Psychology Science, Volume 49, 2007 (4), p. 291-307

Fakability of different measurement methods for achievement motivation: questionnaire, semi-projective, and objective

MATTHIAS ZIEGLER[1], LOTHAR SCHMIDT-ATZERT[2], MARKUS BÜHNER[1] & STEFAN KRUMM[2]

Abstract

Different means can be applied to assess noncognitive personality aspects: projective, semi-projective, self-report, and objective. However, so far little attention has been paid towards the different fakability of these methods. The present study investigated this question with different achievement motivation instruments. The instruments were randomly administered to three student groups: fake bad (n = 41), fake good (n = 37), and control group (n = 41). The faking groups were given specific faking instructions while the control group only received the standard instructions. All instruments were applied computer-assisted. The results show that all tests are fakeable with the exception of the objective measure which could not be faked good as was expected. The effect sizes (d) ranged from .10 to 2.36. Cut-off scores for the detection of faking were computed based on sensitivity as well as specificity. Moreover, they were tested within a second student sample (n = 123). Sensitivity and specificity values are reported. The practical implications for test authors and practitioners are discussed.

Key words: self-rating, objective test, semiprojective test, achievement motivation, faking

[1] Department of Psychology, Ludwig-Maximilians-University of Munich, Germany. Correspondence concerning this article should be addressed to: Matthias Ziegler, MD, Ludwig-Maximilians-University Munich, Leopoldstrasse 13, 80802 Munich, Germany; E-mail: ziegler@psy.uni-muenchen.de; Phone: (+49) (0) 89 / 21 80 6066; Fax: (+49) (0) 89 / 21 80 3000.

[2] Department of Psychology, University of Marburg, Germany

A lot of research exists which demonstrates the effects of faking on noncognitive measures (e.g. Bradley & Hauenstein, 2006; Furnham, 1997; Khorramdel & Kubinger, 2006; Marcus, 2006; Mueller-Hanson, Heggestad, & Thornton, 2003, 2006; Ones, Viswesvaran, Dilchert, & Deller, 2006; Pauls & Crost, 2004; Ramsay, Schmitt, Oswald, Kim, & Gillespie, 2006; Rosse, Stecher, Miller, & Levin, 1998; Topping & O'Gorman, 1997; Viswesvaran & Ones, 1999). However, most of the research used self-reports. While these results are very informative they overlook that noncognitive personality aspects can also be assessed with other measures: projective, semi-projective, and objective. Therefore, the present study investigated in how far these methods differ in regard to their fakability, and how people rate the subjective ease of faking. Moreover, we tried to use cut-off scores to detect faking. There are only few constructs which can be assessed with all of these methods. We used achievement motivation. In the following sections a short overview of achievement motivation and results from faking research will be given before the hypotheses will be presented.

Achievement Motivation. One of the first researchers who displayed an interest in achievement motivation was Henry Murray (1938). His definition of need for achievement was: "To accomplish something difficult. To master, manipulate or organize physical objects, human beings, or ideas. To do this as rapidly and as independently as possible. To overcome obstacles and attain a high standard. To excel one's self. To rival and surpass others. To increase self-regard by the exercise of talent." (Murray, 1938, p. 164). While this definition was broad and covered a wide range of human behaviour, newer definitions are smaller. Cassidy and Lynn (1989) define achievement motivation in general as the personal striving of individuals to attain goals within their social environment. According to Spinath (2001) it comprises such dimensions as need for or pursuit of excellence, work ethic, setting and meeting goals, competitiveness, and status aspiration. McClelland, Koestner, and Weinberger (1989) differentiated between implicit and explicit motives. Implicit motives can only be assessed using projective or objective tests and supposedly predict spontaneous behavioral trends over time. Explicit motives on the other hand can be measured with self-reports and predict immediate specific responses to specific situations or choice behaviour (for an extensive overview see McClelland et al., 1989). Research focused mostly on maximizing reliability and validity of these tests (Spangler, 1992) because such parameters provide a good basis for predicting behavior. The prediction of behavior is especially relevant in a personnel selection context. A large body of evidence exists, showing intelligence as one of the best predictors for job performance (Schmidt & Hunter, 1998). Nevertheless, if there were only slight differences between applicants' intelligence, achievement motivation could play an important role: In his classic study about gifted children, Terman (1959, p. 148) found that differences in achievement motivation, among other aspects, can explain why some gifted people are more successful than others. Although this finding was not derived from a selection context, it implies that assessing achievement motivation (in addition to cognitive ability) might be of practical use.

Faking. When talking about faking one needs to differentiate social desirability, self and other deception, and impression management. These topics have been studied extensively (e.g., Moorman & Podsakoff, 1992; Paulhus, 2002; Pauls & Crost, 2004; Marcus, 2006; Mueller-Hanson et al., 2006; Ones et al., 2006). A discussion of the different results and views is clearly beyond the scope of this paper, and, moreover the focus of this paper is on differential fakability of assessment tools and not on interindividual differences in faking

behavior. Rogers (1997, p. 12) defines deception (faking) as: ”(…) an all-encompassing term to describe any and all attempts by an individual to distort or misrepresent his or her self-reporting”. However, as the definition by Rogers shows, faking has two faces. While social desirability only deals with attempts to appear positive, faking can be distinguished: Overreporting (fake good) and underreporting (fake bad) meaning that people present themselves better or worse than they actually are. An important point which has to be noted here is the difference observed between results from faking studies using students versus real applicants. Effect sizes tend to be smaller (Hough et al., 1990) in the latter samples (for a slightly different view see Griffith, Chmielowski, & Yoshita, 2007). Thus, results from the present study should not be generalized to applicant settings.

