

Faking self-reports of health behavior: a comparison between a within- and a between-subjects design

Viktoria S. Egele , Laura H. Kiefer and Robin Stark

Department of Educational Research, Saarland University, Saarbrücken, Germany

ABSTRACT

Background: This study examines people's ability to fake their reported health behavior and explores the magnitude of such response distortion concerning faking of preventive health behavior and health risk behavior. As health behavior is a sensitive topic, people usually prefer privacy about it or they wish to create a better image of themselves (Fekken et al., 2012; Levy et al., 2018). Nevertheless, health behavior is often assessed by self-report questionnaires that are prone to faking. Therefore, it is important to examine the possible impact of such faking.

Methods: To replicate the findings and test their robustness, two study designs were realized. In the within-subjects-design, 142 participants repeatedly answered a health behavior questionnaire with an instruction to answer honestly, fake good, and fake bad. In the between-subjects design, 128 participants were randomly assigned to one of three groups that filled out the health behavior questionnaire with only one of the three instructions.

Results: Both studies showed that successful faking of self-reported preventive and health risk behavior was possible. The magnitude of such faking effects was very large in the within-subjects design and somewhat smaller in the between-subjects design.

Conclusion: Even though each design has its inherent merits and problems, caution is indicated regarding faking effects.

ARTICLE HISTORY

Received 17 June 2021


Accepted 6 October 2021

KEYWORDS

Health behavior; assessment; self-report; faking; response distortion

It is the main goal to prevent non-communicable diseases and improve people's health behavior, not only for researchers but also for general practitioners and health care workers (World Health Organization [WHO], 2013). Health behavior is defined as "overt behavioral patterns, actions, and habits that relate to health maintenance, to health restoration and to health improvement" (Gochman, 1997, p. 3). Thus, various behaviors are covered by this definition. Often, preventive behavior that improves and protects health like eating a healthy diet and performing sufficient physical activity is distinguished from risk behavior like smoking or excessive alcohol use, which endangers health and should be prevented or reduced to a minimum (Kasl & Cobb, 1966).

CONTACT Viktoria S. Egele  viktorias.egele@uni-saarland.de  Department of Educational Research, Saarland University, Saarbrücken 66123, Germany

 Supplemental data for this article can be accessed at <https://doi.org/10.1080/21642850.2021.1991803>

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Although health researchers can use many innovative techniques like wearables, physiological measures, or ambulatory assessment to assess health behavior, self-reports are still the most frequent measure (Sattler et al., 2021). Self-reports are easy to use, economic in terms of management and time and they are cost-efficient (Foa, Cashman, Jaycox, & Perry, 1997). At the same time, they are subject to criticism, among other things because of their susceptibility to errors and wilful response distortion, leading to a limited validity of self-reported data (Griffith, Chmielowski, & Yoshita, 2007). Objective and subjective measures often show significant correlations in relative terms, for example, correlations of $r = .21$ to $r = .52$ are reported for physical activity measured by self-reports and accelerometers (Atienza & King, 2005; Nelson, Taylor, & Vella, 2019). However, it is striking that the absolute measures differ significantly. For example, people report about twice as much physical activity time in self-reports as objectively measured data (Atienza & King, 2005). Systematic reviews claim that self-reported health behavior questionnaires may succeed at ranking individuals concerning their health behavior but cannot provide valid results concerning the absolute quantity of physical activity (Helmerhorst, Brage, Warren, Besson, & Ekelund, 2012). A similar picture emerges for diet, smoking, and alcohol consumption. Participants overreport fruit and vegetable consumption in self-reports compared to objectively assessed intake (Lechner, Brug, & De Vries, 1997). In the German National Nutrition Survey II, the correlation of subjective and objective measures of reported fruit and vegetable consumption ranged from $.24 \leq r \leq .40$ (Straßburg, Eisinger-Watzl, Krems, Roth, & Hoffmann, 2019). In their review, Gorber, Schofield-Hurwitz, Hardt, Levasseur, and Tremblay (2009) claim that smoking is underreported in self-reports compared to objective measures. For example, smoking prevalence is often underestimated when assessed by self-report versus using blood cotinine levels (Lewis et al., 2003) or urinary cotinine concentration (Hwang, Kim, Lee, Jung, & Park, 2018). For the comparison of self-reported and objectively measured alcohol consumption, correlations of $r = .27$ were reported. Yet, over 50% of participants denying consuming alcohol in the last 30 days were tested positive for Phosphatidylethanol, an objective indicator for drinking (Littlefield et al., 2017).

