

University of Groningen

Measuring non-cognitive predictors in high-stakes contexts

Niessen, A. Susan M.; Meijer, Rob R.; Tendeiro, Jorge N.

Published in:
 Personality and Individual Differences

DOI:
[10.1016/j.paid.2016.11.014](https://doi.org/10.1016/j.paid.2016.11.014)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
 Publisher's PDF, also known as Version of record

Publication date:
 2017

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Niessen, A. S. M., Meijer, R. R., & Tendeiro, J. N. (2017). Measuring non-cognitive predictors in high-stakes contexts: The effect of self-presentation on self-report instruments used in admission to higher education. *Personality and Individual Differences*, 106, 183-189. <https://doi.org/10.1016/j.paid.2016.11.014>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.



Measuring non-cognitive predictors in high-stakes contexts: The effect of self-presentation on self-report instruments used in admission to higher education



A. Susan M. Niessen ^{*}, Rob R. Meijer, Jorge N. Tendeiro

Department of Psychometrics and Statistics, Faculty of Behavioral and Social Sciences, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands

ARTICLE INFO

Article history:

Received 17 August 2016

Received in revised form 2 November 2016

Accepted 7 November 2016

Available online 12 November 2016

Keywords:

Self-presentation
Self-report measures
Validity
High-stakes testing
Selection
Admission
Higher education

ABSTRACT

Non-cognitive constructs such as personality traits and behavioral tendencies show predictive validity for academic performance and incremental validity over and above cognitive constructs. Therefore, non-cognitive predictors are increasingly used in admission procedures for higher education, typically measured using self-report instruments. It is well known that self-report instruments are sensitive to self-presentation, especially in high-stakes contexts. However, remarkably few studies investigated the effect of self-presentation on predictive validity. The effect of self-presentation in applicants to an undergraduate psychology program was studied using a repeated measures design. Respondents completed self-report questionnaires measuring non-cognitive predictors of academic performance before admission to the program, and again after admission. Scores were compared between contexts, as well as predictive validity, incremental validity, and potential hiring decisions. Results showed differences in scores between contexts on all scales, attenuated predictive validity for most scales, attenuated incremental validity when scores obtained in the admission context were used, and effects on admission decisions. In conclusion, validity results based on scores measured in low-stakes contexts cannot simply be generalized to high-stakes contexts. Furthermore, results obtained in a high-stakes context may result in self-presentation irrespective of whether participants are informed that their scores are used for selection decisions or not.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Non-cognitive characteristics such as personality and work styles are the most commonly assessed constructs in personnel selection (Ryan et al., 2015). With the increasing interest in using non-cognitive predictors in admission procedures to higher education in addition to cognitive predictors, this industry is expanding to the educational field (e.g. Kyllonen, Lipnevich, Burrus, & Roberts, 2014; Kyllonen, Walters, & Kaufman, 2005; Schmitt, 2012). Research has shown that non-cognitive predictors such as personality traits, motivation, and self-regulation are associated with academic performance and show incremental validity over and above cognitive predictors (e.g., Richardson, Abrahams, & Bond, 2012). Furthermore, non-cognitive measures are also used to predict broader outcomes than GPA, like job performance, leadership, and interpersonal skills (Lievens, 2013; Schmitt, 2012). The most common method to assess non-cognitive predictors is through self-report questionnaires. However, many studies have shown that self-report

questionnaires are susceptible to self-presentation behavior (Viswesvaran & Ones, 1999). Very few predictive validity studies of non-cognitive admission instruments have been conducted with *actual* applicants. Data are usually collected for research purposes in low-stakes contexts, where the occurrence of self-presentation behavior is less common than in high-stakes selection contexts.

Self-presentation behavior can be intentional (impression management) or unintentional (self-deception; e.g., Paulhus, 1991; Pauls & Crost, 2004). Since it is difficult to distinguish these two kinds of behavior and we often do not know whether response distortions were deliberate or unconscious, we chose to use the neutral term self-presentation. Self-presentation in self-report questionnaires used for college admissions is rarely investigated. Furthermore, in both the educational literature and in the personnel selection literature there are very few studies that use the recommended (e.g., Ryan & Boyce, 2006) repeated measures design, actual applicants, and representative criterion data (for an exception see Peterson, Griffith, Isaacson, O'Connell, & Mangos, 2011). When self-report questionnaires are used for selection purposes it is important to have an understanding of the size of self-presentation effects on predictor scores, and whether self-presentation behavior affects the validity of predictor scores in operational settings. The aim of this study was to fill this gap and to investigate self-presentation

^{*} Corresponding author.

E-mail addresses: a.s.niessen@rug.nl (A.S.M. Niessen), r.r.meijer@rug.nl (R.R. Meijer), j.n.tendeiro@rug.nl (J.N. Tendeiro).

in non-cognitive questionnaires in a sample of actual college applicants, using a repeated measures design.

1.1. Some findings in the literature

1.1.1. Studies in personnel selection

Many studies concerning self-presentation in self-report instruments have been conducted in the context of personnel selection. In summary, these studies indicated that self-report instruments can easily be faked when respondents are instructed to do so (Viswesvaran & Ones, 1999). Furthermore, in their meta-analysis, Birkeland, Manson, Kisamore, Brannick, and Smith (2006) concluded that applicants showed self-presentation behavior in actual high-stakes selection on all Big Five personality constructs, with the largest effect sizes for Conscientiousness and Emotional stability. Also, there were individual differences in the extent of self-presentation behavior (McFarland & Ryan, 2000; Rosse, Stecher, Miller, & Levin, 1998), which affects the rank-ordering of applicants and influences hiring decisions (Hartman & Grubb, 2011; Rosse, Stecher, Miller, & Levin, 1998). Applicants who show self-presentation tend to rise to the top of the rank order, which can negatively affect the utility of the selection procedure, especially when selection ratios are low (Mueller-Hanson, Heggstad, & Thornton, 2003).