Recent findings indicate a change in the construct validity of personality questionnaires due to faking, when self-report measures are used in selection scenarios because correlations between the assessed traits increase (Pauls & Crost, 2005; Schmit & Ryan, 1993). Thus, one could argue that faked personality questionnaires no longer assess personality. Some would argue they assess an ideal employee factor (Schmit & Ryan, 1993). The idea behind the ideal employee factor is that a sixth personality factor emerges when people fake. This factor includes the attributes an ideal employee would display. Other researchers see evidence for the influence of an ability to recognize the dimensions assessed (Hogan, Hogan, & Roberts, 1996; Kleinmann, 1993). This ability supposedly helps to identify which attributes future employers seek. Moreover, it also helps to be successful in the job later on. However, the predictive validity of personality questionnaires does not change in selection scenarios.

Within the present study we used achievement motivation instruments. Considering that achievement motivation is extensively measured by self-reports, it is reasonable to assume that the mentioned impact on construct validity might occur here as well. In contrast to personality measures, which are mainly questionnaires, achievement motivation has the advantage that other measurement methods can be applied as well. Thus, if someone is interested in measuring “real” achievement motivation it should be of great practical interest to find out which of the commonly used measures make faking impossible or at least detectable. Consequently, the present study compares the fakability of a semi-projective, a self-report, and an objective achievement motivation test.

Assessment techniques. Self-reports are subjective measures typically applied in questionnaire form (Murphy & Davidshofer, 2001, p. 7 ff.). Projective tests represent the oldest method of measuring achievement motivation. Following the tradition of Murray’s (1943) Thematic-Apperception-Test (TAT), these tests consist of ambiguous stimuli - mostly pictures. The task is to tell a creative story about the given picture (see Heckhausen, 1991). It has been argued that such techniques lack reliability (Entwisle, 1972) and we therefore decided against using projective techniques. In terms of adequate reliability and objectivity, semi-projective measures are a better choice and were therefore preferred. As with projective tests, ambiguous stimuli are presented. However, instead of writing a story, participants have to rate statements with regard to the pictures. Semi-projective tests are also called grid technique (Schmalt, 1999; Sokolowski, Schmalt, Langens, & Puca, 2000). According to Murphy and Davidshofer (2001, p. 7), “(…) objective tests are seen as containing highly structured, clear, unambiguous test items that are objectively scored.” Cattell (1965) suggested that it would be useful to assess the human personality using objective measures. Examples of objective personality tests are Kubinger and Ebenhöh’s „Arbeitshaltungen [work-attitudes; Kubinger & Ebenhoeh, 1996]” and Schmidt-Atzert’s “Objektiver

Leistungsmotivationstest [Objective achievement motivation test, OLMT; Schmidt-Atzert, 2004]".

Dealing with faking. Numerous approaches exist to ensure the detection of faking in performance tests. Many of them stem from neuropsychological research on clinical samples (e.g., Franzen, 2000; Steck, Reuter, Meir-Korrell, & Schönle, 2000), although non-clinical samples have been studied as well (Schmidt-Atzert, Bühner, Rischen, & Warkentin, 2004). The results typically show that it is nearly impossible to fake good on a performance test. When people try to fake bad, they often underestimate a poor performance and deliver results, which make it easy to spot faking bad. This occurs because they have no frame of reference for what is a good or a poor performance. Similar results have been found in faking studies related to brain damage (Vickery et al., 2004). These studies revealed that patients were not able to fake their results reasonably and produced results that were worse than those by neuropsychological injured and even worse than chance.

An often used approach to detect faking in personality questionnaires is to incorporate control or lie scales as it is done in the MMPI (Hathaway & McKinley, 2000). However, such scales cannot be the ultimate solution because: (1) they can be faked as well (Viswesvaran & Ones, 1999); (2) one can never be sure that high faking scores are truly due to faking; and as a consequence (3) one does not exactly know what to do with persons scoring high on a faking scale. Several test authors tried to deal with faking by using forced choice formats (Christiansen, Burns, & Montgomery, 2005). Studies show small successes (Jackson, Wroblewski, & Ashton, 2000; Martin, Bowen, & Hunt, 2002). Unfortunately, forced choice items cannot be used to compare persons (Baron, 1996).

Summing up, faking in performance test is either impossible or easy to spot. However, personality assessment has not been very successful in detecting faking so far. Therefore, the aim of the present study was not only to explore the fakability of three different approaches to measuring achievement motivation (subjective, semi-projective, objective). Furthermore, cut-off scores which help to detect faking will be tested.

In order to compare different assessment techniques the Achievement Motivation Inventory (Leistungsmotivations Inventar, LMI) by Schuler and Prochaska (2001), the Multi-Motive-Grid (MMG) by Schmalt, Sokolowski, and Langens (2003), and the Objective Achievement Motivation Test (Objektiver Leistungsmotivationstest, OLMT) by Schmidt-Atzert (2004) were applied in three groups: fake good (FG), fake bad (FB) and control (CG). All instruments were conducted using computer-administered testing.

Hypotheses. We assumed that only the OLMT as an objective measure should make it impossible for the subjects to fake good because there are no obvious strategies to present oneself in a positive way. The only exception is the *aspiration level*: subjects in the faking good condition can use different strategies to make a good impression. As concerns the faking bad condition, differences to the control group should be large as was mentioned above for performance tests.

According to research results described above, scores for semi-projective and self-report measures should differ significantly between the control group and both faking groups (bad/good). The authors of the MMG mentioned that the MMG can hardly be faked (Schmalt, Sokolowski & Langens, 2003, p. 26). However their conclusion was based on a low correlation with the Crowne-Marlowe-Scale and other self-reports. We assume that these results do not prove that the MMG is not susceptible to faking since instruments like the Crowne-Marlowe-Scale can be faked as well (Pauls & Crost, 2004).

Besides this, we tried to find out whether the three methods differ in the way people rate the success of their own faking. Finally, cut-off scores to detect faking were tested for the three tests. In order to do that sensitivity and specificity were determined and tested in a second sample.