As explanations for the discrepancies between the objective and subjective measures, in addition to memory effects and biases due to reference points, the deliberate distortion of answers in self-reports is often discussed, to present oneself as more socially desirable (Atienza & King, 2005). It is well documented that people sometimes alter their responses to benefit from the creation of the desired impression (Crowne & Marlowe, 1960; Edwards, 1957; Furnham, 1986; Locander, Sudman, & Bradburn, 1976; Mazar & Ariely, 2006; McCrae & Costa, 1983; Mensch & Kandel, 1988; Nederhof, 1985; Norman, 1967).

The prevailing assumption about dishonest behavior is that people act completely purposively in every situation according to the maxim of the greatest gain. Insofar as dishonest behavior means maximizing profit, people behave dishonestly (Henrich et al., 2001; Morgan, 2006). In doing so, they consider three factors: the benefit that could be gained from dishonest behavior, the probability of being caught in the dishonest behavior, and the expected punishment if caught. The action alternative that maximizes personal gain becomes the guiding factor (Becker, 1968).

Health behavior is a highly delicate topic and may thus be particularly susceptible to dishonest reporting, also because it might contain information that can be socially

unacceptable or even illegal (Fekken, Holden, McNeill, & Wong, 2012). Thus, dishonest reporting of health behavior may lead to significant benefits. Since the probability of being caught in dishonest reporting is relatively low as the validity of self-reports can often not be checked, people might create a desired impression of themselves when reporting their health behavior. Previous research showed that the majority of patients admit restricting information given to their clinicians and not being entirely honest concerning their health behavior (Levy et al., 2018), and also that self-report measures of health behavior are susceptible to response distortion (Fekken et al., 2012). It is thus questionable whether self-reported health behavior yields a diagnostic value. Although trying to reduce nondisclosure and faking seems obvious, the degree of dishonesty and the extent of response distortion remain unclear.

Therefore, the following research investigates people's practical skills to distort their responses in a health behavior questionnaire to create the desired impression and subsequently, estimates the magnitude of such response distortion. To test the robustness of the findings, the customary within-subjects design of faking studies is backed up by a between-subjects design to profit from the advantages and insights of both designs concerning the research question. So, the rationale of this study is to investigate how large such faking effects may be when people are instructed to alter their responses accordingly. The studies do not investigate whether faking happens in practice and how large such field effects may be.

Dishonesty and faking

DePaulo, Kashy, Kirkendol, Wyer, and Epstein (1996) claim that people are dishonest in about 30% of their social interactions each week. Dishonesty can take different forms, not only the extent of dishonesty can vary from telling outright untruths to slight self-promotions, but also the direction of response distortion can vary from creating a favorable impression, fake good, to creating an unfavorable impression, fake bad (Cook, 2004). Faking is a response bias in which individuals consciously manipulate their responses to create the desired impression (Griffith et al., 2007; Komar, Brown, Komar, & Robie, 2008; McFarland & Ryan, 2000; Van Hooft & Born, 2012). As a form of other-deceptive enhancement, faking as conscious response biasing has to be differentiated from self-deceptive enhancement, where individuals believe positive self-descriptions to be true (Paulhus, 1984).

For people to control aspects that are relevant to the development of an impression, the capacity, the willingness, and the opportunity to control the information given are important factors (Levashina & Campion, 2006). Those factors are assumed to be linked multiplicatively. The capacity to control the information given consists of cognitive capacity as well as social and verbal competencies. The willingness to modify the information given relies on personality traits and integrity, but also on a cost-benefit analysis that compares the benefits of creating the desired impression to the negative consequences that may arise if a person is caught being more or less dishonest. The opportunity to modify the information given can be illustrated by comparing two assessment methods. Whereas it can be easy to create the desired impression in an interview or a self-report questionnaire, it is nearly impossible to influence objective measurements like

blood parameters. Self-report questionnaires on health behavior are susceptible to faking because all three factors can be given.

