tests. In this session, participants completed a series of psychological tests. In this study, only their responses to the SJT were used. A couple of weeks later, participants received feedback about their test results via email.

3.2. SJT

To increase the realism and face validity of the SJT for our participants we used an SJT of college student success. This SJT consisted of items related to student-related situations and asked students how they would respond to each scenario by picking one response from a list of four alternatives. Note that the scenarios included are not specific to a particular major. Instead, the items covered generic situations related to teamwork, studying for exams, organizing, accomplishing assignments, interpersonal skills, social responsibility, perseverance, integrity, etc. Recently, several SJTs of college student success have been developed (Bess & Mullins, 2002; Oswald *et al*., 2004) as a response to the interest to use SJTs as complements to cognitive predictors in student admissions. We used the SJT that was originally developed by Bess and Mullins (2002) and translated to Dutch by Peeters and Lievens (2005). Prior research confirmed the non-cognitive nature of this SJT as it had a low correlation with a cognitive ability measure (Peeters & Lievens, 2005).

The scoring key had been developed using subject matter experts (see Peeters & Lievens, 2005, for more detailed information). As the purpose of the SJT of college student success was to select students who will successfully complete their (undergraduate) studies, Peeters and Lievens used first-year graduate students (who had just successfully finished their undergraduate studies) as experts. They independently completed the SJT and indicated the most and least effective options

per item. Afterwards they met to compare their answers. The consensus had to be 80% or higher. If this was not the case, discrepancies were resolved through discussion. On the basis of these expert judgments the scoring key was developed (see Motowidlo *et al*., 1990). If participants chose the response option identified by the subject matter experts as best, they received a score of $+1$. If they chose the response option identified by the subject matter experts as worst, they received a score of -1. They received a score of 0 if their responses were one of the other two options.

To determine the familiarity of the items we conducted a pre-study with a pool of students similar to the ones in the main sample. Specifically, 30 third-year psychology students were asked to rate their familiarity with the situations described in the 23 SJT items (0 = *no experience/familiarity with this situation or a very similar situation*; 4 = *a lot of experience/familiarity with this situation or a very similar situation*). It was made clear that familiarity referred to having acted in this situation or a similar situation and having been confronted with this action's consequences. Eight items with the lowest average familiarity rating ($M = .47$, $SD = .35$) and eight items with the highest average familiarity rating ($M = 3.30$, $SD = .38$), $t(29) = -34.92$, $p < .001$, were included in the final SJT, which therefore counted 16 items. Apart from the total SJT score, two composite SJT scores were computed for the familiar and non-familiar items separately. As fatigue might impact on elaboration, four different versions of the SJT (each with randomly determined item orders) were created.

Across all conditions, the internal consistency reliability of the scores of the 16-item SJT was .51. There were no significant differences in internal consistency reliability across conditions. Such internal consistency values are commonly found in SJTs because SJTs measure heterogeneous content and are therefore factorially complex (Chan & Schmitt, 1997; Clause, Mullins, Nee, Pulakos, & Schmitt, 1998).

3.3. Experimental design

Participants were randomly assigned to 16 conditions. A 2 (Honest vs Fake) $\times$ 2 (Non-Elaboration vs Elaboration) $\times$ 4 (SJT Item Order) $\times$ 2 (Non-Familiar Items vs Familiar Items) mixed design was used, with repeated measures on the last factor (all participants in the 16 between-subjects conditions responded to both familiar and non-familiar SJT items).

In this study, we adhered to the experimental paradigm in faking research. This meant that participants were given different instruction sets. Such directed faking studies constitute a worst case scenario (Kuncel & Borneman, 2007; Viswesvaran & Ones, 1999) as participants are instructed to present misleading and deceptive information, thereby eliminating possible variability in faking (McFarland & Ryan, 2000). In the 'honest' conditions, participants were instructed to answer the questions as honestly as possible and to indicate how they would really handle the situation. In the 'fake' conditions, they were instructed to make the best impression, to answer as they were taking part in a college admission exam wherein they tried to get the highest scores. These instructions were adapted from previous studies (e.g., McFarland & Ryan, 2000; Nguyen *et al*., 2005).

In the 'non-elaboration' conditions, participants had to choose one response from a list of alternatives. In the 'elaboration' conditions, they were asked to do the same. In addition, they were required to give short reasons or motivations for why they had chosen this specific alternative or why they would deal with the situation in this specific way. To this end, additional blank lines were added after each item in the elaboration condition.