Summarizing the aims of this study were to provide information about the fakability of three instruments for measuring achievement motivation as well as the subjective rating of faking ease, and to offer some preliminary cut-off scores for the identification of potential fakers.

Method

Materials

Objective Achievement Motivation Test (OLMT). The OLMT is computer-based and has the character of a performance test. It assesses task related motivation (baseline), motivation through personal goals, aspiration level and motivation through competition. Altogether three subtests have to be carried out. The basic task within each subtest is pretty much the same. Participants are shown a path on the screen which is divided into small numbered squares. The numbers start at one and increase with path length. The path is going in left and right bends up the screen. As long as the bend is turning right the path is coloured red. The colour switches to green when the bend turns to the left. By pressing a red coloured key on the keyboard participants can move up the path when the path is red. For the left direction there is a green key. The task itself always asks to cover as big a distance as possible on the path. In each subtest participants have 10 trials each lasting for 10 seconds. In the second subtest, the subject is also asked how many squares he or she will cover in the following trial. The mean of these goals minus the expected goals equals the score for aspiration level. Expected goals are determined with a regression analysis using actual performance as predictor for goals set. The last subtest confronts the subject with a competitor programmed to be 10% faster than the subject had been in his or her last 3 trials. In the second and third subtest, points are given for reaching or outperforming the self-set goal and respectively for beating the competitor. For failing one's self-set goals, points are subtracted. This way, the scores for motivation through personal goals and motivation through competition are computed. Reliabilities for all subtests are high ($\alpha_{baseline} = .96$; $\alpha_{motivation\ through\ personal\ goals} = .90$; $\alpha_{aspiration\ level} = .83$; $\alpha_{motivation\ through\ competition} = .90$). Regarding the construct validity, one could argue that the OLMT only measures motor ability and not achievement motivation. At first glance, results from a validation study are in line with this argument. They show, that the baseline measure (no other OLMT measure though) correlates moderately with an easy tapping task. However, only the baseline score and not the tapping performance is able to predict a criterion like grade point average (Schmidt-Atzert, 2004, p. 26). Furthermore, the baseline correlates only $r = .35$ with the power version of the SPM (Styles & Raven, 1998). These results can be seen as evidence for divergent validity against motor and cognitive ability. Moreover, there is evidence for the criterion validity of aspiration level. A significant correlation with grade point average ($r = .33$) was found in the same study.

Achievement Motivation Inventory (LMI-K). The short form of the self-report questionnaire Achievement Motivation Inventory (LMI-K) by Schuler and Prochaska (2001) consists of 30 items. Each one has to be rated on a 7 point scale ranging from 1 = „not at all“ to 7 = „completely true“. As a sum, the answers represent an overall achievement motivation score. Typical statements are: “My ambition is easy to challenge.”, “Difficult tasks stimulate me more than easy tasks” (translation by the authors). The LMI is an established achievement motivation test in Germany with an excellent reliability (α = .94) and was therefore selected.

Multi-Motive-Grid (MMG). The Multi-Motive-Grid (MMG) by Schmalt, Sokolowski, and Langens (2003) is a semi-projective test. Eighteen ambiguous pictures are presented consecutively (e.g., two people standing on a place resembling a parking lot wearing construction helmets). Along with each picture, several statements are given (e.g., “Feeling pride because one is able to do something” below the picture just described [translated by the authors]). The subject has to rate whether the statement fits the picture or not. The statements belong to three different motivation types (achievement motivation, power motivation, and affiliation motivation). For the present study, only *hope for success* and *fear of failure* as measures of achievement motivation were of importance. Their reliability was $\alpha_{hope\ for\ success}$ = .69 and $\alpha_{fear\ of\ failure}$ = .67.

Study 1

Sample. n = 123 subjects, aged 18 to 38 years (M = 21.8; SD = 3.1), participated for course credit. The sample consisted of students from a German university with different majors (69% psychology students). The percentage of female subjects was 71%. To ensure that any differences found are not due to differences in the demographic variables, the groups were compared. Age was compared with an ANOVA. Differences in gender and major were tested by Chi² tests. Since the Nullhypothesis is more important in this case (no differences), the beta error is also more important and should be smaller or equal to .05. With the help of G*Power, a compromise power analysis was conducted and new critical p-values were calculated (see Statistical analyses). The three groups (fake good, fake bad, and control) did not differ significantly in their majors. However, there were small but significant differences for gender (w = .21) and age (d < .45). Therefore, all further group comparisons were conducted with age and gender as covariates. Four persons were excluded from the analysis, since they were either not German native speakers or they did not understand the instructions given by the computer. Thus, n = 37 people remained in the fake good group, and n = 41 in each of the other two groups.

Design. Participants were randomly assigned to one of the three groups. All subjects received an extra instruction prior to the first test instruction. Within the control group this extra instruction simply included phrases which typically appear in achievement motivation test instructions such as “do your best”. However, within the faking groups the extra instruction was that the subjects were asked to imagine that they were currently unemployed and applying for a job. Now, they are invited to participate in a selection test. In the fake good group, the job was highly desirable. However, in the fake bad group the job was highly undesirable. Furthermore, subjects were told they would lose financial support from unemployment insurance if they rejected the job. As a consequence, the subjects were asked

to present themselves in a way that would assure that they get the job (fake good) respectively will not get the job (fake bad). To avoid unrealistic scores, subjects were given a warning that a test expert would look at the results to find fakers. The control group also received the same warning in addition to the usual instructions described above.

The test battery consisted of the demographic questions, the three achievement motivation instruments, and follow-up questions. The latter were used to rate the success of one's own faking and to check if the instructions were carried out. To circumvent sequencing effects, the achievement tests were permuted randomly.

