The utility of overclaiming questionnaires depends on the fit between test content and application context

Adrian Hoffmann[1] · Birk Diedenhofen[1] · Sascha Müller[2,3]

Abstract

Overclaiming questionnaires (OCQs) were proposed as a means to counteract social desirability bias by capturing individual differences in participants' self-enhancement tendencies in self-report assessments. Previous studies that evaluated OCQs reported mixed results. However, fit between the content of an OCQ in terms of its items and the context in which the measure is presented has not been tested systematically. In a mock application study ($N = 432$), we compared different levels of content-context fit between conditions. Results show that the utility of a general knowledge OCQ varied as a function of its content fit to different application contexts. Expectedly, overclaiming was most pronounced in an application context with optimal content fit to the OCQ, followed by a context with lower fit and an honest control condition without application context. Furthermore, participants in the application conditions were shown to successfully fake on conventional personality scales while incorporating specific requirements of the application context into their faking behavior. Our results thus corroborate previous findings suggesting a high susceptibility of personality scales to deliberate faking. In contrast, when content-context fit is taken into account, OCQs may be a promising method for assessing applicant faking.

Keywords Overclaiming · Faking · Personnel selection · Social desirability · Validity

Introduction

Social desirability bias threatens the validity of psychological measurements as some individuals tend to respond in a way that will leave a positive impression rather than truthfully. This influence is particularly harmful in situations in which individual evaluations are based on self-reports, as, for example, in personnel selection (Krumpal, 2013; Phillips & Clancy, 1972; Tett & Simonet, 2021; Tourangeau & Yan, 2007). Such scenarios are often referred to as high-demand situations in which individuals are thought to have a high incentive to present themselves favorably, for instance, to increase their chances of being offered a job because they fit the perceived profile the employer is looking for. Empirical evidence suggests that participants in job application situations can and will successfully fake on personality tests provided they have the necessary skills and perceive potential benefits to outweigh associated risks (Huber et al., 2021; Klehe et al., 2012; Rosse et al., 1998; Roulin et al., 2016).

Overclaiming questionnaires (OCQs) have been developed as a potential means to measure and control for individual faking tendencies independent of cognitive ability (Paulhus & Harms, 2004; Paulhus et al., 2003). To this end, participants are typically presented with an alleged general knowledge test consisting of a list of words, and they are instructed to rate their familiarity with each word. Unknown to the participants, the list contains both real ("reals") and fictitious words ("foils"). From the participants' answers, *OCQ bias* can be calculated as a measure of participants' tendency to overstate their general knowledge by claiming familiarity with nonexistent words. Thus, individuals that overclaim are thought to misrepresent themselves with respect to a socially desirable characteristic.

Studies that aimed to locate OCQ bias in a nomological network of social desirability facets or personality

Adrian Hoffmann and Birk Diedenhofen contributed equally to this work

✉ Adrian Hoffmann
adrian.hoffmann@uni-duesseldorf.de

1 Department of Experimental Psychology, University of Duesseldorf, Universitaetsstrasse 1, 40225 Duesseldorf, Germany

2 Department of Psychology, University of Kassel, Kassel, Germany

3 Department of Psychosomatic Medicine and Psychotherapy, Rostock University Medical Center, University of Rostock, Rostock, Germany

traits provided rather inconclusive findings. Some studies reported positive associations with self-enhancement measures such as impression management, narcissism, or risk-taking (Bensch et al., 2019; Paulhus et al., 2003; Randall & Fernandes, 1991; Ziegler et al., 2013); others reported little to no overlap with self-enhancement tendencies, dark personality traits, or risk preferences (Goecke et al., 2020; Grosz et al., 2017; Keller et al., 2021; Müller & Moshagen, 2019a, b).

Overclaiming questionnaires as a measure of applicant faking

Despite these theoretical challenges, overclaiming questionnaires have been applied in the context of personnel selection due to their potential to capture applicant faking behavior. Prominently, Bing et al. (2011) could show in a sample of undergraduate students that an overclaiming measure successfully assessed individual differences in faking. Using overclaiming as a suppressor even substantially improved the predictive validity of an ability measure and a personality measure for future scholastic success. Feeney and Goffin (2015), however, attested the overclaiming technique limited utility in measuring applicant faking. In a study simulating a job application for a retail sales position, the overclaiming bias was outperformed by the competing measure of residualized individual change scores (RICS) in predicting self-reported faking tendencies. Notably, the RICS scores were based on an extraversion test that matched the requirements of the application scenario well, whereas the general knowledge OCQ arguably showed a lower fit to the job requirements because extraversion may be more relevant for a salesperson as compared to word knowledge.

The content-context fit of overclaiming questionnaires was at the very core of a recent series of studies by Dunlop et al. (2020). Keeping the application context constant while experimentally manipulating whether an OCQ captured job-relevant versus job-irrelevant content, the authors could show that the OCQ bias was positively associated with other measures of faking and positive self-presentation if—and only if—the content-context fit of the OCQ was high. Building on this observation, Diedenhofen et al. (2022) recently revisited the findings of Feeney and Goffin (2015). In a mock application study in which both OCQ and RICS fit the application scenario well, Diedenhofen et al. (2022) found that OCQ bias and RICS scores actually predicted participants' self-reported faking tendencies with comparable accuracy.