3.4. Post-test measures

After taking the SJT, participants completed various post-test measures. All items used a five-point scale (1 = *strongly disagree*, 2 = *disagree*, 3 = *neither agree nor disagree*, 4 = *agree*, 5 = *strongly agree*). Three items were constructed to measure participants' perceived capacity to present themselves realistically (e.g., 'The test made my memories come back about the way I really handled these situations in the past'; $\alpha = .77$). Next, two items were used to assess participants' perceptions of their opportunity to perform on this test ('This test gives applicants the opportunity to show what they can really do'; $\alpha = .89$). These items were adapted from Bauer *et al*. (2001). Satisfaction with the SJT was measured with three items ('I would be satisfied if this type of test is used in an admission exam or selection process'; $\alpha = .83$). Finally, we included a self-reported faking scale in the post-test questionnaire as a manipulation check. Four items were constructed to check whether participants had engaged in faking on the SJT (e.g., 'I consciously tried to get the highest score on the SJT'; $\alpha = .88$).

4. Results

4.1. Manipulation checks

We began by checking whether participants elaborated on the SJT items when instructed to do so. The elaborations written down by the participants contained on average 18 words per item, suggesting that they took the elaboration task seriously. Only two

participants in the elaboration condition did not follow these instructions (i.e., they left the spaces to elaborate blank) and therefore were excluded from further analyses.

Our faking manipulation was also successful. Self-reported faking was significantly higher in the fake condition ($M = 3.21$, $SD = .88$) than in the honest condition ($M = 1.83$, $SD = .50$), $t(237) = -14.60$, $p < .001$, $d = 1.37$.

4.2. Effects of elaboration on mean scores

An ANOVA with item order as fixed factor and with total SJT score as dependent variable showed no multivariate effect of item order, $F(3, 243) = .46$, NS (partial $\eta^2 = .01$). Therefore, this factor was no longer considered in our analyses.

Hypotheses 1 and 2 dealt with the effects of elaboration on mean SJT scores. Hypothesis 1 specified a main effect of elaboration on SJT performance, whereas Hypothesis 2 posited that that familiarity would moderate the effect. Table 1 presents means and *SD*s of SJT scores, broken down by faking instructions and item familiarity. We conducted a $2 \times 2 \times 2$ (honest/fake $\times$ (non-) elaboration $\times$ item familiarity) mixed ANOVA with repeated measures on the third factor. There was a significant main effect of faking instructions on mean SJT scores, $F(1, 243) = 69.13$, $p < .001$ (partial $\eta^2 = .22$). This result is consistent with prior studies that indicate that SJTs are prone to faking good (Hooper *et al*., 2006). This main effect was qualified by an interaction effect between faking instructions and item familiarity, $F(1, 243) = 6.56$, $p = .01$, Wilks's $\lambda = .98$ (partial $\eta^2 = .03$). The effect of faking instructions was larger for familiar items ($M_{honest} = 3.94$, $SD = 2.03$ and $M_{fake} = 5.67$, $SD = 1.81$, $d = .82$) than for non-familiar items ($M_{honest} = 4.22$, $SD = 1.46$ and $M_{fake} = 5.31$, $SD = 1.42$, $d = .71$). Contrary to Hypothesis 1, the main effect of elaboration was not significant. Finally, results of the ANOVA revealed an interaction effect between faking instructions, elaboration, and item familiarity, $F(1, 243) = 6.16$, $p = .01$, Wilks's $\lambda = .98$ (partial $\eta^2 = .03$). Consistent with Hypothesis 2, Table 1 indicates that the effect of faking on SJT scores was smaller in the elaboration condition than in the non-elaboration condition, but only for familiar items. Specifically, the faking effect on familiar items was smaller ($d = .61$) in the elaboration condition than in the non-elaboration condition ($d = 1.04$).

To assess the practical relevance of this significant interaction effect between faking and elaboration on familiar items, we examined whether elaboration had effects on who would be selected. To this end, we investigated the percentages of fakers and honest respondents at different selection ratios (see also Mueller-Hanson, Heggestad, & Thornton, 2003; Rosse, Stechner, Levin, & Miller, 1998), broken down by elaboration condition. Table 2 shows that faking had the largest practical effects on who is getting selected at small selection ratios, as fakers rose to the top of the distribution. This effect decreased as the selection ratio increased. However, when elaboration was required the percentages of fakers who would be selected on the basis of test scores decreased (and the percentages of honest respondents increased) especially at smaller selection ratios. For example, if the selection ratio was .20, 85% would be fakers and 15% would be honest respondents when no elaboration on the SJT was required, whereas only 68% would be fakers (and 32% honest) when they were instructed to elaborate.