Manipulation check and faking performance. After finishing the achievement motivation tests, the subjects were asked whether they had constantly acted in accordance to their extra-instruction. For the faking groups this means, they were asked whether they had constantly paid attention to their faking instruction. All subjects included in the following analysis answered this question with yes. Additionally, the subjects had to rate how successful they thought they were in their faking performance (7 point scale from 1 = „not at all" to 7 = „very good"). This information was only obtained in study 1.

Cut-off test study

Sample. One of the goals of this study is to find cut-off values which indicate faking. Since such values are highly sample dependent they were tested in a second sample which consisted of n = 123 subjects. All of them were first semester psychology students at the university of Marburg aged 18 to 46 years (M = 21, SD = 3.2). 81% of the sample were women.

Design. The data were obtained from a study that all first semester psychology students at the University of Marburg took part in. No faking instruction was given and the test results had no further implications for students. Therefore, the results should not be distorted.

Statistical analyses

All computations were conducted with SPSS 14.0 for Windows or the program G*Power by Erdfelder, Faul, and Buchner (1996). First of all, a multivariate analysis of variance with all the achievement motivation test scores as dependent variables was conducted. Since these variables all have different metrics, the raw scores were transformed into T-values according to the test norms. Following MANOVA, post hoc group comparisons were conducted.

The next step was to calculate effect sizes and power for all comparisons between control and faking groups. Because this was done separately for each of the test variables no artefacts due to different metrics were possible and therefore, raw scores were used. We chose Glass' d as effect size measure (Rustenbach, 2003). This measure uses the control group standard deviation as divisor. This standard deviation represents the results of honest test answering, which is why we preferred it to other standard deviations. As mentioned before, we assumed that there is no difference between the fake good and the control group for the OLMT scores. When trying to prove that a difference does not exist, the beta risk is of importance and should be held rather low. With the help of the G*Power program we conducted a compromise power analysis (Erdfelder, Faul & Buchner, 1996). Within this

analysis the actual effect size found as well as the group sizes, and a specified ratio of alpha to beta are entered into G*Power. Beta was set to be 5 % and alpha 20%. This follows the idea that beta is the more important error. A new critical value for alpha as well as the power for such a test is then computed. We also conducted a Bonferroni adjustment for the number of significance tests. In this kind of analyses the beta error has to be divided by the number of tests. We corrected for 14 significant tests conducted. Thus, the ratio of beta to alpha entered into the software was (.05/14)/.20.

Results

Table 1 displays means, standard deviations, and the corresponding T-values of the norm group for the achievement test scores for each group and for the cut-off test sample as well as the results of the group comparisons.

Since the groups differ significantly in age and gender we performed a MANCOVA with the two variables as covariates. The MANCOVA with the T-values of the achievement motivation test scores as dependent variable and the 3 groups as factor yielded a Wilk's Lambda of .28 ($F(14; 216) = 13.70$, power = 1.00, $p \leq .001$, $\eta^2 = .47$). Thus, there was a significant difference between the three groups concerning their results in the achievement motivation tests. The Box-M-Test yielded a significant result $F(56; 37739) = 1.83$, $p \leq .001$. Thus, actual power is decreased (Weinfurt, 1995, p. 245). However, the power found was 1.00, so even a decrease should leave sufficient power. The results of the group comparisons are also shown in table 1.

It is obvious that all three tests were susceptible to faking bad. With one exception, all comparisons between the fake bad and the control group were significant with moderate to large effect sizes. The exception was *fear of failure* (MMG). However, the effect size here was also moderate. Concerning faking good, the results show that the OLMT scores did not differ significantly between the fake good and control group. The LMI-K, on the other hand, was susceptible to faking good. Even though the effect size was smaller than for the comparison of the fake bad and the control group, it still amounted to one standard deviation. For the MMG, results were mixed. *Hope for success* showed no significant difference between the fake good and the control group, but *fear of failure* did. The insignificant comparison still reached a small effect size. The power for most tests was above .80 and thus, satisfying. The only exceptions were the comparisons with the MMG scores. This also explains the insignificance for moderate effect sizes. This means, the probability of finding significant effects of such magnitude was too small given the sample size. In other words, the sample size was probably too small for the found effect. It has to be noted, that Levene Tests for all OLMT scores (except *baseline*) as well as for the MMG score *fear of failure* and the *LMI-K* score were significant ($p \leq .04$). However, since group sizes are about equal and larger than $n = 10$, the post hoc tests are robust (Howell, 2002, p. 213).

Table 2 provides information on the convergent validity for the different tests. Overall the correlations are mostly small and insignificant. A few significant correlations can be found which were small to moderate and in the expected direction.

Table 1:
Means, standard deviations, T-values for the test scores and group comparisons

		Fake Good			Fake Bad			Control Group			Cut - off test				FG vs. CG	FB vs. CG
		M	*SD*	*T*	*M*	*SD*	*T*	*M*	*SD*	*T*	*M*	*SD*	*T*			
OLMT[1]	b	64.1	7.4	55	38.8	15.6	20	62.7	5.8	53	61.6	6.4	52	*d*	.12[a]	-2.23***
														1-²	.98	1.00
	mgoal	0.7	3.0	53	-0.9	9.2	46	0.4	2.5	52	0.3	2.8	52	*d*	.10[b]	-.86*
														1-²	.99	.84
	al	2.4	1.7	60	-2.2	5.9	40	3.4	1.7	66	2.6	1.7	61	*d*	-.46[c]	-1.93***
														1-²	.99	1.00
	mcom	-1.4	4.2	47	-7.9	9.0	20	-0.5	2.5	49	-1.8	5.8	47	*d*	-.34[d]	-2.06***
														1-²	.99	1.00
LMI - K[2]		172.5	20.2	62	109.5	23.3	36	149.5	22.1	53	146.9	22.1	52	*d*	1.06***	-1.54***
														1-²	.98	1.00
MMG[3]	hs	7.8	2.7	51	5.0	2.6	40	6.8	2.6	47	7.0	2.2	47	*d*	.50[5]	-.78*
														1-²	.75	.92
	ff	2.9	2.6	47	6.1	2.6	60	4.4	2.5	54	5.1	2.5	57	*d*	-.75*	.50
														1-	.90	.76