The current study

The central objective of the current study was to further validate and extend the recent finding that an optimized content-context fit is a necessary condition for the utility of OCQs in application settings. First, we wanted to investigate the sensitivity of the OCQ bias measure to participants' faking behavior in different application situations with varying contextual fit to the OCQ content. To this end, and expanding on the findings by Dunlop et al. (2020), we kept the OCQ content constant while experimentally manipulating the application context. Participants were instructed to respond to our questionnaire either as if they were applying for a job as a science journalist, as if they were applying for a job as a psychotherapist, or honestly in a control condition without any application context. We expected generally higher bias scores in both application conditions compared to the honest control condition. Because the OCQ in the current study allegedly measured general knowledge, we furthermore expected bias scores under application conditions with a high content-context fit (science journalist) to exceed those under application conditions with a lower content-context fit (psychotherapist). This resulted in three empirically testable hypotheses:

H1a: *OCQ bias* (science journalist) > *OCQ bias* (honest control)
H1b: *OCQ bias* (psychotherapist) > *OCQ bias* (honest control)
H1c: *OCQ bias* (science journalist) > *OCQ bias* (psychotherapist)

Second, we aimed at investigating how the different application contexts would affect participants' responses to conventional personality scales. To this end, we examined a global measure of achievement motivation. Because achievement motivation should potentially be relevant to any job application context, we expected scores for this measure in both application conditions to exceed those in the honest control condition; for the difference in achievement motivation between application conditions with high (science journalist) versus low content-context fit (psychotherapist), we did not expect a specific direction and therefore formulated an undirected hypothesis.

H2a: *achievement motivation* (science journalist) >
　　　achievement motivation (honest control)
H2b: *achievement motivation* (psychotherapist) >
　　　achievement motivation (honest control)
H2c: *achievement motivation* (science journalist) ≠
　　　achievement motivation (psychotherapist)

As a second personality scale, we assessed a composite measure of traits specifically favorable for the profession of psychotherapists ("therapist personality"). We expected scores for this measure in both application conditions to exceed those in the honest control condition. Furthermore, if participants were indeed capable of taking the specific requirements of the application situation into account while faking their answers, we expected therapist personality

scores in the psychotherapist application condition to exceed those in the science journalist application condition.

H2d: *therapist personality* (science journalist) >
    therapist personality (honest control)
H2e: *therapist personality* (psychotherapist) >
    therapist personality (honest control)
H2f: *therapist personality* (psychotherapist) >
    therapist personality (science journalist)

Third, we wanted to assess a potential proxy for participants' faking behavior. This proxy was the number of extreme responses in the desirable direction that participants provided across all items of the personality scales (extreme positivity, cf. Dunlop et al., 2020). As participants in the application conditions would presumably have a stronger motivation to fake, the extreme positivity in these conditions was expected to be higher compared to the honest control condition; for the difference between application conditions, we did not expect a specific direction and therefore formulated an undirected hypothesis.

H3a: *extreme positivity* (science journalist) >
    extreme positivity (honest control)
H3b: *extreme positivity* (psychotherapist) >
    extreme positivity (honest control)
H3c: *extreme positivity* (science journalist) ≠
    extreme positivity (psychotherapist)

Furthermore, we investigated potential associations of the assessed variables. To the extent to which scores in the respective measures were influenced by similar processes driving applicant faking, we expected to find positive associations between OCQ bias and achievement motivation, therapist personality, and extreme positivity.

H4a: ρ (OCQ bias, achievement motivation) > 0
H4b: ρ (OCQ bias, therapist personality) > 0
H4c: ρ (OCQ bias, extreme positivity) > 0

Finally, following an exploratory approach without prior hypotheses, we assessed whether the strength of these associations differed between experimental conditions.

Method

Participants and procedure

Initially, 454 participants were recruited ad-hoc by two student assistants as a university community sample. The only criterion for participation was that participants had to be of legal age, that is, 18 years or older. After data entry, a total of 22 participants (4.9%) were excluded because they had not provided an answer to ($n = 6$, 1.3%) or failed the manipulation check ($n = 3$, 0.7%), not provided an answer to ($n = 5$, 1.1%) or failed the seriousness check ($n = 9$, 2.0%), or omitted responses to more than 20% of all items in the OCQ ($n = 3$, 0.7%), the LMI ($n = 3$, 0.7%), or the BIP ($n = 2$, 0.4%). The final sample comprised $N = 432$ participants (95.2% of the initial sample) recruited on the campuses of the University of Bochum ($n = 168$, 38.9%), the University of Wuppertal ($n = 134$, 31.0%), the University of Applied Sciences Niederrhein ($n = 52$, 12.0%), the University of Duisburg-Essen ($n = 41$, 9.5%), and the University of Applied Sciences Düsseldorf ($n = 37$, 8.6%). Of these participants, 213 (49.3%) identified as female and 219 (50.7%) identified as male; age ranged from 18 to 45 years with a mean value of 24.0 years ($SD = 4.0$); 346 participants (80.1%) reported German as their native language, 86 (19.9%) indicated a different native language. With respect to formal education, only one participant (0.2%) reported that they had not yet earned an educational degree; this claim is realistic given the university community sampling approach, since German university campuses are open to academic staff and students as well as, for example, visitors and people working in jobs that do not require any formal school education (cleaning staff, untrained manual workers, etc.). Seven participants (1.6%) reported an intermediate school leaving certificate as their highest educational qualification (referred to as "mittlere Reife" in the German educational system); 288 participants (66.7%) reported that they had earned a high school diploma or specialized baccalaureate ("Abitur" / "Fachabitur"); 134 participants (31.0%) reported that they had completed a university degree; two participants (0.5%) reported that they had received a doctoral degree.

At the beginning of the questionnaire, participants were informed about the procedure and compensation before they gave their informed consent to participate. Subsequently, participants completed the questionnaire individually without further verbal instructions as all instructions were given directly in the questionnaire. After returning the completed questionnaire, participants were debriefed and received €3 in cash as compensation for their participation.

The study was carried out in accordance with the revised Declaration of Helsinki (World Medical Association, 2013), the "Professional ethical guidelines of the German Association of Psychologists and the German Psychological Society" (BDP & DGPs, 2016), and the "Ethical Research Principles and Test Methods in the Social and Economic Sciences" (RatSWD, 2017). Respondents participated voluntarily and after informed consent was obtained. Participation in the present study could not have any negative consequences for the respondents, and anonymity was ensured at all times.