4.3. Effects of elaboration on participants' perceptions

To test Hypotheses 3–5 concerning the effects of elaboration on perceived self-presentation, opportunity to perform, and satisfaction with the SJT, independent sample *t*-tests were conducted. Results showed support for Hypotheses 3 and 4 because the perceived self-presentation ($M = 3.42$, $SD = .81$) and opportunity to perform ($M = 2.71$, $SD = .89$) were significantly higher in the elaboration condition than in the non-elaboration condition ($M = 3.18$, $SD = .82$ and $M = 2.48$, $SD = .91$, respectively), $t(242) = -2.28$ ($p < .05$), $d = .29$ and

Table 1. Means and standard deviations of SJT scores, broken down by study condition and item familiarity

	Honest condition			Fake condition		
	Non-elaboration ($n = 55$)	Elaboration ($n = 60$)	Total ($n = 115$)	Non-elaboration ($n = 63$)	Elaboration ($n = 69$)	Total ($n = 132$)
Familiar						
M	3.69	4.17	3.94	5.97	5.41	5.67
SD	2.16	1.90	2.03	1.60	1.96	1.81
Non-familiar						
M	4.22	4.22	4.22	5.17	5.43	5.31
SD	1.40	1.52	1.46	1.41	1.42	1.42

Note: SJT, situational judgment test.

Table 2. Percentages of fakers and honest respondents who could gain access to higher education with different selection ratios, when using familiar SJT items

Selection ratio	Non-elaboration condition		Elaboration condition	
	SJT score familiar items	Type of respondents	SJT score familiar items	Type of respondents
.10	≥ 8	8% honest 92% fakers	≥ 7	20% honest 80% fakers
.20	≥ 7	15% honest 85% fakers	≥ 6	32% honest 68% fakers
.30	≥ 6	22% honest 78% fakers	≥ 6	32% honest 68% fakers
.40	≥ 6	22% honest 78% fakers	≥ 6	32% honest 68% fakers
.50	≥ 5	29% honest 71% fakers	≥ 5	37% honest 63% fakers
.60	≥ 4	35% honest 65% fakers	≥ 5	37% honest 63% fakers
.70	≥ 4	35% honest 65% fakers	≥ 4	38% honest 62% fakers
.80	≥ 3	39% honest 61% fakers	≥ 3	43% honest 57% fakers
.90	≥ 2	42% honest 58% fakers	≥ 2	46% honest 54% fakers

Note: SJT, situational judgment test.

$t(244) = -2.06$ ($p < .05$), $d = .26$, respectively. Hypothesis 5 was not supported as there was no significant difference in satisfaction between the SJT with an elaboration format ($M = 2.72$, $SD = .88$) and the SJT without an elaboration format ($M = 2.99$, $SD = .80$), $t(244) = .01$ (NS), $d = .00$.

5. Discussion

The aim of our study consisted of investigating whether an approach (elaboration in the form of reason-giving) that has been found to be successful in social psychology in reducing self-enhancement would also be successful in reducing faking (as invoked by our directed faking instructions). Generally, our results indicate that elaboration decreased the effect of faking on mean SJT scores, but only among familiar items. In other words, fakers had lower scores when they were asked to elaborate on familiar items. Furthermore, this was not a trivial finding as this effect was not only statistically significant but also practically relevant. In fact, elaboration on familiar SJT items was found to produce positive effects on who would be hired (i.e., a higher percentage of honest respondents), especially in the case of low selection ratios. The fact that elaboration produced lower scores only among familiar items confirms that prior familiarity with the situations is an important factor in the context of the effects of elaboration on mean scores (Schmitt & Kunce, 2002). In fact, elaboration on familiar SJT items is somewhat comparable with elaboration on biodata items, given that both elaborations are based on the availability of behavioral examples, which may be necessary for producing the effect. In addition, the moderating effect of familiarity maps well into social psychological research about reason giving (Sedikides *et al*., 2007), showing that the activation of a more impartial set of behaviors through the requirement to write down reasons depends on the ability to engage in autobiographical memory searches.

The present study also identified item familiarity as an additional important variable (see Hooper *et al*., 2006) that determines the degree to which an SJT can be faked. We found that familiar items could be faked easier ($d = .82$) than non-familiar items ($d = .71$). One plausible explanation might be that experience and familiarity with situations lead to the development of strong schemas about the type of behaviors best suited in these situations, as shown by prior social cognitive research (Fiske & Cox, 1979; Klein, Loftus, Trafton, & Fuhrman, 1992). In turn, test-takers possessing relevant job knowledge and/or experience have been found to be better able to fake than inexperienced test-takers (Frei, Griffith, Snell, McDaniel, & Douglas, 1997; Vasilopoulos *et al*., 2000). Our results for familiarity are also consistent with research that shows that people can better fake more obvious (transparent) items than more subtle ones in personality inventories (e.g., Peterson, Clark, & Bennett, 1989; Posey & Hess, 1984). Apparently, possessing relevant schemata of the job domain and its requirements provides candidates with cues that make faking somewhat easier (Bowen *et al*., 2002; Christiansen *et al*., 2005). Transparent items are typically items that reveal these job requirements.