Construct validity is also often affected by self-presentation; instruments measuring the Big Five often yield a sixth 'ideal employee' factor in applicant samples, with high loadings for items that describe desirable personality dimensions (Klehe et al., 2012; Schmit & Ryan, 1993). In addition, based on the literature it is difficult to draw a clear conclusion about the effect of self-presentation on predictive validity (Morgeson, Campion, Dipboye, Hollenbeck, Murphy, & Schmitt, 2007a, 2007b). Based on a meta-analysis correcting scale scores for social desirability, Ones, Viswesvaran, and Reiss (1996) concluded that self-presentation does not affect predictive validity, while other studies did find attenuating effects of self-presentation or test-taking motivation on predictive validity (e.g., O'Neill, Goffin, & Gellatly, 2010; Peterson, Griffith, Isaacson, O'Connell, & Mangos, 2011). However, an important observation is that studies that found an attenuating effect mostly used between-subjects designs with an honest condition and a faking condition, whereas studies that found no effect mostly used one-sample designs and controlled non-cognitive scores for scores on a social desirability scale. Peterson et al. (2011) found that scores on a social desirability scale were not related to applicant faking, so results based on this approach may have underestimated the effect of self-presentation on predictive validity (Griffith & Peterson, 2008).

1.1.2. Studies in educational selection

Many individual studies and meta-analyses have shown that scores on non-cognitive predictors can predict academic performance and have incremental validity over and above cognitive tests scores and high school GPA. In their meta-analysis, Richardson, Abrahams, and Bond (2012) found correlations around $r = 0.30$ between college GPA and Conscientiousness, procrastination, academic self-efficacy, and effort regulation, and correlations of $r \geq 0.50$ between college GPA and performance self-efficacy and grade goal. Such results promote the use of non-cognitive predictors in admission decisions (e.g., Kappe & van der Flier, 2012), and supplementing cognitive tests with non-cognitive questionnaires for admission or matching purposes is increasingly popular (e.g., Kyllonen, Walters, & Kaufman, 2005; Kyllonen, Lipnevich, Burrus, & Roberts, 2014; Schmitt, 2012). However, most predictive validity studies were not conducted in actual admissions contexts, but used volunteers for whom the stakes were low. The question is whether results of such studies can be generalized to high-stakes admission contexts. The literature on assessing non-cognitive predictors, either in personnel selection or in educational selection, does not provide an answer to this question. Furthermore, results based on personnel selection samples may not generalize to educational selection samples. Several

studies have found a positive relationship between cognitive ability and self-presentation score inflation (e.g. Tett, Freund, Christiansen, Fox, & Coaster, 2012; Pauls & Crost, 2004). Given the above average cognitive ability of applicants to higher education they may show more score inflation than applicants in a personnel selection context.

In a study using respondents who were instructed to fake, self-presentation attenuated the predictive validity of GPA for a situational judgment test measuring study-related behavioral tendencies (Peeters & Lievens, 2005). Similar results were found for Big Five personality constructs (Huws, Reddy, & Talcott, 2009), except when an ipsative scoring format was used (Hirsh & Peterson, 2008). However, these studies may overestimate the extent and effect of self-presentation because respondents who were instructed to fake tend to show more score inflation than actual applicants (Birkeland et al., 2006). The only study that used actual applicants instead of instructed self-presentation and a repeated-measures design was Griffin and Wilson (2012). In a sample of medical school applicants, they found higher scores in the high-stakes context than in the low-stakes context for all Big Five personality scales except for Agreeableness. Almost two-thirds of the applicants had higher scores in the selection context than in the research context on at least one subscale, and scores on the Conscientiousness scale showed the largest mean difference between the two settings. However, effects on predictive validity were not examined in this study.

1.1.3. Aim of the present study

So, in spite of the large body of literature about self-presentation, we still do not know if and to what extent self-presentation behavior affects predictive validity in operational contexts. As noted by Peeters and Lievens (2005), results based on participants who were instructed to fake may show the results of a worst-case scenario rather than realistic outcomes. Therefore, the main aim of this study was to investigate the effect of self-presentation on the predictive validity and incremental validity of non-cognitive predictors using actual applicants, using a repeated measures design. The Big Five personality traits, procrastination tendencies, perceived academic skills and academic competence, and grade goal were measured using self-report Likert-format questionnaires. We examined (1) to what extent self-presentation behavior occurred, (2) the effect of self-presentation on the predictive validity of the self-reported non-cognitive predictors, (3) the effect of self-presentation on the incremental validity of the self-reported non-cognitive predictors, and (4) the effect of self-presentation behavior on potential admission decisions and criterion performance of admitted applicants. Agreeableness, Neuroticism, Extraversion, and Openness tend to show no or small relationships with academic performance (Richardson et al., 2012), so they were not included in analyses involving predictive validity, incremental validity, or admission decisions, but we did study if self-presentation behavior occurred on these predictors.

2. Method

2.1. Respondents and procedure

All applicants to an undergraduate psychology program at a Dutch university in the academic year 2014–2015 were invited to complete several questionnaires before admission tests were administered and admission decisions were made. We refer to this measurement as the admission context. The applicants were informed that the aim of filling out these questionnaires was to measure non-cognitive constructs that were related to academic performance, but that their scores would not be used in admission decisions and were collected for research purposes only. Standard instructions for filling out the questionnaires were provided to the respondents. Five months later, after the start of the academic year, all students who completed the questionnaires before admission and who enrolled in the program could voluntarily participate in filling out the questionnaire a second time for course credit. Participants were told that the second administration of the questionnaires

was conducted to study the consistency of responses over time; they did not know that the specific interest was self-presentation, but they were informed about this after completing the questionnaires. We refer to this measurement as the research context. The instructions and questionnaires were identical to the first administration. Both administrations were conducted online via a survey tool.