Considerations about the adequate study design

Since there is often no way of checking the validity of self-reports, directed-faking designs are employed to study faking (Viswesvaran & Ones, 1999). In these studies, participants receive instructions to distort their responses in a particular manner. Although it is disputable whether directed faking accurately represents faking in practice directed-faking designs benefit from a high degree of control and a direct comparison of honest and faked scores (Furnham, 1990). Most studies about response distortion rely on a within-subjects design where participants are being tested multiple times with different external stimuli (e.g. Fell & König, 2016). The advantages of this design are obvious. For example, the within-subjects design is usually characterized by higher internal validity and higher statistical power. However, concerning the practical relevance of the topic in question, it seems plausible to consider a between-subjects design, too, where each participant is assigned to one faking condition. Between-subjects designs are often discarded for their reliance on randomization and the risk of baseline differences between the different experimental groups, possibly leading to substantial noise and therewith reducing the statistical power. But the strength of between-subjects designs cannot be ignored, as they yield a higher external validity and might be more naturally aligned with the phenomenon in practice (Charness, Gneezy, & Kuhn, 2012). With regards to response distortion, it seems highly implausible that a person repeatedly answers the same questions albeit giving different responses in real settings. The case where participants are exposed to a motivational cue to present themselves either more positively or more negatively seems closer to reality.

Both designs have their merits regarding the scientific insights concerning the research topic. Thus, in the following study, the extent to which it is possible to fake health behavior in a self-report questionnaire is examined both by a within-subjects design (Study 1) as well as by a between-subjects design (Study 2). The two designs do not only permit to answer slightly different research questions: While the within-subjects design allows determining the maximum limits of response distortion of the health behavior scales, the between-subjects design sheds light on the operational level of faking with higher external validity (Viswesvaran & Ones, 1999). Further, this approach allows to compare the results of the two study designs and thus leads to a deeper understanding of faking self-reported health behavior. For example, it could previously be shown that the responses to a stimulus differ significantly dependent on whether participants were evaluating just the one stimulus or evaluating multiple stimuli (Hsee & Zhang, 2004). This referencing effect might cause participants of a within-subjects design to alter their responses according to the instructions. As participants in the between-subjects design are only confronted with one instruction, they do probably not conduct the same careful balancing as the participants in the within-subjects design who may have to adjust their responses to previous responses given. A possible result may be that participants in the within-subjects design do not report their honest health behavior, but rather a constructed concept of behavior that represents the middle of fake good and fake bad behavior.

Therefore, we investigate the following hypotheses: Instructed to report a favorable health behavior, people report significantly healthier dietary habits (H1), more physical activity (H2), less smoking (H3), and less alcohol consumption (H4) than people instructed to report their actual health behavior. Analogously, people instructed to report unfavorable health behavior report significantly unhealthier dietary habits (H1), less physical activity (H2), more smoking (H3), and more alcohol consumption (H4) than people instructed to report their actual health behavior.

In addition, we compare the results of the two study designs and investigate first indicators of referencing effects concerning the reported honest behavior.

Method 1

Sample

As it was not clear whether the effect sizes of directed faking studies in personality inventories are comparable to the effects of faking in health behavior questionnaires, small effects were anticipated (Cohen, 1988). Therefore, the intended sample size calculated by G*Power was 134 participants (Faul, Erdfelder, Lang, & Buchner, 2007). The final sample included 142 German participants (73.2% female) between the ages of 18 and 67 ($M = 25.5$, $SD = 11.4$). 93% of the sample had graduated from high school and nearly 20% of those participants had a university degree.

Instrument

To assess health behavior tailored to the German sample, a questionnaire based on existing questionnaires and newly developed items was developed to assess diet, physical activity, smoking, and alcohol consumption. Independent experts checked that responses to all items could potentially be distorted.

Diet was measured with 15 items based on the recommendations of the German Nutrition Society (German Nutrition Society [DGE], 2010). Analogous to the assessment of physical activity, the average number of days a week in which participants consumed a certain category of food (vegetables, fruit, grains, dairy products, meat, fish, and eggs), as well as the amount of food eaten, was assessed. In addition, the amount of drinking per day was assessed. For comparability, following other inventories for the assessment of dietary habits, the amount of food was assessed in portions and drinking was assessed in liters (e.g. Emanuel, McCully, Gallagher, & Updegraff, 2012).