The moderating effect of familiarity on faking and elaboration leads to interesting practical implications.

From a practical point of view, elaboration seems a fruitful strategy to decrease faking when SJTs primarily consist of familiar items. In such applications, candidates can be expected to have already experienced most situations presented in the respective SJT items. Examples of such settings include the use of SJTs for licensing or certification purposes, for job incumbents, and for job seekers with considerable work experience in the domain of interest. Requiring test-takers to elaborate on their answers to familiar SJT items might then decrease the amount of faking and decrease the percentage of fakers in the top of the distribution. Accordingly, SJTs will provide a better assessment of whether test-takers have acquired the necessary knowledge and skills, which is the typical aim of certification and licensure exams (Raymond, Neustel, & Anderson, 2007; Shimberg, 1981). Conversely, elaboration does not seem to be useful when SJTs primarily consist of non-familiar items (e.g., the use of SJTs in a college admission context or training and developmental context). It should be noted, though, that our results show that the non-familiarity of such SJT items slightly impedes faking.

This study also began to examine participants' reactions toward the use of elaboration in SJTs. In line with our hypotheses, participants in the elaboration condition rated the SJT significantly higher in terms of allowing them to present themselves more realistically and to demonstrate their knowledge, skills, and abilities. These results are consistent with self-presentation theory (Johnson, 1981) and procedural justice theory (Gilliland, 1993), respectively. Although we hypothesized that participants would be less satisfied when they know that their elaborations on the SJT would not be evaluated no significant differences were found in participants' satisfaction with the SJT across the elaboration and non-elaboration condition. These results are encouraging for the use of elaboration. Yet, it is important to acknowledge that the effects found were relatively small. In addition, the mean participant perceptions were not very high (regardless of the condition).

Other limitations of this study are also in order. First, this study was not conducted in a real selection setting. Our study adhered to the experimental research paradigm for studying faking (Hooper *et al*., 2006). Directed faking studies constitute a worst case scenario as they reduce possible variability in faking tendencies among participants. Hence, we believe that an intervention that reduces faking in this worst case scenario, is also likely to reduce faking in more realistic selection settings. Yet, future studies should examine the generalizability of our findings in actual selection settings. Second, this study was conducted with students because they could be randomly assigned to either an honest or fake condition. Students might have completed the SJT items without the procedural and declarative knowledge of experienced employees. To sidestep this potential problem, we used an SJT that presented students with student-related problems. Accordingly, the SJT of college student success was relevant and realistic to them. As a third limitation, the familiarity of the SJT items was determined *a priori* in a pilot study (with psychology students). Hence, it is possible that some 'familiar items' were not familiar for a given participant or that some 'non-familiar' items were familiar for another participant. However, it should be noted that SJT tests are not an adaptive test format. This means that the same SJT is typically given to all candidates. When developing SJTs, *a priori* expert judgments (e.g., about the job relatedness of the situations) are typically made. In this study, *a priori* judgments about the familiarity of the items were also made.

In terms of future research, it is important to examine how elaboration might impact faking across various SJT formats. In this study, we found that elaboration (operationalized as reason-giving) had different effects on responding to an interpersonally oriented SJT depending on whether familiar vs non-familiar items were included. In a similar vein, future research might test whether elaboration has different effects as a function of the response instructions given to participants. This study employed behavioral tendency response instructions ('indicate what you would do'). It would be interesting to investigate the effects of elaboration on SJTs with knowledge-based instructions ('indicate the best answer'). Such SJTs have already been found to be less susceptible to faking (Nguyen *et al*., 2005). Another direction for future research is to examine whether elaboration is still useful when the SJT content is more cognitively oriented (e.g., SJT as a measure of procedural job knowledge). Thus, a comparison of the effectiveness of elaboration depending on the content and instructions of SJTs seems useful.

Another intriguing avenue for future research consists of examining the effects of elaborations on SJT validity. Although elaboration seems to reduce faking on familiar items, we do not know how elaboration might affect criterion-related validity. This is an important question because another proactive approach to faking (the use of warnings) has been found in some instances (e.g., in customer service contexts) to lead to lowered criterion-related validity (Harold, McFarland, Dudley, & Odin, 2004). In the case of the use of warnings, the lowered predictive validity may stem from the fact that applicants overcompensate their answers to ensure that they are not detected as fakers (Dwight & Donovan, 2003). Similar effects might be possible in the context of elaboration.