Material and design

Participants received a questionnaire with a total of eight pages. As a single, between-subjects independent variable, we manipulated the application context under which respondents participated in the study. To this end, participants were instructed to respond either in an application context with high contextual fit to the content of the OCQ (application as scientific journalist), in a context with low contextual fit (application as psychotherapist), or without an application context (honest control condition).

The first page of the questionnaire provided basic information about the approximately 15-minute duration of participation and instructed respondents to complete the questionnaire in a place where they would be undisturbed as well as not to use any assistive devices. On the same page, demographic variables were assessed (age, gender, native language, and highest educational qualification). Depending on the experimental condition, the second page instructed participants to respond to the following parts of the questionnaire as if they were applying for a job as a science journalist or for a job as a psychotherapist, or to respond honestly. In the two application conditions, these instructions were accompanied by a corresponding mock job advertisement. Directly below the advertisement or the instruction to answer honestly, the manipulation check question was presented. Pages 3 through 7 comprised the overclaiming questionnaire (one page) and the two personality scales (two pages each). These three scales were presented in one of six possible orders to eliminate potential order effects. At the top of pages 3 through 7, participants in the application conditions were additionally reminded that they should answer as if they were applying for a job as a science journalist or a job as a psychotherapist, respectively. On the last page, participants were presented with the seriousness check and thanked for their participation.

The combination of three job application contexts and six orders of overclaiming and personality scales resulted in 18 different questionnaire versions; participants were assigned to questionnaire versions at random. This resulted in an even distribution across experimental conditions, with 141 participants (32.6%) allocated to the science journalist application condition, 143 participants (33.1%) to the psychotherapist application condition, and 148 participants (34.3%) to the honest control condition.

Mock job advertisements

The mock job advertisements used in the two application conditions included brief information on the desired profile of potential applicants, the job tasks, and the offers of the potential employer. In the science journalist application condition, the desired profile included extensive general knowledge; successful applicants would allegedly be required to process scientific content and to independently conduct research in complex subject areas. In contrast, the mock job advertisement in the psychotherapist application condition focused on requirements such as high empathy and emotional stability; successful applicants in this condition would allegedly have to provide therapy sessions and emergency consultation in crisis situations. An English translation of the mock job advertisements, originally presented in German language, is provided in Fig. 1.

Fig. 1 Mock job advertisements used in the application conditions

To expand our team, we are currently looking for a
Science Journalist (m/f/d)

Your profile:
- University degree
- Comprehensive general knowledge
- Interest in a wide range of scientific and social topics
- Ability to communicate complex issues in a comprehensible way

Your tasks:
- Autonomous research into complex subject areas
- Processing of scientific content into exciting articles
- Strategic planning and editing of print and online media

Our offer:
- Multifaceted employment at the intersection of science and society
- Multidisciplinary team with many years of experience
- Opportunities for professional and personal development

To expand our team, we are currently looking for a
Psychotherapist (m/f/d)

Your profile:
- University degree and license to practice as a psychotherapist
- Experience in dealing with individuals with mental health issues in a sensitive, empathic, and respectful manner
- Empathy, distinct ability to take on different perspectives
- Emotional stability, also in stressful situations

Your tasks:
- Planning and conducting patient-centered individual & group therapies
- Providing emergency consultation services for individuals in acute crisis

Our offer:
- Self-determined and multifaceted employment
- Multidisciplinary team with many years of experience
- Opportunities for professional and personal development

Note. Left panel: science journalist application condition, right panel: psychotherapist application condition. Originally presented in German language, English translation provided by the authors.

Manipulation check

To ensure that participants had read and understood the instructions pertaining to their experimental condition, a manipulation check was presented directly below the mock job advertisements (application conditions) or the instruction to answer honestly (honest control condition). The manipulation check read: "To verify that you have read and understood these instructions, please mark the correct statement below." Participants had to choose from one of four answer options: (a) "I am to present myself as negatively as possible when answering this questionnaire," (b) "I am to answer this questionnaire as if I were in a job application situation," (c) "I am to simply tick anything without thinking about it when answering this questionnaire", or (d) "I am to answer this questionnaire in a way that applies to me personally at the moment." In the application conditions, participants who checked option (b) were identified as having passed the manipulation check; in the honest control condition, participants had to check option (d) to pass the manipulation check. Participants who failed the manipulation check were excluded from further analyses.

Overclaiming questionnaire

The overclaiming questionnaire employed in the current study was developed by the authors in an unpublished pilot study and included a total of 42 items from six general knowledge domains: (1) language and literature, (2) history and politics, (3) sports, health, and nutrition, (4) arts and culture, (5) natural sciences and mathematics, and (6) leisure and entertainment. Each domain comprised seven items, five of which were real words ("reals", e.g., "Oxymoron") and two of which were fictitious words ("foils", e.g., "Randel herb"), respectively. Participants were instructed to rate how well they knew each item on a 7-point scale ranging from "not at all" (1) to "very well" (7). The full questionnaire is available from the authors upon request.

A measure for participants' tendency to overclaim their general knowledge (*OCQ bias*) was calculated based on the principals of signal detection theory (Green & Swets, 1966; Macmillan & Creelman, 2005) following the procedure detailed in Paulhus et al. (2003). To this end, we determined hit (H_i) and false alarm rates (FA_i) for all six possible thresholds on the familiarity scale (that is, familiarity ratings for reals and foils higher than 1, 2, 3, 4, 5, or 6, respectively) and calculated the response criterion c for each threshold using the formula $c_i = -0.5 * (z[H_i] + z[FA_i])$. The OCQ bias score was then computed as the mean of these criteria c_1 to c_6, and reversed so that higher scores would reflect a stronger tendency to overclaim (cf. Paulhus et al., 2003), *OCQ bias* $= -1 * c$. In addition to the OCQ bias, overclaiming questionnaires also allow for the computation of OCQ accuracy, an

ability measure of participants' general knowledge. While beyond the scope of the current manuscript, results from an exploratory analysis of OCQ accuracy are provided as an electronic supplement via the Open Science Framework (see https://osf.io/5ysrm/).