Notes. b = Baseline, mgoal = motivation through personal goals, al= aspiration level, mcom = motivation through competition, LMI-K = achievement motivation index (minimum 30, maximum 210 points), hs = hope for success (minimum 0, maximum 12 points), ff = fear of failure (minimum 0, maximum 12 points). The table shows means and standard deviations. Norm groups for T-values:[1] = OLMT norm group with $n = 170$ (83 females); [2] = LMI-K norm group with $n = 3660$ (1635 females); [3] = MMG norm group with $n = 361$ (115 females). *d* = Glass' d (division by the control group standard deviation); *1-β* = Power; Bonferroni adjusted significance levels are * = $p \leq .05$; ** = $p \leq .01$; *** = $p \leq .001$. Exceptions are the comparisons FG vs. CG in the OLMT: [a] significant for $p \leq .67$; [b] significant for $p \leq .69$; [c] significant for $p \leq .29$; [d] significant for $p \leq .41$. One tailed tests.

Table 2:
Convergent Validities

group		Mgoal	al	mcom	lmi	hs	ff
Fake Good	b	.01	.12	.05	.12	-.02	.09
	mgoal		.25	.37*	-.15	-.22	.04
	al			.30	.18	.30	-.02
	mcom				.15	-.23	-.10
	lmi					.39*	-.57**
	he						-.18
Fake Bad	b	-.24	.34*	.19	.10	.26	-.10
	mgoal		-.17	.61**	-.04	.02	-.11
	al			.33*	-.17	-.06	-.25
	mcom				-.13	.18	-.21
	lmi					.13	-.14
	he						-.21
Control Group	b	.06	.31*	.09	-.13	-.05	.22
	mgoal		.36*	.64**	-.01	.12	.12
	al			.28	.01	.13	.32*
	mcom				-.10	.04	.12
	lmi					.51**	-.06
	he						.03
Cut - off test sample	b	-.05	.10	.15	-.13	.01	-.05
	mgoal		.52**	.55**	.13	-.03	.11
	al			.20*	-.03	-.07	.09
	mcom				.00	-.04	.18*
	lmi					.22*	-.19*
	he						-.15

Notes. b = Baseline, mgoal = motivation through personal goals, al = aspiration level, mcom = motivation through competition, LMI-K = achievement motivation index, hs = hope for success, ff = fear of failure. Significance levels are * = $p \leq .05$; ** = $p \leq .01$; *** = $p \leq .001$.

The subjective ratings indicate how good the two faking groups rated their own faking performance. These scores were compared for the two faking groups on each test. Moreover, comparisons for the tests within each faking group were conducted. Using ANOVAs, the two groups' ratings of faking success ($M_{Fake\ good}$ = 4.8 and $M_{Fake\ bad}$ = 4.6) did not differ significantly for the LMI (p = .63). For the OLMT ($M_{Fake\ good}$ = 3.2 and $M_{Fake\ bad}$ = 4.8), the fake good group rated their faking performance significantly worse than the fake bad group ($p \leq .001$). The effect size using the pooled standard deviation was d = -.91. Results for the MMG ($M_{Fake\ good}$ = 4.5 and $M_{Fake\ bad}$ = 4.0) were opposite ($p \leq .05$), and the effect size was d = .44). The p-values were Bonferroni-adjusted. It has to be noted that the means of all ratings were at least average or above the scale mean. The within group comparisons of the test were conducted using t-tests for dependent measures. The p-values were Bonferroni-adjusted.

Regarding their own faking performance, subjects in the fake good group rated their performance on the OLMT worse than on the LMI-K ($p \leq .001$; $d = -.65$) and MMG ($p \leq .001$; $d = -.67$). No significant difference occurred for the comparison of MMG and LMI-K. In the fake bad group, the success in faking the MMG was rated significantly worse compared to the OLMT ($p \leq .002$; $d = -.51$) and the LMI-K ($p \leq .001$; $d = -.52$). No significant difference occurred for the comparison of OLMT and LMI-K ($p > .008$).

Sensitivity and specificity for different cut-off scores were calculated. Those cut-off scores, which showed the highest specificity while providing a reasonable sensitivity, were selected. We paid more attention to the specificity, since we believe it is worse to classify someone as a faker when in fact he or she is not. Table 3 shows the cut-off-scores along with their sensitivity and specificity for both kinds of faking.

Detecting faking bad with the OLMT *baseline* score as indicator, only 7% are unjustly classified as fakers, while none of the fakers remains unnoticed. Faking bad in the LMI-K can be detected with almost equally good sensitivity and specificity. Unfortunately, the cut-off scores for the MMG did not reach an acceptable level of sensitivity. Only the specificity

Table 3:
Cut-off scores for detecting faking

	Faker		**Non Faker**			
Cut-off	+	-	+	-	Sensitivity	Specificity
Faking Bad						
OLMT						
Baseline ≤ 55	41	0	38	3	1.00	0.93 (0.85)
aspiration level ≤ 1	30	11	39	2	0.73	0.95 (0.86)
motivation through competition ≤ -3	32	9	37	4	0.78	0.90 (0.68)
LMI-K						
LMI-Index ≤ 119	29	12	38	3	0.71	0.93 (0.89)
MMG						
hs ≤ 4	15	26	32	9	0.37	0.78 (0.86)
ff ≥ 7	12	29	34	7	0.26	0.83 (0.72)
Faking Good						
LMI-K						
LMI-Index ≥ 165	27	10	33	8	0.73	0.81 (0.81)
MMG						
hs ≥ 9	16	21	29	12	0.43	0.71 (0.72)
ff ≤ 2	21	16	33	8	0.57	0.81 (0.84)
*Classification type**	TP	FN	TN	FP	TP/(TP+FN)	TN/(TN+FP)