Physical activity was measured with seven items that were based on the International Physical Activity Questionnaire – Short Form in German but presented in written format (IPAQ-SF, Booth, Owen, Bauman, & Gore, 1996). The IPAQ-SF is a retrospective self-report questionnaire that assesses the physical activity of the past seven days. The questionnaire assesses the number of days and the average time (hours and minutes) spent on physical activity with an open response format. More specifically, the questionnaire assesses moderate and vigorous physical activity as well as walking and sitting behavior. The IPAQ-SF was chosen because of its good psychometric qualities and its implementation in multiple previous studies (Craig et al., 2003; Hagströmer, Oja, & Sjöström, 2006).

Smoking and alcohol consumption were only assessed if participants answered positively to a filter question assessing their basic consumption (i.e. 'Do you/ did you ever smoke?'). If the answer was 'yes', the frequency of substance consumption was asked and the number of alcoholic beverages, and the amount of smoking. The items are listed in the questionnaire in the supplementary material (Questionnaire).

Design

Following most previous studies on response distortion, faking was investigated in a repeated-measures design with three conditions. The online questionnaire was implemented using SoSci Survey (Leiner, 2019) and made available to participants at www.soscisurvey.de Participants were recruited via notices on campus and various online platforms. Participation in the study was not monetarily rewarded. In total, the questionnaire was accessed 756 times. 314 participants started working on the questionnaire and 142 participants completed the survey entirely.

In one condition, participants were asked to answer the questionnaire honestly (in the following: honest condition). In the two other conditions, participants were asked to fake their responses to appear as healthy as plausible (fake good condition), or as unhealthy as plausible (fake bad condition). A pilot study confirmed the effectiveness of the instruction and indicated the necessity to add the phrase 'as (un-)healthy as plausible' to the instruction to prevent unrealistic ceiling- or bottom-effects in the response behavior. The order of conditions was fully randomized to prevent sequencing order effects. The conduct of the study complied with the ethical standards of the responsible committee (The Ethics Committee of the Faculty of Empirical Human and Economic Sciences of Saarland University). Written informed consent was obtained from all subjects before the study.

Analytic Strategies

Data analysis was conducted using IBM SPSS Statistics 24. First, descriptive measures were calculated. For vegetables, fruit, grains, dairy, and eggs, the average amount of portions per day were calculated by multiplying the number of days by the amount of food eaten and then dividing that by seven days. For meat and fish, the recommendations of the DGE are based on weekly consumption, therefore, the average amount of meat and fish consumed per week was calculated by multiplying the number of days of consumption with the reported number of portions. Analogously, the time spent on physical activity per week was calculated by multiplying the number of days of vigorous physical activity per week, respectively moderate physical activity, and walking, with the particular amount of time spent. Drinking and sitting were assessed as daily behavior, thus these measures were not modified. To detect intraindividual differences in the reported diet and physical activity, two repeated-measures multivariate analyses of variance (MANOVA) were conducted. Individual comparisons on each facet of the constructs as well as planned contrasts were conducted to specify the results. For smoking, measures for the frequency of smoking behavior and the number of cigarettes per day were assessed. A repeated-measures analysis of variance (ANOVA) was conducted to

analyze intraindividual differences. Similarly, for alcohol consumption, the frequency and quantity of alcohol consumption were compared across the three conditions through two ANOVAs.

Results 1

To examine intraindividual differences in the reported dietary habits, a repeated-measures MANOVA was conducted. The reported dietary habits differed significantly over the three conditions, *Wilks Lambda* = .19, $F(16, 125) = 32.47$, $p < .001$, $\eta_p^2 = .81$, confirming the hypotheses. Furthermore, all subscales showed similar differences: vegetables, $F(1.70, 238.04) = 185.87$, $p < .001$, $\eta_p^2 = .57$, fruit, $F(1.89, 264.23) = 181.52$, $p < .001$, $\eta_p^2 = .57$, grains, $F(1.42, 199.39) = 59.19$, $p < .001$, $\eta_p^2 = .30$, dairy products, $F(1.82, 255.00) = 37.96$, $p < .001$, $\eta_p^2 = .21$, meat, $F(1.44, 201.28) = 158.18$, $p < .001$, $\eta_p^2 = .53$, fish, $F(1.57, 220.02) = 32.55$, $p < .001$, $\eta_p^2 = .19$, eggs, $F(1.49, 209.04) = 26.85$, $p < .001$, $\eta_p^2 = .16$ and drinking, $F(1.34, 188.12) = 61.00$, $p < .001$, $\eta_p^2 = .31$.