Personality scales

Achievement motivation was measured via the short form of the *Achievement Motivation Inventory* (Leistungsmotivationsinventar-Kurzform [LMI-K]; Schuler & Prochaska, 2001). This inventory includes 30 statements with positive polarity (e.g., "I like to find tasks where I can test my skills.") to which participants had to indicate their agreement on a 7-point scale ranging from "strongly disagree" (1) to "strongly agree" (7). To form a global measure of *achievement motivation*, the mean score was calculated across all 30 items; higher scores reflected a generally higher achievement motivation. The internal consistency of the personality scales in the current study was determined using McDonald's ω (Hayes & Coutts, 2020). For the LMI-K, the observed consistency was high, ω = .92, indicating that the scale measured a homogeneous construct.

To measure personality traits favorable for the profession of a psychotherapist (*therapist personality*), 30 items from three subscales of the *Bochum Business-Focused-Inventory of Personality* (Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung [BIP]; Hossiep & Paschen, 1998) were used. These items were selected by the authors based on their content fit to the psychotherapist application context. Ten items each were taken from the BIP subscales sensitivity (e.g., "I can adapt very well to a wide variety of people."), sociability (e.g., "I avoid provoking others."), and emotional stability (e.g., "I don't spend a lot of time thinking about personal problems."). Of the 30 items, 16 had positive polarity (i.e., with agreement indicating high trait levels) and 14 had negative polarity (i.e., with agreement indicating low trait levels). Participants were required to indicate their agreement with each statement on a 7-point scale ranging from "strongly disagree" (1) to "strongly agree" (7). After inverting the answers to the 14 items with negative polarity, an overall score for therapist personality was formed by calculating the mean across all 30 items; higher scores therefore reflected higher sensitivity, sociability, and emotional stability, all of which were considered personality traits favorable for the profession of a psychotherapist. The internal consistency of the therapist personality scale was high, ω = .93.

Additionally, responses to the LMI-K and the therapist personality scale were used to form an overall *extreme positivity* score as a proxy for participants' tendency to fake (cf. Dunlop et al., 2020, p. 787f). For each participant, we calculated how many of the 60 items from the two personality scales they had answered with the most extreme response

option in the desirable direction (i.e., "strongly agree" [7] for items with positive polarity, "strongly disagree" [1] for items with negative polarity). Extreme positivity therefore had a theoretical range of 0 to 60, with higher values assumed to reflect higher faking tendencies.

Seriousness check

At the end of the questionnaire, a seriousness check was used to determine whether participants had diligently participated in the study. Seriousness checks have been demonstrated to enhance data quality in surveys, as they control for nonserious answering behavior, thereby decreasing statistical noise and improving experimental power (Aust et al., 2013). The seriousness check read: "Have you answered all the questions in this questionnaire seriously and according to the instructions? Please answer honestly; your answer will not affect your financial compensation." Participants had to choose from one of two answer options, that is, (a) "Yes, I have answered all questions seriously and according to the instructions.", or (b) "No, I just wanted to 'have a look', or I got distracted; my responses should better not be analyzed." Irrespective of experimental condition, participants choosing option (a) were identified as having passed the seriousness check; participants who failed the seriousness check were excluded from further analyses.

Statistical analyses

All measures were tested for normality before selecting appropriate inferential statistical procedures. Shapiro–Wilk tests indicated violations of the normality assumption for OCQ bias, achievement motivation measured via the LMI-K, therapist personality measured via the BIP, and extreme positivity, respectively. A graphical inspection of the distribution of these variables gave reason to believe that these violations could possibly be explained by minor floor (OCQ bias and extreme positivity) and ceiling effects (therapist personality), as well as by the influence of a few outliers (achievement motivation), for which Shapiro-Wilks tests are known to be highly sensitive. Against this background and for the sake of consistency, we decided to use nonparametric procedures for all analyses, as these procedures are considered robust to violations of the normality assumption and to the influence of outliers. Overall mean differences between experimental groups were assessed via Kruskal–Wallis tests; in case of a significant overall effect, pairwise group differences were assessed using Mann–Whitney-U tests with Bonferroni-Holm correction to prevent α error inflation (Holm, 1979). Statistical associations between measures were assessed via Spearman's rank correlation coefficients (ρ). Analyses for which directed hypotheses had been formulated before data collection were performed as one-tailed tests

(relating to H1a-H1c; H2a, H2b, H2e-H2g; H3a, H3b; H4a-H4c); in contrast, analyses for which no directed hypotheses had been formulated were performed as two-tailed tests (relating to H2c; H3c; exploratory analyses of differences in strength of associations between experimental conditions). For all analyses, a significance level of $\alpha = .05$ was assumed.

Results

Descriptive statistics for all measures in the total sample and separately for the three experimental conditions are provided in Table 1.

Differences between experimental conditions

For OCQ bias, a significant overall effect of experimental condition was observed, $H(2) = 71.71$, $p < .001$. Pairwise comparisons showed significant differences between all conditions. Expectedly, and in line with H1a, H1b and H1c, OCQ bias scores were higher in the science journalist application condition than in the psychotherapist application condition, $U = 8866.00$, $z = -1.76$, $p = .040$ (one-tailed), $d = 0.21$; higher in the science journalist application condition than in the honest control condition, $U = 4766.00$, $z = -7.98$, $p < .001$ (one-tailed), $d = 1.14$; and higher in the psychotherapist application condition than in the honest control condition, $U = 6040.00$, $z = -6.33$, $p < .001$ (one-tailed), $d = 0.89$.