Notes. Faker is a subject of the fake bad or fake good group and non fakers are from the control group. „+“ means that the subject was correctly classified and „-“ wrongly classified. * TP = true positive, FP = false positive, TN = true negative, FN = false negative. The numbers in brackets are the results of the cut-off test. Cut-off scores to detect faking bad for *motivation through personal goals* and faking good for any OLMT scores were not computed since no significant difference occurred. hs = hope for success, ff = fear of failure.

was acceptable. The same holds true for the detection of faking good. Sensitivity and specificity of detecting faking good is satisfying for the LMI-K, while only specificity is acceptable for the MMG.

In a second step, the cut-off scores were tested. For this, a different sample was used. In this sample, faking was not very likely. Therefore, the number of subjects classified as fakers should be very low. An analysis with the cut-off scores for faking bad and faking good, respectively, was conducted. Since no faking instruction was given, sensitivity cannot be determined. The results of the first study could extensively be replicated (see table 3, specifity scores in brackets): specificity values remain high in the test sample for all cut-off scores. Only for the OLMT score *motivation through competition,* the specificity of the cut-off score decreased when looking for faking bad.

Discussion

The main goals of this study were to investigate the faking susceptibility of different noncognitive measures and the subjective ease of faking as well as cut-off scores for the detection of faking. The study revealed that the methods differed in their susceptibility to faking. The objective test did not allow faking good and provided useful cut-off scores to detect faking bad. Both other measures (subjective and semi-projective) were susceptible to faking but to a different extent. The self-report allowed faking to a higher degree. Cut-offs for the subjective and semi-projective measures provided an acceptable specificity. However, only the sensitivity for the self-report measure was sufficient. The good specificity values could be replicated. Furthermore, it could be shown that the faking groups differed in the way they rated their faking success. The objective measure was perceived to be easier to fake under fake bad conditions, and the semi-projective measure under fake good conditions. Within-group comparisons revealed that the fake good group found the objective measure hardest to fake while it was the semi-projective measure in the fake bad group.

Faking the Objective Test. The most obvious reason for these results is that people were able to recognize what was being measured. Thus, they intentionally distorted their responses. However, if the test contains an upper limit for performance, like the OLMT, faking upwards is not possible. Interestingly, the subjects in this study admitted that faking good in the OLMT was harder than in both other tests. Yet, the mean rating of faking success was still average. In other words, people were aware that somehow the objective measure made faking harder, yet, they still believed they did a good job. The reason for this probably is the unknown computation of test scores. The *baseline* only takes the last 3 trials of the first subtest into account. And for *motivation through personal goals* and *motivation through competition,* the *baseline* is partialled out of the performance score. It is reasonable to assume that the unawareness of actual score computation led people into believing they did a good job. After all, they saw little increases in their performances compared with the first trials. The unawareness of actual score computation also can be hold responsible for a surprising result: The moderate effect size for the comparison between the fake good and the control group in the *aspiration level* score. The *aspiration level* score is the only OLMT score, which is not directly dependent on motor ability, subjects can set their own goals. The mean of these goals minus the goals expected due to a regression on the actual performance defines the *aspiration level* score. Thus, if one sets higher goals the score will be larger. If

this was obvious to the participants, it would be a chance for faking. But interestingly, the mean *aspiration level* score in the fake good group was smaller than in the control group. This means, the subjects in the fake good group tried to make a good impression by setting more realistic goals.

We hypothesized that faking bad within the OLMT was easy to detect. Subjects do not have a frame of reference and badly misjudge a poor performance. One has to keep in mind that the subjects had been told to fake reasonably, since a test expert would analyse their results. Despite this, the differences were really large and faking bad in the OLMT could be detected easily.

Faking the Self-Report. Concerning the self - report measure (LMI-K), our results are in line with previous findings (Martin, Bowen, & Hunt, 2002; Viswesvaran & Ones, 1999). Obviously, the test intention and score computation are easy to guess and thus, the test can be faked easily. However, looking at the effect sizes and the cut - off scores, it is also obvious that faking can be detected. This occurred although subjects were told to fake reasonably. Thus, even with such an easy answer format, subjects seemed to lack a frame of reference for their faking and did not fake "smart".

Faking the Semi-Projective Test. The results for the semi - projective measure (MMG) were mixed. The significance levels and effect sizes show that faking was possible in spite of the results reported by the test authors. Unexpectedly, *hope for success* in the faking good group and *fear of failure* scores in the faking bad group did not differ significantly from the control group, but effect sizes of these differences were moderate and power was below .80. It can be assumed that the sample size was too small. As discussed in many papers (for a good overview see Nickerson, 2000), one should not only look at the significance level but also at the effect size when judging a hypothesis. Doing this, we strongly believe that the differences found are substantial. However, the fake bad group rated their faking performance regarding the MMG worst but still in an average area. Thus, the test's intention seems to be harder to guess. Nevertheless, faking was possible in both groups. Unfortunately, the cut-off scores did not have sufficient sensitivity. Thus, faking on the MMG was possible, but the shift in group means was smaller than observed with the self-report measure. This is in part due to the fact that the MMG raw scores are not as differentiated as the LMI scores. The range is only 12. It is reasonable to assume that a longer MMG version would bear larger faking effects and make the detection of faking easier.

Limitations and outlook. One could argue that the most critical limitation of this study is the sample of university students. But, as Rogers (1997) pointed out, the sample in simulation studies should have practical relevance. Since students are a very likely clientele in personnel selection scenarios, we believe our sample has this practical relevance. Nevertheless, as already mentioned above, effect sizes tend to be larger in student studies compared to real applicant settings. Thus, the cut-offs reported here cannot be transferred to such settings.