There were 140 students who filled out the questionnaires at both time-points (21% of all first-year students). The mean age was 19 ($SD = 1.4$), and 81% was female. The proportion of females enrolled in this program is traditionally high, around 70%, thus women were slightly overrepresented in this study. The program consists of an English track and a Dutch track with similar course content. Thirty-one percent of the respondents were enrolled in the Dutch track. Thirty-four percent of the respondents were Dutch, 53% were German, 10% had another European nationality, and 3% had a non-European nationality.

2.2. Measures

The Big Five personality dimensions (Agreeableness, Extraversion, Neuroticism, Conscientiousness, and Openness to Experience) were assessed using the Big Five Inventory (BFI; John, Donahue, & Kentle, 1991). The BFI consists of 44 items with eight to ten items for each dimension. Each item was answered on a five-point Likert-scale (1 = strongly disagree through 5 = strongly agree). Cronbach's alpha ranged from 0.64 (Openness) to 0.84 (Extraversion) when administered in a research context, and from 0.54 (Openness) to 0.78 (Extraversion) when administered in an admission context.

Procrastination tendencies were measured with Lay's Procrastination Scale (Lay, 1986). This scale consists of 20 items and each item was answered on a five-point Likert-scale (1 = never through 5 = all of the time). Cronbach's alpha was 0.82 in the research context and 0.72 in the admission context.

The Study Management and Academic Results Test (Kleijn, van der Ploeg, & Topman, 1994) was used to measure perceived academic skills (academic competence, test competence, time management, and strategic studying). The SMART contains 29 items and each of the four scales consists of 4 to 6 items with a four-point Likert scale (1 = almost never through 4 = nearly always). Cronbach's alpha ranged from 0.60 (strategic studying) to 0.79 (academic competence) in the research context and from 0.51 to (academic competence) to 0.68 (test competence) in the admission context. Grade goal was measured by one item asking for the expected mean grade in the first year of the psychology program on a scale from 1 to 10, with a 5.5 or higher representing a pass.

Furthermore, to study incremental validity, scores on a trial-studying exam were collected. All respondents took this exam as part of the admission procedure to the program and it consisted of questions about two chapters from an introductory book about psychology. This exam was a good predictor of first year academic performance, comparable to high school GPA (for further details see Niessen, Meijer, & Tendeiro, 2016; Meijer & Niessen, 2015).

Data for two relevant criteria were collected: academic performance and dropout during the first year. Academic performance was operationalized as first year GPA (FYGPA), that is, the mean grade across all course-grades obtained in the first year. Respondents provided informed consent and permission to obtain their grades and dropout records through the university administration.

2.3. Data-analysis

To assess whether self-presentation behavior occurred, mean scores for all scales obtained at each administration were computed and dependent samples t-tests were conducted. To investigate how many respondents showed self-presentation behavior, 95% confidence intervals were computed around each individual score obtained during administration in the research context. The confidence intervals were computed around the estimated true score (based on the reliability of

the test, the observed score, and the mean observed score¹; Gulliksen, 1950) using Cronbach's alpha. Respondents who obtained a score above the upper bound for positive predictors or below the lower bound for negative predictors in the admission context were classified as having shown self-presentation behavior. Grade goal was not included in this analysis since it was measured using one item.

In all further analyses, Agreeableness, Extraversion, Neuroticism, and Openness were not included because earlier research showed no or very small relationships with academic performance (e.g., Richardson et al., 2012). These predictors were included in the survey because Big Five questionnaires are usually not administered as separate scales and only administering the Conscientiousness items could affect self-presentation behavior. To assess the effect of self-presentation on predictive validity, correlations were computed between predictor scores obtained in the admission context or in the research context, and FYGPA and dropout. The correlations were compared using t-tests for dependent correlations (Steiger, 1980). The effect on incremental validity for predicting FYGPA was studied by conducting two hierarchical multiple regression analyses. In the first step, scores on a cognitive trial-studying exam were added. In the second step, either the non-cognitive predictor scores obtained in the admission context or the predictor scores obtained in the research context were added.

To investigate the influence of self-presentation on admission decisions we investigated to what extent the same decisions would be made about the applicants based on scores obtained in both contexts. All respondents were ranked according to their scores obtained in both contexts. Because scores on several predictors are often combined in actual admission procedures, and using single scales would result in many ties, we investigated admission decisions based on two composite scores. First, a composite score across all relevant non-cognitive predictors was computed (Conscientiousness, procrastination, the four SMART scales, and grade goal), with equal weights for each predictor. Second, a composite score based on the non-cognitive composite and the score on the trial-studying exam (that was not sensitive to self-presentation), with equal weights for the trial-studying score and the non-cognitive composite score. The scores on each predictor were standardized first. The sign of the standardized procrastination score was reversed, so that a high score on this scale indicated a low tendency to procrastinate and a low score indicated a high tendency to procrastinate. Based on the rankings and using several selection ratios, percentages of incorrectly rejected respondents and percentages of incorrectly admitted respondents were computed, assuming that scores obtained in the research context were true responses. Incorrectly rejected respondents were defined as respondents who would have been rejected based on their rank obtained in the admission context, but would have been accepted based on their rank obtained in the research context. Incorrectly accepted respondents were defined as respondents who would have been accepted based on their rank obtained in the admission context, but would have been rejected based on their rank obtained in the research context.