For achievement motivation measured via the LMI-K, a significant difference between experimental conditions was observed, $H(2) = 102.56$, $p < .001$. Pairwise comparisons showed that LMI-K scores did not differ between the science journalist and the psychotherapist application conditions, thus providing no support for the undirected H2c, $U = 9573.00$, $z = -0.73$, $p = .462$, $d = 0.10$; in line with H2a and H2b, however, LMI-K scores were expectedly higher in the science journalist application condition than in the honest control condition, $U = 4142.00$, $z = -8.86$, $p < .001$ (one-tailed), $d = 1.17$; and higher in the psychotherapist application condition than in the honest control condition, $U = 4425.00$, $z = -8.58$, $p < .001$ (one-tailed), $d = 1.11$.

For the therapist personality measure assessed via selected items from the BIP, a significant overall effect of experimental condition was observed, $H(2) = 180.59$, $p < .001$. Expectedly, and providing support for hypotheses H2d to H2f, scores on the therapist personality scale were highest in the psychotherapist application condition, and higher than in the science journalist application condition, $U = 7440.00$, $z = -3.82$, $p < .001$ (one-tailed), $d = 0.42$; or the honest control condition, $U = 1905.00$, $z = -12.09$, $p < .001$ (one-tailed), $d = 2.02$, respectively; scores were also higher in the science journalist application condition than in the honest control condition, $U = 2993.00$, $z = -10.48$, $p < .001$ (one-tailed), $d = 1.56$.

Table 1 Descriptive statistics for all measures in the total sample and by experimental condition

Measure	M	SD	Min	Max	Md	Range
OCQ bias: total sample	-0.4	0.8	-1.8	2.0	-0.6	3.7
application as science journalist	-0.1	0.9	-1.6	2.0	-0.2	3.6
application as psychotherapist	-0.3	0.8	-1.8	2.0	-0.4	3.7
honest control condition	-0.9	0.4	-1.7	0.4	-0.9	2.1
Achievement motivation (LMI-K): total sample	5.4	0.7	3.0	7.0	5.5	4.0
application as science journalist	5.7	0.6	3.1	7.0	5.7	3.9
application as psychotherapist	5.6	0.6	3.0	6.7	5.7	3.7
honest control condition	4.9	0.7	3.0	6.7	4.9	3.7
Therapist personality (BIP): total sample	5.1	0.9	2.1	6.8	5.2	4.7
application as science journalist	5.4	0.8	3.5	6.7	5.5	3.2
application as psychotherapist	5.7	0.7	3.7	6.8	5.8	3.1
honest control condition	4.2	0.7	2.1	5.9	4.3	3.8
Extreme positivity: total sample	15.9	12.9	0	51	12	51
application as science journalist	18.1	12.8	0	50	15	50
application as psychotherapist	22.8	12.4	0	51	22	51
honest control condition	7.0	7.5	0	50	4.5	50

Overall $N = 432$; application as science journalist condition: $n = 141$; application as psychotherapist condition: $n = 143$; honest control condition: $n = 148$

Finally, for the extreme positivity measure which served as a proxy of participants' faking behavior (cf. Dunlop et al., 2020, p. 787f), a significant overall effect of experimental condition was observed, $H(2) = 136.05$, $p < .001$. Pairwise comparisons via Mann–Whitney-U tests revealed significant differences between all groups; notably, the number of extreme responses was higher in the psychotherapist than in the science journalist application condition, thus providing support for the undirected H3c, $U = 7761.50$, $z = -3.35$, $p < .001$, $d = 0.37$; expectedly, and in line with H3a and H3b, participants in both the science journalist application condition, $U = 4437.00$, $z = -8.45$, $p < .001$ (one-tailed), $d = 1.07$, and the psychotherapist application condition, $U = 2721.00$, $z = -10.96$, $p < .001$ (one-tailed), $d = 1.54$, showed higher extreme positivity than participants in the honest control condition.

Associations of overclaiming with personality scales and extreme positivity

An overview of spearman rank correlations for the association of OCQ bias with all other measures is provided in

Table 2. In the total sample, OCQ bias expectedly showed substantial positive associations with both personality measures and with extreme positivity, thus supporting H4a to H4c.

An exploratory comparison of correlation coefficients based on Fisher's z-test (Fisher, 1925; Myers & Sirois, 2006) implemented in the software package *cocor* (Diedenhofen & Musch, 2015) revealed that the association of OCQ bias with achievement motivation was comparable across experimental conditions: application as science journalist vs. application as psychotherapist, $z = -0.40$, $p = .689$; application as science journalist vs. honest control, $z = 0.27$, $p = .790$; application as psychotherapist vs. honest control, $z = 0.67$, $p = .501$. On a descriptive level, the association of OCQ bias with therapist personality and with extreme positivity appeared to be higher in the application conditions compared to the honest control condition. However, pairwise comparisons between conditions remained insignificant for the association of OCQ bias and therapist personality: application as science journalist vs. application as psychotherapist: $z = 0.75$, $p = .456$; application as science journalist vs. honest

Table 2 Spearman rank correlations (ρ) of OCQ Bias with all other measures

	OCQ bias			
	Total sample	Application as science journalist	Application as psychotherapist	Honest control condition
Achievement motivation (LMI-K)	.38*	.23*	.28*	.20*
Therapist personality (BIP)	.37*	.25*	.16*	.05
Extreme positivity	.33*	.19*	.23*	.06

* $p < .05$ (one-tailed)

control: $z = 1.77$, $p = .077$; application as psychotherapist vs. honest control: $z = 1.02$, $p = .307$; as well as for the association of OCQ bias and extreme positivity: application as science journalist vs. application as psychotherapist: $z = -0.34$, $p = .734$; application as science journalist vs. honest control: $z = 1.13$, $p = .257$; application as psychotherapist vs. honest control: $z = 1.48$, $p = .138$.