Across all variables, participants reported significantly different dietary habits under the instruction to fake good compared to the instruction to answer the questions honestly respectively to fake bad. For all dietary facets except meat consumption, there were significant differences between the instructions to be honest or fake good. Also, under the instruction to fake bad, the responses given differed significantly from the honest condition, thus confirming the hypotheses (H1). For example, instructed to fake good, participants reported eating significantly more fruit and vegetables than in the honest and the fake bad condition. Similarly, instructed to fake bad, participants reported eating significantly less fruit and vegetables per day than instructed to answer honestly. The descriptive values, as well as the planned contrasts, are shown in [Table 1](#).

For physical activity, differences in vigorous physical activity, moderate physical activity, walking, and sitting were investigated. A repeated-measures MANOVA showed that there were significant differences between the three conditions concerning their reported levels of physical activity, *Wilks Lambda* = .27, $F(8, 133) = 44.41$, $p < .001$, $\eta_p^2 = .73$. All facets of physical activity showed the expected differences (H2). The three conditions differed significantly for the reported vigorous physical activity, $F(2, 289) = 68.28$, $p < .001$, $\eta_p^2 = .33$, moderate physical activity, $F(1.76, 246.92) = 56.00$, $p < .001$, $\eta_p^2 = .29$, walking, $F(1.42, 271.40) = 116.46$, $p < .001$, $\eta_p^2 = .30$ and sitting, $F(1.41, 197.37) = 116.46$, $p < .001$, $\eta_p^2 = .45$. Again, [Table 1](#) displays the descriptive values and the planned contrasts between the three conditions. In support of our hypotheses, the responses given in each of the conditions differed significantly. For vigorous and moderate physical activity as well as walking, higher values signify healthier behavior, contrary to sitting, where higher values equal unhealthier behavior. Thus, instructed to fake bad and report an unfavorable level of physical activity, participants reported less time spent on vigorous and moderate physical activity and walking and reported more time spent sitting than instructed to present themselves honestly respectively fake good. Also, consistent with our hypotheses, the fake good condition differed significantly from the honest condition. For example, instructed to fake good, participants reported spending less time sitting and spending significantly more time walking and with moderate and vigorous physical activity than instructed to be honest.

Table 1. Means, standard deviations, and planned contrasts statistics for reported diet and physical activity in the within-subjects design.

Variable	Fake good		Honest		Fake bad		Fake good vs. Honest		Fake good vs. Fake bad		Fake bad vs. Honest	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>F</i> (1, 140)	η_p^2	<i>F</i> (1, 140)	η_p^2	<i>F</i> (1, 140)	η_p^2
Diet												
Vegetables ^a	2.90	1.44	1.90	1.19	0.69	0.72	90.71***	.39	259.02***	.65	151.53***	.52
Fruit ^a	2.43	1.24	1.42	1.12	0.49	0.70	92.31***	.40	302.99***	.68	112.99***	.45
Grains ^a	1.52	0.84	1.93	1.14	2.98	2.07	21.86***	.14	74.36***	.35	52.36***	.27
Dairy Products ^a	1.03	0.72	1.24	1.28	2.04	1.41	4.22*	.03	59.25***	.30	36.36***	.21
Meat ^b	4.06	5.57	4.15	4.77	15.22	9.37	0.04	.00	182.12***	.57	179.38***	.56
Fish ^b	5.37	5.29	1.45	2.61	2.37	5.11	92.48***	.40	22.89***	.14	3.94*	.03
Eggs ^a	0.77	0.59	0.59	1.02	1.46	1.46	4.30*	.03	27.56***	.16	33.33***	.19
Drinking ^a	2.63	0.75	2.16	0.81	1.44	1.35	64.61***	.32	89.29***	.39	33.24***	.19
Physical Activity												
Vigorous ^c	221.08	379.40	129.60	347.19	46.77	277.60	38.54***	.22	117.01***	.46	35.88***	.20
Moderate ^c	240.01	387.06	149.33	333.64	62.87	240.10	39.79***	.22	81.53***	.37	30.00***	.18
Walking ^c	368.09	504.44	276.76	493.43	135.54	469.74	17.55***	.11	101.73***	.42	54.58***	.28
Sitting ^d	5.2	2.43	7.30	2.75	9.27	4.07	169.07***	.55	162.10***	.55	45.43***	.25

Note.