To study the effects on criterion performance, the mean FYGPA for selected applicants for different hypothetical selection ratios were computed, based on both the non-cognitive composite and the composite of the non-cognitive scores and the trial-studying exam obtained in both contexts.

3. Results

3.1. Occurrence, extent, and prevalence of self-presentation

Descriptive statistics for each predictor in both contexts, t-test results, and effect sizes for the differences in scores between both contexts are presented in Table 1. The differences between mean scores obtained

¹ $CI = (\bar{X} + r_{xx'}(X - \bar{X})) \pm (z^*)(S_X)(\sqrt{1 - r_{xx'}})(\sqrt{r_{xx'}})$

Table 1
Descriptive statistics and differences between scores obtained during the two administrations.

Predictor	Administration		$t_{(139)}$	d	Identified SP
	Admission	Research			
Personality					
Agreeableness	3.71 (0.48)	3.60 (0.50)	3.81*	0.30	20%
Extraversion	3.43 (0.56)	3.25 (0.65)	5.10*	0.46	18%
Neuroticism	2.78 (0.59)	2.92 (0.57)	-3.67*	-0.28	21%
Conscientiousness	3.54 (0.49)	3.34 (0.50)	5.56*	0.47	27%
Openness	3.72 (0.36)	3.56 (0.43)	5.64*	0.49	19%
SMART					
Academic competence	3.51 (0.29)	3.39 (0.44)	3.45*	0.29	21%
Test competence	3.02 (0.40)	2.86 (0.48)	3.89*	0.36	25%
Time management	3.19 (0.52)	3.02 (0.60)	3.28*	0.27	23%
Strategic studying	3.12 (0.46)	2.98 (0.49)	3.31*	0.28	21%
Procrastination	2.94 (0.41)	3.08 (0.48)	-4.05*	-0.36	25%
Grade goal	7.28 (1.01)	7.06 (0.66)	2.55*	0.24	

Note. Means with standard deviations between brackets. d = Cohen's d . Identified SP = percentage identified as having shown self-presentation.

* $p < 0.05$.

in each context are all in the direction that was expected when self-presentation occurred. When comparing scores obtained in the admission context to scores obtained in the research context, it is clear that respondents scored higher on positively evaluated predictors like Conscientiousness and test competence, and lower on the negatively evaluated predictors Neuroticism and procrastination tendencies. All differences were statistically significant, with small to moderate effect sizes in terms of Cohen (1988). Thus, these findings suggested that self-presentation behavior occurred when the questionnaires were administered in an admission context.

Self-presentation behavior also occurred on the Big Five scales that are not or only weakly related to academic performance (Agreeableness, Extraversion, Neuroticism, and Openness), but to a slightly lesser extent. Table 1 shows the percentage of respondents that showed self-presentation for each scale, based on the 95% confidence intervals around their scores obtained in the research context, compared to their scores obtained in the admission context. Conscientiousness, test competence, and procrastination showed the largest percentages of self-presentation, but the differences in the prevalence of self-presentation between the subscales were small. Seventy-three percent of all respondents were identified as having shown self-presentation on at least one scale.

3.2. Self-presentation and predictive validity

Table 2 presents the correlations between predictor scores obtained in both contexts and the criteria. In the admission context, only Conscientiousness had a significant but small positive correlation with FYGPA. For scores obtained in the research context, test competence, time management showed small positive correlations with FYGPA, and

procrastination showed small negative correlations with FYGPA. Conscientiousness showed a moderate positive correlation with FYGPA, and grade goal showed a large positive correlation with FYGPA. For predicting dropout, none of the predictors showed a significant correlation when obtained in the admission context, whereas Conscientiousness and grade goal showed significant small to moderate negative correlations when obtained in the research context. In addition, there were significant differences between the correlations in the administration contexts for five of the seven predictors for FYGPA and four of the seven predictors for drop out. These results showed that administering these self-report scales in an admission context significantly attenuated the predictive validity of most predictors.

3.3. Self-presentation and incremental validity

Hierarchical multiple regression analyses showed that the trial-studying exam alone predicted FYGPA with $R^2 = 0.20$ ($F_{(1138)} = 34.31$, $p < 0.01$). Adding the non-cognitive predictor scores obtained in the admission context in a second step yielded $\Delta R^2 = 0.06$ ($F_{(7131)} = 1.49$, $p = 0.18$, model $R^2 = 0.26$, $R^2_{adj.} = 0.21$). Adding the non-cognitive predictor scores obtained in the research context in a second step yielded $\Delta R^2 = 0.19$ ($F_{(7131)} = 5.76$, $p < 0.01$, model $R^2 = 0.39$, $R^2_{adj.} = 0.35$). Thus, in the admission context, the non-cognitive predictors added little explained variance over and above the trial-studying exam, but in the research context they explained an additional 19% of the variance in FYGPA.

3.4. Self-presentation and admission decisions

Table 3 shows the percentages of incorrectly rejected respondents and incorrectly accepted respondents for several selection ratios, assuming that scores obtained in the admission context were true responses. When a composite of the non-cognitive predictors would be used to select the respondents and the selection ratio is low, substantial proportions of respondents would be incorrectly admitted based on their scores obtained in the admission context. When the selection ratios are high, fewer respondents would be incorrectly accepted, but a substantial proportion of the rejected respondents would be incorrectly rejected. For example, when 10% of the applicants would be selected based on their rank obtained in the admission context, 57% of the selected applicants would not have been selected based on their rank obtained in the research context. When 80% of the applicants would be selected based on their rank obtained in the admission context, 61% of the rejected applicants would not have been rejected based on their rank obtained in the research context. Even when the non-cognitive composite scores were combined with a test score that was not susceptible to self-presentation, this result was still observed, but percentages of incorrectly rejected and incorrectly accepted applicants were smaller. Thus, self-presentation would affect the rank ordering of respondents and, as a result, would affect admission decisions.