Finally, we need to know precisely why and how elaboration affects faking and test-taker responses. In

this study, applicants appreciated that elaborations provided them with the opportunity to evaluate themselves more realistically and to demonstrate their knowledge, skills, and abilities. However, we were not able to examine whether these perceptions were also related to the decrease in faking on familiar SJT items. Future research is needed to test this link between perceptions and faking. In a similar vein, other explanations regarding the effects of elaboration should be explored. For example, it is possible that elaboration reduces faking because it increases the cognitive load while responding, making it more difficult to engage in faking.

References

Alliger, G.M. and Dwight, S.A. (2000) A Meta-Analytic Investigation of the Susceptibility of Integrity Tests to Faking and Coaching. *Educational and Psychological Measurement*, **60**, 59–72.

Anderson, C.D., Warner, J.L. and Spencer, C.C. (1984) Inflation Bias in Self-Assessment Examinations: Implications for valid employee selection. *Journal of Applied Psychology*, **69**, 574–580.

Arvey, R.D. and Sackett, P.R. (1993) Fairness in Selection: Current developments and perspectives. In: Schmitt, N. and Borman, W. (eds), *Personnel Selection in Organizations*. San Francisco: Jossey-Bass, pp. 171–202.

Bauer, T., Truxillo, D., Sanchez, R., Craig, J., Ferrara, P. and Campion, M. (2001) Applicant Reactions to Selection: Development of the Selection Procedural Justice Scale (SPJS). *Personnel Psychology*, **54**, 387–419.

Becker, T.E. and Colquitt, A.L. (1992) Potential Versus Actual Faking on a Biodata Form: An analysis along several dimensions of item type. *Personnel Psychology*, **45**, 389–406.

Bess, T.L. and Mullins, M.E. (2002, April). *Exploring The Dimensionality of Situational Judgment: Task and contextual knowledge*. Paper presented at the 17th Annual Conference of the Society for Industrial and Organizational Psychology, Toronto, ON.

Bowen, C.C., Martin, B.A. and Hunt, S.T. (2002) A Comparison of Ipsative and Normative Approaches for Ability to Control Faking in Personality Questionnaires. *International Journal of Organizational Analysis*, **10**, 240–259.

Chan, D. and Schmitt, N. (1997) Video-Based Versus Paper-and-Pencil Method of Assessment in Situational Judgment Tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology*, **82**, 143–159.

Christiansen, N.D., Burns, G.N. and Montgomery, G.E. (2005) Reconsidering Forced-Choice Item Formats for Applicant Personality Assessment. *Human Performance*, **18**, 267–307.

Clause, C.S., Mullins, M.E., Nee, M.T., Pulakos, E. and Schmitt, N. (1998) Parallel Test Form Development: A procedure for alternate predictors and an example. *Personnel Psychology*, **51**, 193–208.

Clevenger, J., Pereira, G.M., Wiechmann, D., Schmitt, N. and Harvey, V.S. (2001) Incremental Validity of Situational Judgment Tests. *Journal of Applied Psychology*, **86**, 410–417.

Converse, P.D., Oswald, F.L., Imus, A., Hedricks, C., Roy, R. and Butera, H. (2008) Comparing Personality Test Formats and Warnings: Effects on criterion-related validity and test-taker reactions. *International Journal of Selection and Assessment*, **16**, 155–169.

Crowne, D.P. and Marlowe, D. (1960) A New Scale of Social Desirability Independent of Psychopathology. *Journal of Consulting Psychology*, **24**, 349–354.

Dalen, L.H., Stanton, N.A. and Roberts, A.D. (2001) Faking Personality Questionnaires in Personnel Selection. *Journal of Management Development*, **20**, 729–742.

Dodd, D.K. and Leal, L. (1988) Answer Justification: Removing the 'Trick' from multi-choice questions. *Teaching of Psychology*, **15**, 37–38.

Dwight, S.A. and Donovan, J.J. (2003) Do Warnings not to Fake Reduce Faking? *Human Performance*, **16**, 1–23.

Ellingson, J.E., Sackett, P.R. and Hough, L.M. (1999) Social Desirability Corrections in Personality Measurement: Issues of applicant comparison and construct validity. *Journal of Applied psychology*, **84**, 155–166.