^aportions per day.^bportions per week.^cminutes per week.^dhours per day.* $p < .05$ ** $p < .01$ *** $p < .001$.

Assuming that the experimental instruction influences the reported smoking behavior, a one-way repeated-measures ANOVA was conducted (H3). The expected differences were detected, $F(1.65, 232.07) = 304.23$, $p < .001$, $\eta_p^2 = .68$. Contrast tests were consistent with the previous findings: instructed to fake bad, participants reported significantly more smoking ($M = 3.22$, $SD = 1.09$) than instructed to present themselves honestly ($M = 1.52$, $SD = 0.91$), $F(1, 141) = 264.58$, $p < .001$, $\eta_p^2 = .63$, and instructed to fake good ($M = 1.13$, $SD = 0.41$), $F(1, 141) = 481.76$, $p < .001$, $\eta_p^2 = .77$. The fake good condition differed significantly from the honest condition, $F(1, 141) = 35.42$, $p < .001$, $\eta_p^2 = .20$.

Concerning the number of smoked cigarettes per day, the three conditions differed significantly, $F(1.13, 98.3) = 59.92$, $p < .001$, $\eta_p^2 = .41$. Contrast tests were consistent with the previous findings: instructed to fake bad ($M = 8.47$, $SD = 0.86$), participants reported smoking more cigarettes per day than instructed to present themselves honestly ($M = 0.73$, $SD = 2.83$), $F(1, 87) = 61.01$, $p < .001$, $\eta_p^2 = .41$, and instructed to fake good ($M = 0.05$, $SD = 0.43$), $F(1, 87) = 63.60$, $p < .001$, $\eta_p^2 = .42$. The fake good condition also differed significantly from the honest condition, $F(1, 87) = 5.76$, $p = .019$, $\eta_p^2 = .06$.

A one-way repeated-measures ANOVA was conducted to investigate whether the reported amount of alcohol consumption differed according to the experimental instruction (H4). In support of our hypotheses, the reported frequency of alcohol consumption in the three conditions differed significantly, $F(1.68, 237.10) = 161.86$, $p < .001$, $\eta_p^2 = .53$. Instructed to fake good, participants reported a lower frequency of alcohol consumption ($M = 1.56$, $SD = 0.87$) than instructed to report their health behavior honestly ($M = 2.29$, $SD = 0.60$), $F(1, 141) = 89.32$, $p < .001$, $\eta_p^2 = .39$, and differed significantly from the fake bad condition ($M = 3.09$, $SD = 0.58$), $F(1, 141) = 232.89$, $p < .001$, $\eta_p^2 = .62$. Also, the reported frequency of alcohol consumption differed significantly between the instruction to report health behavior honestly and fake bad, $F(1, 141) = 119.56$, $p < .001$, $\eta_p^2 = .46$.

Concerning the amount of alcoholic beverages, a repeated-measures ANOVA again showed that the three conditions differed significantly, $F(2, 40) = 55.22$, $p < .001$, $\eta_p^2 = .73$. On average, participants reported consuming less portions of alcoholic beverages when they were instructed to fake good ($M = 1.83$, $SD = 1.14$) than when instructed to respond honestly ($M = 2.90$, $SD = 1.43$), $F(1, 20) = 16.54$, $p = .001$, $\eta_p^2 = .45$, and when instructed to fake bad ($M = 4.33$, $SD = 0.86$), $F(1, 20) = 111.70$, $p < .001$, $\eta_p^2 = .85$. They also reported consuming less portions of alcoholic drinks when they were instructed to respond honestly than when instructed to fake bad, $F(1, 20) = 44.78$, $p < .001$, $\eta_p^2 = .69$. Thus, the responses given in each condition indicated different amounts of consumption of alcoholic beverages and differences in the reported frequency of consumption.