Table 2
Correlations between the criterion variables and the predictor scores obtained during the two administrations.

Predictor	FYGPA			Dropout		
	Administration		$t_{(137)}$	Administration		$t_{(137)}$
	Admission	Research		Admission	Research	
Conscientiousness	0.21*	0.31*	-1.43	-0.02	-0.17*	2.08*
Academic competence	-0.03	0.13	-1.68*	0.15	-0.12	2.89*
Test competence	0.02	0.19*	-1.99*	-0.03	-0.08	0.57
Time management	-0.01	0.17*	-1.93*	0.06	-0.02	0.84
Strategic studying	-0.07	0.14	-2.34*	0.14	-0.06	2.23*
Procrastination	-0.08	-0.20*	1.62	-0.04	-0.06	0.27
Grade goal	0.05	0.49**	-5.51*	-0.05	-0.27**	2.46*

* $p < 0.05$.

Table 3
Percentage of incorrectly admitted and incorrectly rejected students for different predictors under different selection ratios.

SR	Non-cognitive composite		Non-cognitive/test composite	
	IA%	IR%	IA%	IR%
0.05	100	5	43	2
0.10	57	6	36	4
0.20	46	12	25	6
0.30	43	18	26	11
0.40	38	25	18	12
0.50	31	31	17	17
0.60	27	41	16	23
0.70	21	50	8	19
0.80	15	61	5	21
0.90	9	79	3	29

Note. SR = selection ratio. IA% is the percentage of incorrectly admitted respondents (as a percentage of all admitted respondents given the selection ratio). IR% is the percentage of incorrectly rejected respondents (as a percentage of all rejected respondents given the selection ratio). IA% and IR% are not independent, $IR\% = IA\% \cdot (SR/1-SR)$, and $IA\% = IR\% \cdot (1-SR/SR)$.

3.5. Self-presentation and criterion performance

If self-presentation behavior negatively affects predictive validity and admission decisions, it may also negatively affect the criterion performance of admitted applicants. This hypothesis was explored by looking at the mean criterion performance of admitted applicants for several selection ratios, based on the non-cognitive composite scores and the non-cognitive and trial-studying test composite scores, in both contexts. Fig. 1 shows that overall applicants selected based on scores obtained in a research context performed better than applicants selected based on scores obtained in the admission context, as was expected based on the validity results discussed earlier. These differences were smaller when a composite including the trial-studying test was used. What stands out is the distribution of the FYGPA's across selection ratios. Theoretically, stricter selection leads to higher criterion performance of the selected group, given that the validity coefficient of the selection criterion is positive. So, we would expect a descending trend in all graphs. Although the trends based on scores obtained in the research context show some deviations, they roughly reflect the expected ordering. However, when inspecting the composite scores based on scores obtained in the admission contexts, the stricter selection was not

associated with better criterion performance. For very low selection ratios, mean FYGPA was equal to or even below the level of very lenient selection. So, an important message is that using instruments that allow self-presentation can negatively affect the utility of admission procedure.

4. Discussion

The results showed that scores on non-cognitive predictors were inflated when obtained in an admission context compared to scores obtained in a research context, but with smaller effect sizes than often found in instructed faking studies (Birkeland et al., 2006). On each scale, approximately 20% of the respondents significantly inflated their scores in the admission context. Also, the reliabilities of the scales were often lower in the admission context than in the research context. This shows that the same respondents behaved differently on non-cognitive questionnaires depending on the context in which these questionnaires were administered. Furthermore, the non-cognitive predictors showed substantial predictive validity and incremental validity over and above the trial-studying test when measured in a low-stakes research context, as expected based on the results discussed in existing studies (e.g. Richardson et al., 2012). However, while score inflation in the admission context was not large, the predictive validity and incremental validity were attenuated for most scales when scores were obtained in the admission context. The most striking difference was found for grade goal, which did not significantly predict FYGPA or dropout when measured in the admission context, but showed a high positive correlation with FYGPA and a moderate negative correlation with dropout when measured in a research context. An explanation for the seemingly contradictory findings of this attenuation effect on the one hand and the small to moderate differences between effect sizes across contexts on the other, may be that self-presentation behavior is not restricted to score inflation (Kiefer & Benit, 2016). For instance, an applicant who perceives an item as irrelevant for the position they are applying to may choose a neutral response option instead of an option that reflects the applicant's characteristics, changing the score without necessarily inflating it (Ziegler, 2011). Finally, admission decisions based on a composite score of the non-cognitive predictors, or on a composite score also including a non-self-report exam, were affected by self-presentation behavior. This showed the possible impact of using self-presentation measures on individual applicants. Applicants can be rejected as a result of other applicant's self-presentation behavior, leading to fairness issues. In addition, this can also lead to decreased criterion performance