A further exploration of our data showed that the two personality scales, achievement motivation and therapist personality, were strongly associated in the total sample, $\rho = .56$, $p < .001$. This strong association was mainly driven by participants in the science journalist application condition, $\rho = .60$, $p < .001$; associations were also significant, but weaker, in the psychotherapist application condition, $\rho = .36$, $p < .001$, and in the honest control condition, $\rho = .19$, $p = .020$. Pairwise comparisons revealed that the association between achievement motivation and therapist personality was significantly stronger in the science journalist application condition compared to the psychotherapist application condition, $z = 2.61$, $p = .009$; and compared to the honest control condition, $z = 4.18$, $p < .001$. A comparison between the psychotherapist application condition and the honest control condition remained insignificant, $z = 1.55$, $p = .121$.

The dataset generated during the current study and the program code used for the main analyses are available from the Open Science Framework (see https://osf.io/5ysrm/).

Discussion

The current study aimed at validating and expanding the recent finding that the overclaiming technique (OCT; Paulhus et al., 2003) is capable of capturing applicant faking behavior if the content of the overclaiming questionnaire (OCQ) used fits the application context well (Diedenhofen et al., 2022; Dunlop et al., 2020). In contrast to the overall disappointing validity evidence with respect to the utility of OCQs as an indicator of individual differences in self-presentation tendencies (e.g., Ludeke & Makransky, 2016; Müller & Moshagen, 2019b), this would suggest that such measures may be more valid in the specific high-demand context of application scenarios in which individuals' self-enhancement motivation should be generally elevated, as compared to low-demand situations in which other factors such as cognitive biases may be driving individual over-claiming tendencies (Müller & Moshagen, 2018; Paulhus, 2011). In a one-factorial experiment, we collected participants' responses to an OCQ in an application context with high content-context fit (application as a science journalist), in a context with low content-context fit (application as a psychotherapist), or in an honest control condition. Additionally, we employed two conventional personality scales. One scale captured achievement motivation which was presumably relevant to any application situation; the other scale measured a composite of traits specifically relevant to a psychotherapist (therapist personality). Finally, as a proxy for applicant faking, we assessed the number of extreme responses in the desirable direction to the items of the personality scales.

Sensitivity of OCQ bias to applicant faking behavior

Our results generally support the assumption that the utility of OCQs is a function of their content fit to the specific application context. As expected, the OCQ bias scores were higher in both application conditions compared to the control condition (H1a and H1b). More importantly, they were also higher in the application condition which fit the "general knowledge" content of the OCQ best (science journalist) compared to the application condition with lower content fit (psychotherapist; H1c). This difference should be interpreted with some caution; had H1c been formulated as an undirected hypothesis and the difference been tested using a two-tailed test respectively, it would not have reached statistical significance. However, because a priori expectations clearly justified a one-tailed hypothesis test, the corresponding difference between application groups implies that participants who allegedly applied for a science journalism job specifically felt that overclaiming their general knowledge would be instrumental to a successful application—and most importantly, that the OCQ was sensitive to their faking behavior. Positive and substantial associations of the OCQ bias with both personality measures and with extreme positivity in the total sample (H4a to H4c) and in the application conditions further support the conclusion that OCQs may be a promising means of capturing applicant faking, especially if the content-context fit is taken into account (Diedenhofen et al., 2022; Dunlop et al., 2020). To some extent, this last finding may be qualified by studies that attribute stronger correlations between self-report measures in fake good compared to control conditions to some inattentive participants not following the fake good instructions, producing outlier values and thus artificially inflating parametric correlations (Paulhus et al., 1995). We however argue that a full explanation of our correlation pattern with this phenomenon is unlikely; first, because we reminded participants in the application conditions repeatedly and at salient points in the questionnaire what they were supposed to do; second, because we used a manipulation check to verify that participants had read and understood the instructions and excluded those who did not pass this check; and third, because we used nonparametric Spearman correlations to assess pairwise associations between variables, which are far less susceptible to the influence of potential outlier values than parametric procedures.

Susceptibility of personality scales to applicant faking behavior

Furthermore, our findings add to the existing body of literature suggesting that applicants can and will fake their answers to conventional personality scales if given the opportunity, even under mock application conditions as in the current study. Achievement motivation as measured via the LMI-K was substantially higher in both application conditions compared to the honest control condition ($d = 1.17$ and $d = 1.11$, respectively; H2a and H2b). The same was true for the therapist personality composite measure of sensitivity, sociability, and emotional stability based on items from the BIP ($d = 2.02$ and $d = 1.56$, respectively; H2d and H2e). Notably, however, scores on this measure were even higher in the psychotherapist application condition as compared to the science journalist application condition ($d = 0.42$; H2f). Due to our experimental design, these differences are hardly attributable to real differences in the underlying, supposedly stable, personality constructs; rather, they are most likely a product of participants successfully incorporating the specific requirements of the mock job application scenarios into their faking behavior. This conclusion is further supported by differences between the application conditions and the honest control condition with respect to extreme positivity ($d = 1.07$ and $d = 1.54$, respectively; H3a and H3b) as well as between psychotherapist and science journalist application conditions in this measure ($d = 0.37$; H3c) of which the latter was presumably driven by participants in the psychotherapist condition excessively overstating positive aspects of their personality relevant to the job demands. In summary, our results add to the findings of previous empirical studies that suggested a strong susceptibility of personality scales to applicant faking (e.g., Huber et al., 2021; Tett & Simonet, 2021).