An interesting question that remains unanswered is, whether some people are better fakers than others. And moreover, if this is the case, what enables them? Are they more intelligent or more familiar with test situations? These questions are certainly worth exploring.

Another limiting aspect is the use of achievement motivation measures which are not used as often as other noncognitive measures such as Big 5 measures. However, as

mentioned above, achievement motivation can be assessed with well known measures using all the different approaches. Moreover, the present study did not aim at exploring criterion validity but faking susceptibility. Thus, the actual construct measured was of secondary interest.

Finally, one could argue that looking at the convergent validities for the various measures that they do not assess the same construct. However, the low convergent validities are in line with other research results. Spangler (1992) reported only low but significant correlations in his meta-analyses of achievement motivation measures. Thrash and Elliot (2002) postulated that different achievement motivation measures despite the low intercorrelations assess the same construct. Thus, the low convergent validities found here are not surprising and do not give evidence against the construct validity of any of the measures.

Practical implications. As we pointed out in the introduction, there is no real good way to ban faking from personality testing. Nevertheless, faking seems to be a threat to construct validity. Whenever it is necessary to really measure the intended construct, there needs to be at least one reliable way to detect fakers. Thus, a certain safety in classifying people as fakers is necessary. The practical implications of these results presented here are: one should use either objective or self-report measures, when the detection of faking is necessary. When it is needed to prevent faking good, objective measures should be applied. Moreover, test authors should try to include cut-off scores for the detection of faking in their manuals. However, our results also show that using these scores does not assure a perfect hit rate. Yet, the objective test did very good and the numbers of false positives respectively false negatives were comparably low.

The results regarding the subjective ease of faking can help in test selection. Thus, in a high stakes situation where faking is highly likely, objective measures can help preventing faking good while semi-projective give participants a harder time to fake bad.

Cut-off scores reported here should not be generalized for any high stakes situation. It is reasonable to assume that participants in a laboratory fake more extreme than people in a real world setting. Thus, the cut-off scores calculated here might be to extreme. Moreover, it has been shown that people fake according to the specific goal they have set for themselves. Thus, faking might differ in an application setting for a nurse job compared with a manager position (Pauls & Crost, 2005). Therefore, if test authors include cut-off scores into their test manuals they should also provide detailed information on the specific situation in which they were obtained. However, the general usefulness of such cut-offs has been proven and we hope that test authors and practitioners will consider them in the future.

Of course, one important question, which has to be answered by each practitioner, remains unanswered. This study cannot determine which measure is the better one for the prediction of real-world criteria. Nevertheless, a good predictor of real-world criteria remains useless, if it can be faked easily and if faking cannot be detected. In Spangler's meta analyses (1992), self-report measures yielded only low criterion validities that were in general also lower than those of projective measures. This fact would speak against self-report measures. However, up to now, there are only few results regarding the practical importance of objective measures. So far the OLMT was shown to be able to predict school grades. Further validity studies should try to strengthen this basis. Moreover, questionnaires have proven their usefulness when other constructs such as the Big 5 are assessed (e.g.

Chamorro-Premuzic, 2006; Furnham & Crump, 2005; Lounsbury, Sundstrom, Loveland, & Gibson, 2002; Marcus, Hoft, & Riediger, 2006).

The practical importance of detecting faking has been laid out. This along with the arguments presented by Hough should encourage test authors to put more effort into determining cut-off scores. Furthermore, it should persuade practitioners to put more confidence in personality and motivation tests. Even though the prognostic validity of such tests seems rather low, as Schmidt and Hunter (p. 273, 1998) say: ”In economic terms, the gains from increasing the validity of hiring methods can amount over time to literally millions of dollars.” Thus, even small increases in overall validity might be worth a lot of money.

References

Baron, H. (1996). Strengths and limitations of ipsative measurement. *Journal of Occupational and Organizational Psychology*, 69, 49-56.

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.

Bowen, C.C., Martin, B.A., & 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.

Bradley, K. M., & Hauenstein, N. M. A. (2006). The moderating effects of sample type as evidence of the effects of faking on personality scale correlations and factor structure. *Psychology Science. Vol, 48*(3), 313-335.

Caldwell-Andrews, A., Baer, R.A., & Berry, D.T.R. (2000). Effects of response sets on NEO-PI-R scores and their relations to external criteria. *Journal of Personality Assessment*, 74 (3), 472-488.

Cassidy, T., & Lynn, R. (1989). A multifactorial approach to achievement motivation: The development of a comprehensive measure. *Journal of Occupational Psychology, 62*(4), 301-312.

Chamorro-Premuzic, T. (2006). Creativity versus conscientiousness: Which is a better predictor of student performance? *Applied Cognitive Psychology, 20*(4), 521-531.

Christiansen, N. D., Burns, G. N., & Montgomery, G. E. (2005). Reconsidering forced-choice item formats for applicant personality assessment. *Human Performance, 18*, 267-307.

Entwisle, D. R. (1972). To Dispel Fantasies About Fantasy-Based Measures of Achievement Motivation. *Psychological Bulletin, 77*(6), 377-391.

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods Instruments & Computers, 28*, 1-11.

Furnham, A. (1997). Knowing and faking one's Five-Factor personality score. *Journal of Personality Assessment, 69*(1), 229-243.

Furnham, A., & Crump, J. (2005). Personality traits, types, and disorders: An examination of the relationship between three self-report measures. *European Journal of Personality, 19*(3), 167-184.

Griffith, R. L., Chmielowski, T., & Yoshita, Y. (2007). Do applicants fake? An examination of the frequency of applicant faking behavior. *Personnel Review, 36*(3), 341-357.