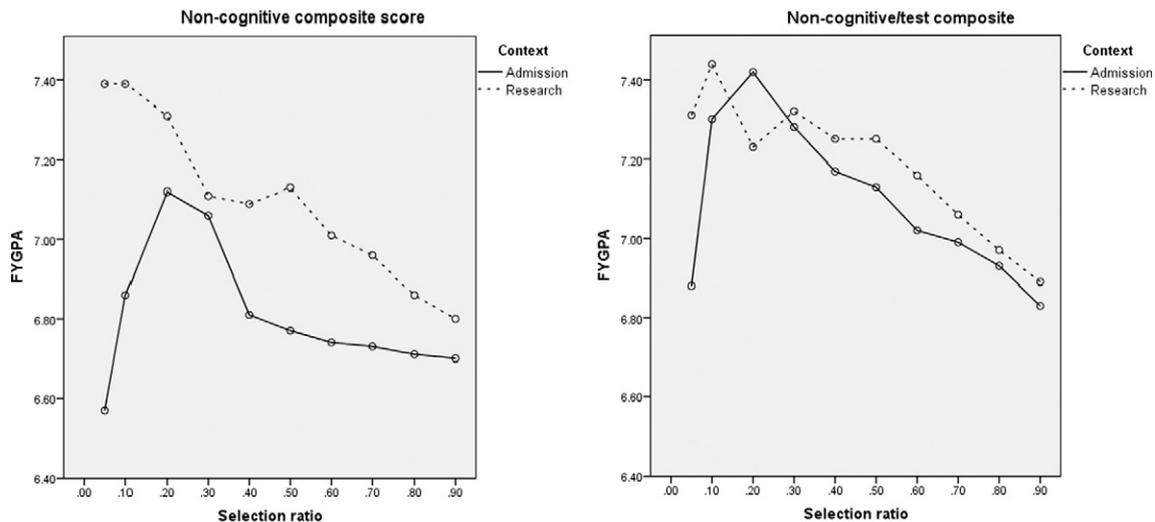


Fig. 1. Mean criterion performance (FYGPA) for each hypothetical selection ratio, for predictor composites obtained in the admission context and in the research context.

in the selected group, making the utility of the selection procedures smaller or even negative, especially when selection ratios are low.

4.1. Practical solutions for self-presentation

Our results confirmed that certain personality traits and behavioral tendencies predict academic performance, and have incremental validity over and above cognitive predictors, such as scores on an exam. However, these results cannot simply be applied and generalized to selection practices and high-stakes assessment, as some studies have implied (e.g., Chamorro-Premuzic & Furnham, 2003; Schmitt, 2012).

Self-report measures using Likert-scales may be the most 'fakable' instruments available, and there are some methods that may be used to reduce self-presentation behavior. Examples are providing warnings that self-presentation can be detected (e.g. Burns, Filipowski, Morris, & Shoda, 2015), using other-ratings rather than self-reports (Ziegler, Danay, Schölmerich, & Bühner, 2010), indirect measures, such as conditional reasoning tests (James, 1998) and, often mentioned as the most viable solution, using forced-choice or ipsative item formats (Hirsh et al., 2008; O'Neill et al., in press; Stark et al., 2014). However, these methods remain understudied in actual applicant samples and thus far results are far from unanimous. Respondents tend to show the same response biases in other ratings as in self-rating (Brown, 2016), and indirect measures were found susceptible to faking when their purpose was disclosed (Bowler & Bowler, 2014). O'Neill et al. (in press) compared actual applicants' scores to honest respondents' on a Likert-format personality questionnaire and a forced-choice personality questionnaire and found differences with medium to large effects sizes for both instruments.

Furthermore, developing a psychometrically sound instrument based on forced-choice items takes a lot of time and effort. The development and research of forced-choice instruments is still preliminary. These instruments might contribute to a solution to the self-presentation problem, but more research is needed before such conclusions can be drawn and we are certainly not there yet (Brown, 2016). Another risk is that, as Kyllonen et al. (2005, p. 177) discussed, "The coaching industry would quickly crack the code on any non-cognitive measure and translate that into a successful strategy for defeating the measure if it were a self-assessment". This could lead to even more problematic results. For instance, Zickar and Robie (1999) found that respondents who were coached to fake showed more score inflation than respondents who were instructed to fake, but not coached.

4.2. Limitations

Our sample was obtained from one cohort of psychology students and female applicants were overrepresented. Therefore, we encourage replication studies that use larger and more representative samples from educational and personnel selection contexts. Furthermore, only applicants who enrolled in the program were asked to participate in the second administration. It would have been valuable to include applicants who decided not to enroll. Furthermore, an alternative explanation for our findings is that the applicants actually changed, for example, with respect to their motivation between both measurements, as a result of their transition to higher education. However, given the direction of the change (lower score on desirable traits and higher scores on undesirable traits after starting higher education), this seems unlikely and is also inconsistent with findings in the literature about changes in personality across the lifespan (Roberts, Walton, & Viechtbauer, 2006). It could also be possible that scores obtained during the second measurement were better predictors because the participants gained more insight in their own characteristics and behavior after starting higher education, and were able to provide more accurate self-assessments. This would, however, yield the same conclusion about the predictive validity of self-reported non-cognitive predictors in admission contexts.

In this study, applicants completed the questionnaires for research purposes in both administrations and the scores obtained in the admissions context were not used to make admission decisions. This may not be entirely representative for situations where such questionnaires are actually used to select students. This implies that the self-presentation effect will likely to be at least as strong when these questionnaires are actually used for selection. Note that we assumed that completing these instruments in an applicant context, even when it was clearly communicated that it was for research purposes, would induce self-presentation and, therefore served as a proxy for high-stakes assessment. However, Stark et al. (2014) assumed that such a situation serves as a proxy for low-stakes assessment. Stark et al. (2014) studied self-presentation in forced-choice personality tests developed for the U.S. military by comparing scores obtained by applicants for selection purposes to scores obtained by applicants for research purposes. They found small differences between scores obtained in both groups. The results of our study indicated that responses collected for research purposes in an applicant situation resemble those typical for a high-stakes context, and may not be suitable to serve as a proxy for low-stakes assessment in self-presentation research. An important alternative explanation for the small differences found by Stark et al. (2014), is that self-presentation occurred in both samples, opposed to concluding that the instrument was 'self-presentation proof'.