Correlation analyses further revealed that the association of OCQ bias with the personality measures and with extreme positivity were positive and substantial in the overall sample (H4a to H4c), and descriptively higher in both application conditions as compared to the honest control condition. However, these differences failed to reach statistical significance, possibly due to insufficient statistical power. Assuming an acceptable power of at least .80, the sample sizes of the experimental groups (ranging from 141 to 148) only allow for the detection of effect sizes of at least $q = 0.33$ with Fisher's z-test (1925), which corresponds to a difference between $r = .20$ and $r = .49$, for example. Nonetheless, the presence of significant associations in the application conditions and the absence of such associations in the control condition (at least for therapist personality and extreme positivity, see Table 2) allow for the tentative conclusion that the respective variables share common variance specifically, or possibly even exclusively, under application conditions.

Although the overlap is admittedly small, it may have been due to cognitive processes that influenced applicants' faking behavior on all measures collected, including OCQ bias.

Limitations and future research directions

One apparent limitation of our study is that we had to apply mock application contexts rather than collecting data in real application situations. Thus, participants in our study were probably only exposed to a somewhat weaker pressure to present themselves in a positive light than applicants to real jobs may feel. On the other hand, our mock applications were free of any risks associated with being exposed as a faker in real personnel selection contexts, possibly enabling some participants to exaggerate positive aspects of their personality even more than they would in real application processes. Against this background, we can only speculate as to whether the direction and magnitude of the effects we found can be transferred to real job applications. However, the medium to very large effect sizes especially with regard to group differences in the collected measures allow for the cautious assumption that the OCQ bias would also be sensitive to participants' faking behavior in real job application situations. We hence encourage future studies to further investigate the influence of content-context fit on the utility of OCQs in real-world application situations—although such studies are expected to entail lower experimental control in conjunction with considerably increased monetary, time, and organizational costs.

A second limitation can be found in the demographic characteristics and size of our sample. We argue that a community sample mainly consisting of university students may be considered as representative of potential applicants to jobs such as those from the application contexts we induced. However, our results on the utility of the OCQ bias measure cannot be generalized to application contexts in general and may thus not apply to other situations such as more mature applicants seeking higher-level positions. Regarding sample size, it should be noted that it proved to be sufficient for detecting effects of the experimental manipulation as well as substantial associations between the measures used. However, exploratory pairwise comparisons of associations between conditions mostly remained at non-significant levels. Given sufficient resources, future studies should consider the collection of larger samples including more heterogeneous participants and application scenarios to address the apparent sample limitations of the current study.

In conclusion, the current study provided further experimental evidence supporting the assumption that the utility of OCQs in detecting applicant faking may be a function of their content fit to the specific requirements of the application context. While the apparent susceptibility of conventional personality scales to faking is problematic in terms of

validity, the sensitivity of the OCQ bias measure to faking indicates its potential usefulness in job application situations, especially when content-context fit is high. Against this background, we argue that OCQs can be considered as a promising means to capture applicant faking, provided that their content meets the specific requirements of the application context.

Data availability The dataset generated during the current study and the program code used for the main analyses are available from the Open Science Framework (see https://osf.io/5ysrm/).

Declarations

Ethics approval The study was carried out in accordance with the revised Declaration of Helsinki (World Medical Association, 2013), the "Professional ethical guidelines of the German Association of Psychologists and the German Psychological Society" (BDP & DGPs, 2016), and the "Ethical Research Principles and Test Methods in the Social and Economic Sciences" (RatSWD, 2017).

Consent Respondents participated voluntarily and after informed consent was obtained. Participation in the present study could not have any negative consequences for the respondents, and anonymity was ensured at all times.

Competing interests The authors have no relevant financial or non-financial interests to disclose.

References

Aust, F., Diedenhofen, B., Ullrich, S., & Musch, J. (2013). Seriousness checks are useful to improve data validity in online research. *Behavior Research Methods, 45*(2), 527–535. https://doi.org/10.3758/s13428-012-0265-2

BDP & DGPs. (2016). *Berufsethische Richtlinien des Berufsverbandes Deutscher Psychologinnen und Psychologen e.V. und der Deutschen Gesellschaft für Psychologie e.V. [Professional ethical guidelines of the German Association of Psychologists and the German Psychological Society]*. Retrieved November 7, 2022, from https://www.dgps.de/fileadmin/user_upload/PDF/berufsethik-foederation-2016.pdf

Bensch, D., Paulhus, D. L., Stankov, L., & Ziegler, M. (2019). Teasing apart overclaiming, overconfidence, and socially desirable responding. *Assessment, 26*(3), 351–363. https://doi.org/10.1177/1073191117700268

Bing, M. N., Kluemper, D., Davison, H. K., Taylor, S., & Novicevic, M. (2011). Overclaiming as a measure of faking. *Organizational Behavior and Human Decision Processes, 116*(1), 148–162. https://doi.org/10.1016/j.obhdp.2011.05.006

Diedenhofen, B., & Musch, J. (2015). cocor: A comprehensive solution for the statistical comparison of correlations. *PLoS One, 10*(4), e0121945. https://doi.org/10.1371/journal.pone.0121945

Diedenhofen, B., Hoffmann, A., & Müller, S. (2022). Detecting applicant faking with a context-specific overclaiming questionnaire. *Manuscript submitted for publication*.

Dunlop, P. D., Bourdage, J. S., de Vries, R. E., McNeill, I. M., Jorritsma, K., Orchard, M., Austen, T., Baines, T., & Choe, W.-K. (2020). Liar! Liar! (when stakes are higher): Understanding how the overclaiming technique can be used to measure faking in personnel selection. *Journal of Applied Psychology, 105*(8), 784–799. https://doi.org/10.1037/apl0000463

Feeney, J. R., & Goffin, R. D. (2015). The Overclaiming Questionnaire: A good way to measure faking? *Personality and Individual Differences, 82*, 248–252. https://doi.org/10.1016/j.paid.2015.03.038

Fisher, R. A. (1925). *Statistical methods for research workers* (11th ed.). Oliver and Boyd.