Fiske, S.T. and Cox, M.G. (1979) Person Concepts: The effect of target familiarity and descriptive purpose of the process of describing others. *Journal of Personality*, **47**, 136–161.

Frei, R.L., Griffith, R.L., Snell, A.F., McDaniel, M.A. and Douglas, E.F. (1997, April) *Faking of Non-Cognitive Measures: Factor invariance using multiple groups LISREL*. Paper presented at the 12th Annual Meeting of the Society for Industrial and Organizational Psychology, St Louis, MO.

Fritzsche, B.A., Stagl, K.C., Salas, E. and Burke, C.S. (2006) Enhancing the Design, Delivery and Evaluation of Scenario-Based Training: Can situational judgment tests contribute? In: Weekley, J.A. and Ployhart, R.E. (eds), *Situational Judgment Tests: Theory, measurement and application*. Mahwah, NJ: Lawrence Erlbaum, pp. 301–318.

Gilliland, S.W. (1993) The Perceived Fairness of Selection Systems: An organizational justice perspective. *Academy of Management Review*, **18**, 694–734.

Graham, K.E., McDaniel, M.A., Douglas, E.F. and Snell, A.F. (2002) Biodata Validity Decay and Score Inflation with Faking: Do item attributes explain variance across items? *Journal of Business and Psychology*, **16**, 573–592.

Halberstadt, J.B. and Levine, G.M. (1999) Effects of Reasons Analysis on the Accuracy of Predicting Basketball Games. *Journal of Applied Social Psychology*, **29**, 517–530.

Harold, C.M., McFarland, L.A., Dudley, N. and Odin, E.P. (2004, April) *Personality and Faking Behavior: Does warning moderate validity?* Paper presented at the 17th Annual Conference of the Society for Industrial and Organizational Psychology, Chicago, IL.

Heggestad, E.D., Morrison, M., Reeve, C.L. and McCloy, R.A. (2006) Forced-Choice Assessments of Personality for Selection: Evaluating issues of normative assessment and faking resistance. *Journal of Applied Psychology*, **91**, 9–24.

Hogan, R.T. (1991) Personality and Personality Measurement. In: Dunnette, M.D. and Hough, L.M. (eds), *Handbook of Industrial and Organizational Psychology* (2nd edn, Vol. 2). Palo Alto, CA: Consulting Psychologists Press, pp. 873–919.

Holden, R.R., Wood, L.L. and Tomashewski, L. (2001) Do Response Time Limitations Counteract the Effect of Faking on Personality Inventory Validity? *Journal of Personality and Social Psychology*, **81**, 160–169.

Hooper, A.C., Cullen, M.J. and Sackett, P.R. (2006) Operational Threats to the Use of SJTs: Faking, coaching, and retesting issues. In: Weekley, J.A. and Ployhart, R.E. (eds), *Situational Judgment Tests: Theory, measurement and application*. Mahwah, NJ: Lawrence Erlbaum, pp. 205–232.

Jackson, D.N., Wroblewski, V.R. and Ashton, M.C. (2000) The Impact of Faking on Employment Tests: Does forced choice offer a solution? *Human Performance*, **13**, 371–388.

Johnson, J.A. (1981) The 'Self-Disclosure' and 'Self-Presentation' Views of Item Response Dynamics and Personality Scale Validity. *Journal of Personality and Social Psychology*, **40**, 761–769.

Juraska, S.E. and Drasgow, F. (2001, April). *Faking Situational Judgment: A test of the Conflict Resolution Skills Assessment*. Paper presented at the 16th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Kanning, U.P., Grewe, K., Hollenberg, S. and Hadouch, M. (2006) From the Subjects' Point of View – Reactions to different types of situational judgment items. *European Journal of Psychological Assessment*, **22**, 168–176.

Klein, S.B., Loftus, J., Trafton, J.G. and Fuhrman, R.W. (1992) Use of Exemplars and Abstraction in Trait Judgments: A model of trait knowledge about the self and others. *Journal of Personality and Social Psychology*, **63**, 739–753.

Kluger, A.N. and Collela, A. (1993) Beyond the Mean Bias: The effect of warning against faking on biodata item variances. *Personnel Psychology*, **46**, 763–780.

Kuncel, N.R. and Borneman, M.J. (2007) Toward a New Method of Detecting Deliberately Faked Personality Tests: The use of idiosyncratic item responses. *International Journal of Selection and Assessment*, **15**, 220–231.

Lerner, J.S. and Tetlock, P.E. (1999) Accounting for the Effects of Accountability. *Psychological Bulletin*, **125**, 55–275.

Levashina, J. and Campion, M.A. (2006) A Model of Faking Likelihood in the Employment Interview. *International Journal of Selection and Assessment*, **14**, 299–316.