4.3. Conclusion

In conclusion, personality traits and behavioral tendencies are vital in understanding performance in various domains, including academic performance. Therefore, it is appealing to include these variables when making high-stakes decisions, such as admission decisions. However, we conclude that at the moment there are no undisputed valid ways to measure these constructs validly in high-stakes contexts. Ignoring the context when using self-report instruments in high-stakes selection can have negative effects with respect to fairness and utility. In future research, authors should explicitly state the intended use of non-cognitive self-report instruments, and most importantly, conduct validation research in the intended context of use. Such a practice would aid the generalizability and replicability of the findings in our field.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

- Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment*, 14, 317–335. <http://dx.doi.org/10.1111/j.1468-2389.2006.00354.x>.
- Bowler, J. L., & Bowler, M. C. (2014). Evaluating the fakability of a conditional reasoning test of addiction proneness. *International Journal of Psychology*, 49, 415–419. <http://dx.doi.org/10.1002/ijop.12030>.
- Brown, A. (2016, July). *Response distortions in self-reported and other-reported measures: Is there light at the end of the tunnel? Paper presented at the 10th International Test Commission Conference, Vancouver, Canada.*
- Burns, G. N., Filipowski, J. N., Morris, M. B., & Shoda, E. A. (2015). Impact of electronic warnings on online personality scores and test-taker reactions in an applicant simulation. *Computers in Human Behavior*, 48, 163–172. <http://dx.doi.org/10.1016/j.chb.2015.01.051>.
- Chamorro-Premuzic, T., & Furnham, A. (2003). Personality predicts academic performance: Evidence from two longitudinal university samples. *Journal of Research in Personality*, 37, 319–338. [http://dx.doi.org/10.1016/S0092-6566\(02\)00578-0](http://dx.doi.org/10.1016/S0092-6566(02)00578-0).
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Griffin, B., & Wilson, I. G. (2012). Faking good: Self-enhancement in medical school applicants. *Medical Education*, 46, 485–490. <http://dx.doi.org/10.1111/j.1365-2923.2011.04208.x>.
- Griffith, R. L., & Peterson, M. H. (2008). The failure of social desirability measures to capture applicant faking behavior. *Industrial and organizational psychology: Perspectives*