Goecke, B., Weiss, S., Steger, D., Schroeders, U., & Wilhelm, O. (2020). Testing competing claims about overclaiming. *Intelligence, 81*, 101470. https://doi.org/10.1016/j.intell.2020.101470

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. John Wiley.

Grosz, M. P., Lösch, T., & Back, M. D. (2017). The narcissism-overclaiming link revisited. *Journal of Research in Personality, 70*, 134–138. https://doi.org/10.1016/j.jrp.2017.05.006

Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. But…. *Communication Methods and Measures, 14*(1), 1–24. https://doi.org/10.1080/19312458.2020.1718629

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65–70.

Hossiep, R., & Paschen, M. (1998). *Das Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung [The Bochum Business-Focused-Inventory of Personality]*. Hogrefe.

Huber, C. R., Kuncel, N. R., Huber, K. B., & Boyce, A. S. (2021). Faking and the validity of personality tests: An experimental investigation using modern forced choice measures. *Personnel Assessment and Decisions, 7*(1), 20–30. https://doi.org/10.25035/pad.2021.01.003

Keller, L., Bieleke, M., Koppe, K.-M., & Gollwitzer, P. M. (2021). Overclaiming is not related to dark triad personality traits or stated and revealed risk preferences. *PLoS One, 16*(8), e0255207. https://doi.org/10.1371/journal.pone.0255207

Klehe, U.-C., Kleinmann, M., Hartstein, T., Melchers, K. G., Konig, 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*(4), 273–302. https://doi.org/10.1080/08959285.2012.703733

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: A literature review. *Quality & Quantity, 47*, 2025–2047. https://doi.org/10.1007/s11135-011-9640-9

Ludeke, S. G., & Makransky, G. (2016). Does the Over-Claiming Questionnaire measure overclaiming? Absent convergent validity in a large community sample. *Psychological Assessment, 28*(6), 765–774. https://doi.org/10.1037/pas0000211

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Erlbaum.

Müller, S., & Moshagen, M. (2018). Overclaiming shares processes with the hindsight bias. *Personality and Individual Differences, 134*, 298–300. https://doi.org/10.1016/j.paid.2018.06.035

Müller, S., & Moshagen, M. (2019a). Controlling for response bias in self-ratings of personality: A comparison of impression management scales and the overclaiming technique. *Journal of Personality Assessment, 101*(3), 229–236. https://doi.org/10.1080/00223891.2018.1451870

Müller, S., & Moshagen, M. (2019b). True virtue, self-presentation, or both?: A behavioral test of impression management and overclaiming. *Psychological Assessment, 31*(2), 181–191. https://doi.org/10.1037/pas0000657

Myers, L., & Sirois, M. J. (2006). Spearman correlation coefficients, differences between. In *Encyclopedia of statistical sciences*. John Wiley & Sons, Inc. https://doi.org/10.1002/0471667196.ess5050.pub2

Paulhus, D. L. (2011). Overclaiming on personality questionnaires. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 151–164). Oxford University Press.

Paulhus, D. L., & Harms, P. D. (2004). Measuring cognitive ability with the overclaiming technique. *Intelligence, 32*(3), 297–314. https://doi.org/10.1016/j.intell.2004.02.001

Paulhus, D. L., Bruce, M. N., & Trapnell, P. D. (1995). Effects of self-presentation strategies on personality profiles and their structure. *Personality and Social Psychology Bulletin, 21*(2), 100–108. https://doi.org/10.1177/0146167295212001

Paulhus, D. L., Harms, P. D., Bruce, M. N., & Lysy, D. C. (2003). The over-claiming technique: Measuring self-enhancement independent of ability. *Journal of Personality and Social Psychology, 84*(4), 890–904. https://doi.org/10.1037/0022-3514.84.4.890

Phillips, D. L., & Clancy, K. J. (1972). Some effects of social desirability in survey studies. *American Journal of Sociology, 77*, 921–940. https://doi.org/10.1086/225231

Randall, D. M., & Fernandes, M. F. (1991). The social desirability response bias in ethics research. *Journal of Business Ethics, 10*(11), 805–817. https://doi.org/10.1007/Bf00383696

RatSWD. (2017). *Forschungsethische Grundsätze und Prüfverfahren in den Sozial- und Wirtschaftswissenschaften [Ethical research principles and test methods in the social and economic sciences].* Rat für Sozial- und Wirtschaftsdaten (RatSWD).

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. https://doi.org/10.1037/0021-9010.83.4.634

Roulin, N., Krings, F., & Binggeli, S. (2016). A dynamic model of applicant faking. *Organizational Psychology Review, 6*(2), 145–170. https://doi.org/10.1177/2041386615580875

Schuler, H., & Prochaska, M. (2001). *Leistungsmotivationsinventar: Dimensionen berufsbezogener Leistungsorientierung [Achievement motivation inventory: Dimensions of job-related achievement orientation].* Hogrefe.

Tett, R. P., & Simonet, D. V. (2021). Applicant faking on personality tests: Good or bad and why should we care? *Personnel Assessment and Decisions, 7*(1), 6–19. https://doi.org/10.25035/pad.2021.01.002

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133*, 859–883. https://doi.org/10.1037/0033-2909.133.5.85917723033

World Medical Association. (2013). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA, 310*, 2191–2194. https://doi.org/10.1001/jama.2013.281053

Ziegler, M., Kemper, C., & Rammstedt, B. (2013). The Vocabulary and Overclaiming Test (VOC-T). *Journal of Individual Differences, 34*(1), 32–40. https://doi.org/10.1027/1614-0001/a000093