Lievens, F., Buyse, T. and Sackett, P.R. (2005) The Operational Validity of a Video-Based Situational Judgment Test for Medical College Admissions: Illustrating the importance of matching predictor and criterion construct domains. *Journal of Applied Psychology*, **90**, 442–452.

Martin, B.A., Bowen, C.C. and Hunt, S.T. (2002) How Effective are People at Faking on Personality Questionnaires? *Personality and Individual Differences*, **32**, 247–256.

McDaniel, M.A., Hartman, N.S., Whetzel, D.L. and Grubb, W.L. (2007) Situational Judgment Tests, Response Instructions, and Validity: A meta-analysis. *Personnel Psychology*, **60**, 63–91.

McDaniel, M.A., Morgeson, F.P., Finnegan, E.B., Campion, M.A. and Braverman, E.P. (2001) Use of Situational Judgment Tests to Predict Job Performance: A clarification of the literature. *Journal of Applied Psychology*, **86**, 730–740.

McFarland, L.A. and Ryan, A.M. (2000) Variance in Faking Across Non-Cognitive Measures. *Journal of Applied Psychology*, **85**, 812–821.

McFarland, L.A., Ryan, A.M. and Ellis, A. (2002) Item Placement on a Personality Measure: Effects on faking behavior and test measurement properties. *Journal of Personality Assessment*, **78**, 348–369.

McKeachie, W.J., Pollie, D. and Speisman, J. (1955) Relieving Anxiety in Classroom Examinations. *Journal of Abnormal and Social Psychology*, **50**, 93–98.

Mesmer-Magnus, J. and Viswesvaran, C. (2006) Assessing Response Distortion in Personality Tests: A review of research designs and analytic strategies. In: Griffith, R.L. and Peterson, M.H. (eds), *A Closer Examination of Applicant Faking Behavior*. Greenwich, CT: Information Age Publishing, pp. 85–114.

Motowidlo, S.J., Dunnette, M.D. and Carter, G.W. (1990) An Alternative Selection Procedure: The low-fidelity simulation. *Journal of Applied Psychology*, **75**, 640–647.

Mueller-Hanson, R., Heggestad, E.D. and Thornton, G.C. III (2003) Faking and Selection: Considering the use of personality from select-in and select-out perspectives. *Journal of Applied Psychology*, **88**, 348–355.

Nguyen, N.T., McDaniel, M.A. and Biderman, M.D. (2005) Effects of Response Instructions on Faking a Situational Judgment Tests. *International Journal of Selection and Assessment*, **13**, 250–260.

Nield, A.F. and Wintre, M. (1986) Multiple Choice Questions with an Option to Comment: Students' attitude and use. *Teaching of Psychology*, **13**, 196–199.

Ones, D.S. and Viswesvaran, C. (1998) The Effects of Social Desirability and Faking on Personality and Integrity Assessment for Personnel Selection. *Human Performance*, **11**, 245–269.

Ones, D.S., Viswesvaran, C. and Reiss, A.D. (1996) Role of Social Desirability in Personality Testing for Personnel Selection: The red herring. *Journal of Applied Psychology*, **81**, 660–679.

Oswald, F.L., Schmitt, N., Kim, B.H., Ramsay, L.J. and Gillespie, M.A. (2004) Developing a Biodata Measure and Situational Judgment Inventory as Predictors of College Student Performance. *Journal of Applied Psychology*, **89**, 187–207.

Pace, V.L. and Borman, W.C. (2006) The Use of Warnings to Discourage Faking on Noncognitive Inventories. In: Griffith, R.L. and Peterson, M.H. (eds), *A Closer Examination of Applicant Faking Behavior*. Greenwich, CT: Information Age Publishing, pp. 283–304.

Paulhus, D.L. (1991) Measurement and Control of Response Bias. In: Robinson, J.P., Shaver, P.R. and Wrightsman, L.S. (eds), *Measures of Personality and Social Psychological Attitudes*. San Diego, CA: Academic Press, pp. 17–59.

Pauls, C.A. and Crost, N.W. (2004) Effects of Faking on Self-Deception and Impression Management Scales. *Personality and Individual Differences*, **37**, 1137–1151.

Peeters, H. and Lievens, F. (2005) Situational Judgment Tests and their Predictiveness of College Student Success: The influence of faking. *Educational and Psychological Measurement*, **65**, 70–89.

Peterson, G.W., Clark, D.A. and Bennett, B. (1989) The Utility of MMPI Subtle, Obvious Scales for Detecting Fake Good and Fake Bad Response Sets. *Journal of Clinical Psychology*, **45**, 575–583.