- on science and practice. 1. (pp. 308–311). <http://dx.doi.org/10.1111/j.1754-9434.2008.00053.x>.
- Gulliksen, H. (1950). *Theory of mental tests*. New York: John Wiley & Sons.
- Hartman, N. S., & Grubb, W. I. (2011). Deliberate faking on personality and emotional intelligence measures. *Psychological Reports, 108*, 120–138. <http://dx.doi.org/10.2466/03.09.28.PR0.108.1.120-138>.
- Hirsh, J. B., & Peterson, J. B. (2008). Predicting creativity and academic success with a “fake-proof” measure of the big five. *Journal of Research in Personality, 42*, 1323–1333. <http://dx.doi.org/10.1016/j.jrp.2008.04.006>.
- Huws, N., Reddy, P. A., & Talcott, J. B. (2009). The effects of faking on non-cognitive predictors of academic performance in university students. *Learning and Individual Differences, 19*, 476–480. <http://dx.doi.org/10.1016/j.lindif.2009.04.003>.
- James, L. R. (1998). Measurement of personality via conditional reasoning. *Organizational Research Methods, 1*, 131–163. <http://dx.doi.org/10.1177/109442819812001>.
- John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The big five inventory—versions 4a and 5a*. Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.
- Kappe, R., & van der Flier, H. (2012). Predicting academic success in higher education: What’s more important than being smart? *European Journal of Psychology of Education, 27*, 605–619. <http://dx.doi.org/10.1007/s10212-011-0099-9>.
- Kiefer, C., & Benit, N. (2016). What is applicant faking behavior? A review on the current state of theory and modeling techniques. *Journal of European Psychology Students, 7*, 9–19 (doi: 10.5334/jeps.345).
- Klehe, U., Kleinmann, M., Hartstein, T., Melchers, K. G., König, C. J., Heslin, P. A., & Lievens, F. (2012). Responding to personality tests in a selection context: The role of the ability to identify criteria and the ideal-employee factor. *Human Performance, 25*, 273–302. <http://dx.doi.org/10.1080/08959285.2012.703733>.
- Kleijn, M. C., van der Ploeg, H. M., & Topman, R. M. (1994). Cognition, study habits, test anxiety, and academic performance. *Psychological Reports, 75*(12), 19–1226. <http://dx.doi.org/10.2466/pr0.1994.75.3.1219>.
- Kyllonen, P. C., Lipnevich, A. A., Burrus, J., & Roberts, R. D. (2014). Personality, motivation, and college readiness: A prospectus for assessment and development (ETS Report No. 14–06). (Retrieved from) <http://onlineibrary.wiley.com/doi/10.1002/ets2.12004/pdf>
- Kyllonen, P. C., Walters, A. M., & Kaufman, J. C. (2005). Noncognitive constructs and their assessment in graduate education: A review. *Educational Assessment, 10*, 153–184. http://dx.doi.org/10.1207/s15326977ea1003_2.
- Lay, C. (1986). At last, my research article on procrastination. *Journal of Research in Personality, 20*, 474–495. [http://dx.doi.org/10.1016/0092-6566\(86\)90127-3](http://dx.doi.org/10.1016/0092-6566(86)90127-3).
- Lievens, F. (2013). Adjusting medical school admission: Assessing interpersonal skills using situational judgement tests. *Medical Education, 47*, 182–189. <http://dx.doi.org/10.1111/medu.12089>.
- 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>.
- Meijer, R. R., & Niessen, A. S. M. (2015). A trial studying approach to predict college achievement. *Frontiers in Psychology, 6*, 1–3. <http://dx.doi.org/10.3389/fpsyg.2015.00887>.
- Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007a). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology, 60*, 683–729. <http://dx.doi.org/10.1111/j.1744-6570.2007.00089.x>.
- Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007b). Are we getting fooled again? Coming to terms with limitations in the use of personality tests for personnel selection. *Personnel Psychology, 60*, 1029–1049. <http://dx.doi.org/10.1111/j.1744-6570.2007.00100.x>.
- Mueller-Hanson, R., Heggstad, E. D., & Thornton, G. I. (2003). Faking and selection: Considering the use of personality from select-in and select-out perspectives. *Journal of Applied Psychology, 88*, 348–355. <http://dx.doi.org/10.1037/0021-9010.88.2.348>.
- Niessen, A. S. M., Meijer, R. R., & Tendeiro, J. N. (2016). Predicting performance in higher education using proximal predictors. *PLoS ONE, 11*. <http://dx.doi.org/10.1371/journal.pone.0153663>.
- O’Neill, T. A., Goffin, R. D., & Gellatly, I. R. (2010). Test-taking motivation and personality test validity. *Journal of Personnel Psychology, 9*, 117–125. <http://dx.doi.org/10.1027/1866-5888/a000012>.
- O’Neill, T. A., Lewis, R. J., Law, S. J., Larson, N., Hancock, S., Radan, J., ... Carswell, J. J. (2016). Forced-choice pre-employment personality assessment: Construct validity and resistance to faking. *Personality and Individual Differences, 106*. <http://dx.doi.org/10.1016/j.paid.2016.03.075> (in press).
- Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81*, 660–679. <http://dx.doi.org/10.1037/0021-9010.81.6.660>.
- Pauls, C. A., & Crost, N. W. (2004). Effects of faking on self-deception and impression management scales. *Personality and Individual Differences, 37*, 1137–1151. <http://dx.doi.org/10.1016/j.paid.2003.11.018>.
- Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, & P. R. Shaver (Eds.), *Measures of personality and social psychological attitudes*. Vol. 1. (pp. 17–59). San Diego, United States: Academic Press.
- Peeters, H., & Lievens, F. (2005). Situational judgment tests and their predictiveness of college students’ success: The influence of faking. *Educational and Psychological Measurement, 65*, 70–89. <http://dx.doi.org/10.1177/0013164404268672>.
- Peterson, M. H., Griffith, R. L., Isaacson, J. A., O’Connell, M. S., & Mangos, P. M. (2011). Applicant faking, social desirability, and the prediction of counterproductive work behaviors. *Human Performance, 24*, 270–290. <http://dx.doi.org/10.1080/08959285.2011.580808>.
- Richardson, M., Abrahams, C., & Bond, R. (2012). Psychological correlates of university students’ academic performance: A systematic review and meta-analysis. *Psychological Bulletin, 138*, 353–387. <http://dx.doi.org/10.1037/a0026838>.
- Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin, 132*, 1–25. <http://dx.doi.org/10.1037/0033-2909.132.1.1>.
- Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology, 83*, 634–644 (doi: 021-9010/98/j3.00).
- Ryan, A. M., & Boyce, A. S. (2006). What do we know and where do we go? Practical directions for faking research. In R. L. Griffith, & M. H. Peterson (Eds.), *A closer examination of applicant faking behavior* (pp. 357–371). Greenwich, CT: Information Age.
- Ryan, A. M., Inceoglu, I., Bartram, D., Golubovich, Y., Grand, J., Reeder, M., ... Yao, X. (2015). Trends in testing: Highlights of a global survey. In I. Nikolaou, & J. K. Oostrom (Eds.), *Employee recruitment, selection, and assessment: Contemporary issues for theory and practice* (pp. 136–153). New York, NY: Psychology Press/Taylor & Francis.
- Schmit, M. J., & Ryan, A. M. (1993). The big-5 in personnel-selection: Factor structure in applicant and non-applicant populations. *Journal of Applied Psychology, 78*, 966–974.
- Schmitt, N. (2012). Development of rationale and measures of non-cognitive college student potential. *Educational Psychologist, 47*, 18–29. <http://dx.doi.org/10.1080/00461520.2011.610680>.
- Stark, S., Chernyshenko, O. S., Drasgow, F., Nye, C. D., White, L. A., Heffner, T., & Farmer, W. L. (2014). From ABLE to TAPAS: A new generation of personality tests to support military selection and classification decisions. *Military Psychology, 26*, 153–164. <http://dx.doi.org/10.1037/mil0000044>.
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87*, 45–51. <http://dx.doi.org/10.1037/0033-2909.87.2.245>.
- Tett, R. P., Freund, K. A., Christiansen, N. D., Fox, K. E., & Coaster, J. (2012). Faking on self-report emotional intelligence and personality tests: Effects of faking opportunity, cognitive ability, and job type. *Personality and Individual Differences, 52*, 195–201. <http://dx.doi.org/10.1016/j.paid.2011.10.017>.
- Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement, 59*, 197–210. <http://dx.doi.org/10.1177/00131649921969802>.
- Zickar, M. J., & Robie, C. (1999). Modeling faking good on personality items: An item-level analysis. *Journal of Applied Psychology, 84*, 551–563. <http://dx.doi.org/10.1037/0021-9010.84.4.551>.
- Ziegler, M. (2011). Applicant faking: A look into the black box. *The industrial and organizational psychologist, 49*. (pp. 29–36).
- Ziegler, M., Danay, E., Schölmerich, F., & Bühner, M. (2010). Predicting academic success with the big 5 rated from different points of view: Self-rated, other rated and faked. *European Journal of Personality, 24*, 341–355. <http://dx.doi.org/10.1002/per.753>.