Hogan, R., Hogan, J., & Roberts, B. W. (1996). Personality measurement and employment decisions: Questions and answers. *American Psychologist, 51*(5), 469-477.

Jackson, D. N., Wroblewski, V. R., & Ashton, M. C. (2000). The impact of faking on employment tests: Does forced choice offer a solution? *Human Performance, 13*, 371-388.

Khorramdel, L., & Kubinger, K. D. (2006). The effect of speediness on personality questionnaires: an experiment on applicants within a job recruiting procedure. *Psychology Science*(48(3)), 378-397.

Kleinmann, M. (1993). Are rating dimensions in assessment centers transparent for participants? Consequences for criterion and construct validity. *Journal of Applied Psychology, 78*(6), 988-993.

Lounsbury, J. W., Sundstrom, E., Loveland, J. L., & Gibson, L. W. (2002). Broad versus narrow personality traits in predicting academic performance of adolescents. *Learning and Individual Differences, 14*(1), 65-75.

Marcus, B. (2006). Relationships between faking, validity, and decision criteria in personnel selection. *Psychology Science, 48*(3), 226-246.

Marcus, B., Hoft, S., & Riediger, M. (2006). Integrity tests and the five-factor model of personality: A review and empirical test of two alternative positions. *International Journal of Selection and Assessment, 14*(2), 113-130.

Martin, B. A., Bowen, C. C., & Hunt, S. T. (2002). How effective are people at faking on personality questionnaires? *Personality and Individual Differences, 32*(2), 247-256.

McClelland, D. C., Koestner, R., & Weinberger, J. (1989). How do self-attributed and implicit motives differ? *Psychological Review, 96*(4), 690-702.

Moormann, R. H., & Podsakoff, P. M. (1992). A meta-analytic review and empirical test of the potential confounding effects of social desirability response sets in organizational behaviour research. *Journal of Occupational and Organizational Psychology, 65*, 131-149.

Mueller-Hanson, R., Heggestad, E. D., & Thornton, G. C. (2003). Faking and selection: Considering the use of personality from select-in and select-out perspectives. *Journal of Applied Psychology, 88*(2), 348-355.

Mueller-Hanson, R., Heggestad, E. D., & Thornton, G. C. (2006). Individual differences in impression management: an exploration of the psychological processes underlying faking. *Psychology Science, 3*, 288-312.

Murphy, K. R., & Davidshofer, C. O. (2001). *Psychological Testing.* New Jersey: Prentice Hall.

Murray, H. A. (1938). *Explorations in personality*. New York: Oxford University Press.

Ones, D. S., Viswesvaran, C., Dilchert, S., & Deller, J. (Eds.). (2006). *Considering response distortion in personality measurement for industrial, work and organizational psychology research and practice* (Vol. 48(3)): Psychology Science (Special Issue).

Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. I. Braun, D. N. Jackson & D. E. Wiley, *The role of constructs in psychological and educational measurement* (pp. 46-49). Mahwah, NJ, US: Lawrence Erlbaum Associates.

Pauls, C. A., & Crost, N. W. (2004). Effects of faking on self-deception and impression management scales. *Personality and Individual Differences, 37*(6), 1137-1151.

Paus, C. A. & Crost, N. W. (2005). Effects of different instructional sets on the construct validity of the NEO-PI-R. *Personality and Individual Differences, 39*, 297-308.

Ramsay, L. J., Schmitt, N., Oswald, F. L., Kim, B. H., & Gillespie, M. A. (2006). The impact of situational context variables on responses to biodata and situational judgment inventory items. *Psychology Science. Vol, 48*(3), 268-287.

Rogers, R. (1997). Researching Dissimulation. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (2 ed., pp. 309-327): New York: Guilford.

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*(4), 634-644.

Rustenbach, S. J. (2003). *Metaanalyse*. [Metaanalyses]. Huber.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*(2), 262-274.

Schmidt-Atzert, L. (2004). OLMT. *Objective Achievement Motivation Test*. Moedling: Schuhfried.

Schmit, M. J., & Ryan, A. M. (1993). The Big Five in personnel selection: Factor structure in applicant and nonapplicant populations. *Journal of Applied Psychology, 78*(6), 966-974.

Spangler, W. D. (1992). Validity of questionnaire and TAT measures of need for achievement: Two meta-analyses. *Psychological Bulletin, 112*(1), 140-154.

Spinath, F. M. (2001). Genetic and environmental influences on achievement motivation and its covariation with personality and intelligence. In R. Riemann, F. M. Spinath & F. Ostendorf (Eds.), *Personality and Temperament: Genetics, Evolution, and Structure* (1 ed., pp. 11-25): Pabst Science Publishers.

Styles, I., & Raven, M. (1998). *Standard Progressive Matrices. Plus Version.*: Dumfries: J.C. Raven.

Terman, L. M., & Oden, M. H. (1959). *Genetic studies of genius. Vol. V. The gifted group at mid-life*. Stanford, CA: Stanford University Press.

Thrash, T. M., & Elliot, A. J. (2002). Implicit and self-attributed achievement motives: Concordance and predictive validity. *Journal of Personality, 70*(5), 729-755.

Topping, G. D., & O'Gorman, J. G. (1997). Effects of faking set on validity of the NEO-FFI. *Personality and Individual Differences, 23*(1), 117-124.

Vickery, C. D., Berry, D. T. R., Dearth, C. S., Vagnini, V. L., Baser, R. E., Cragar, D. E., et al. (2004). Head injury and the ability to feign neuropsychological deficits. *Archives of Clinical Neuropsychology, 19*(1), 37-48.

Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement, 59*(2), 197-210.

Weinfurt, K. P. (1995). In L. G. In Grimm & P. R. Yarnold (Eds.), *Reading and understanding multivariate statistics* (pp. 245-276). Washington, DC: APA.