Ployhart, R.E., Weekley, J.A., Holtz, B.C. and Kemp, C. (2003) Web-Based and Paper-and-Pencil Testing of Applicants in a Proctored Setting: Are personality, biodata, and situational judgment comparable? *Personnel Psychology*, **56**, 733–752.

Posey, D.F. and Hess, A.K. (1984) The Fakeability of Subtle and Obvious Measures of Aggression by Male Prisoners. *Journal of Personality Assessment*, **48**, 137–144.

Raymond, M.R., Neustel, S. and Anderson, D. (2007) Retest Effects on Identical and Parallel Forms in Certification and Licensure Testing. *Personnel Psychology*, **60**, 367–396.

Reynolds, D.H., Winter, J.L. and Scott, D.R. (1999, May). *Development, Validation, and Translation of a Professional-Level Situational Judgment Inventory*. Paper presented at the 14th Annual Conference of the Society for Industrial and Organizational Psychology, Atlanta, GA.

Rosse, J.G., Stechner, M.D., Levin, R.A. and Miller, J.L. (1998) The Impact of Response Distortion on Preemployment Personality Testing and Hiring Decisions. *Journal of Applied Psychology*, **83**, 634–644.

Rothstein, M.G. and Goffin, R.D. (2006) The Use of Personality Measures in Personnel Selection: What does current research support? *Human Resource Management Review*, **16**, 155–180.

Schleicher, D.J., Venkataramani, V., Morgeson, F.P. and Campion, M.A. (2006) So You Didn't Get the Job ... Now What Do You Think? Examining Opportunity-to-Perform Fairness Perceptions. *Personnel Psychology*, **59**, 559–590.

Schmitt, N. and Kunce, C. (2002) The Effects of Required Elaboration of Answers to Biodata Questions. *Personnel Psychology*, **55**, 569–587.

Schmitt, N. and Oswald, F.L. (2006) The Impact of Corrections for Faking on the Validity of Noncognitive Measures in Selection Settings. *Journal of Applied Psychology*, **91**, 613–621.

Schmitt, N., Oswald, F.L., Kim, B.H., Gillespie, M.A., Ramsay, L.J. and Yoo, T.Y. (2003) Impact of Elaboration on Socially Desirable Responding and the Validity of Biodata Measures. *Journal of Applied Psychology*, **88**, 979–988.

Sedikides, C., Herbst, K.C., Hardin, D.P. and Dardis, G.J. (2002) Accountability as a Deterrent to Self-Enhancement: The search for mechanisms. *Journal of Personality and Social Psychology*, **83**, 592–605.

Sedikides, C., Horton, R.S. and Gregg, A.P. (2007) The Why's the Limit: Curtailing self-enhancement with explanatory introspection. *Journal of Personality*, **75**, 783–824.

Shimberg, B. (1981) Testing for Licensure and Certification. *American Psychologist*, **36**, 1138–1146.

Stark, S., Chernyshenko, O.S., Chan, K., Lee, W. and Drasgow, F. (2001) Effects of the Testing Situation on Item Responding: Cause for concern. *Journal of Applied Psychology*, **86**, 943–953.

Vasilopoulos, N.L., Cucina, J.M., Dyomina, N.V., Morewitz, C.L. and Reilly, R.R. (2006) Forced-Choice Personality Tests: A measure of personality and cognitive ability? *Human Performance*, **19**, 175–199.

Vasilopoulos, N.L., Reilly, R.R. and Leaman, J.A. (2000) The Influence of Job Familiarity and Impression Management on Self-Report Measure Scale Scores and Response Latencies. *Journal of Applied Psychology*, **85**, 50–64.

Viswesvaran, C. and Ones, D.S. (1999) Meta-Analyses of Fakability Estimates: Implications for personality measurement. *Educational and Psychological Measurement*, **59**, 197–210.

Weekley, J.A., Ployhart, R.E. and Holtz, B.C. (2006) Scaling, Scoring, and Developing. Situational Judgment Tests. In Weekley, J.A. and Ployhart, R.E. (eds), *Situational Judgment Tests: Theory, measurement and application*. Mahwah, NJ: Lawrence Erlbaum, pp. 157–182.

Wilson, T.D. and LaFleur, S.J. (1995) Knowing What You'll Do: Effects of analyzing reasons on self-prediction. *Journal of Personality and Social Psychology*, **68**, 21–35.

Wittmaier, B.C. (1976) Low Test Anxiety as a Potential Indicator of Underachievement. *Measurement and Evaluation in Guidance*, **9**, 146–151.