Discussion 1

The results of this study match previous findings on response distortion claiming that people are successful at altering their responses according to the instruction. For example, Viswesvaran and Ones (1999) showed in a meta-analysis that people are very successful at altering their responses in studies on personality questionnaires employing a directed-faking design. In the health context, the results also align with the respective body of literature. For example, Fekken et al. (2012) showed that self-reported health

behavior on the Health Behavior Checklist (HBC, Vickers, Conway, & Hervig, 1990) was susceptible to response distortion. The results of Fekken et al. (2012) indicated that all dimensions of the HBC were susceptible to reporting an unfavorable behavior, but faking good was only shown on the preventive health subscale, whereas faking good was not successfully shown on the subscales for health risk behavior. Contrary to these findings, in the present study successful response distortion in both directions was not only shown on preventive health behavior but also health risk behavior like smoking and alcohol consumption. Thus, generally speaking, the current study complies with the existing literature and extends its findings.

Considering the magnitude of response distortion in the current study, it was shown that the effects for reporting a more favorable health behavior ($.03 \leq \eta^2 \leq .55$) were smaller than the effects for reporting a more unfavorable health behavior ($.03 \leq \eta^2 \leq .63$) on nearly all facets of health behavior. This might be an indicator of an egocentric bias in the perception of one's health behavior. More specifically, the above-average effect presumes that people rate themselves more favorably than comparable others (Alicke, Klotz, Breitenbecher, Yurak, & Vredenburg, 1995; Taylor & Brown, 1988). Following this assumption, people would rate their health behavior as above average, thus leaving little room for improvement. This ceiling effect might explain why the effect sizes for positive response distortion were smaller than their counterparts for negative response distortion. The dissimilarity of the magnitude for the two directions of response distortion has previously been reported in faking personality inventories (Viswesvaran & Ones, 1999).

Yet, comparing the detected range of response distortion to previous findings on faking in personality inventories, the effect sizes of the current study seem to be larger than previous findings. For example, Birkeland, Manson, Kisamore, Brannick, and Smith (2006), as well as Viswesvaran and Ones (1999), claim that the extent of response distortion in personality inventories may be up to one standard deviation. For example, conscientiousness scores, as well as neuroticism scores, could be augmented by nearly an entire standard deviation, scores for extraversion, openness, and agreeableness tended to be augmented by around half a standard deviation (corresponding to $\eta^2 = .20$, resp. $\eta^2 = .06$) (Viswesvaran & Ones, 1999). The effect sizes in the present study, however, ranged from $.03 \leq \eta^2 \leq .63$ and can mostly be interpreted as very large (Cohen, 1988; Funder & Ozer, 2019). Although it is unclear whether personality inventories and health behavior questionnaires can be regarded as comparable, two conclusions can be drawn: first, apparently, health behavior scales, in general, seem to be very susceptible to faking. Second, some facets of health behavior seem to be more prone to response distortion than others. Maybe, the knowledge on some facets of health behavior is larger than on others, rendering faking a more difficult task concerning those facets, where the knowledge is lower (Levashina & Campion, 2006). Possibly, the sample of the study contributed to the magnitude of the faking effects. As the sample had a relatively high level of education, participants might have been particularly good at distorting their responses because their capacity to fake was high. Another explanation for the magnitude of the effect sizes in the present study is related to design characteristics of the within-subjects design. As participants had to contrast multiple scenarios, they might have been more sensitized to distort their responses more drastically following the instructions, leading to increases in effect sizes (Charness et al., 2012). Similar assumptions were made by Viswesvaran and Ones (1999), who also highlighted the demand characteristics of within-subjects designs.

Method 2

To circumvent the inherent design difficulties of the interpretation of the previous results and to replicate the findings, the research questions of study 1 were also examined using a between-subjects design. The questionnaire was the same in both study designs, solely the set-up was adjusted.

Sample

For study 2, the calculation of the intended sample size was adjusted to the effects detected in the within-subjects design. At least medium effects were anticipated in the between-subjects design (Cohen, 1988). Therefore, the intended sample size calculated by G*Power was 123 participants (Faul et al., 2007). The final sample included 128 German participants (55.5% female) with a mean age of 26 years ($SD = 7.71$, range 18-58). 90% of the participants at least graduated from high school and 36% of those had a university degree. Participants were randomly distributed to one of three groups. Due to selective drop-out, there were 38 participants in the fake good group, 49 participants in the honest group, and 41 participants in the fake bad group. No significant differences were found between the three groups concerning age, gender, or education ($p > .2$).

Design

The same questionnaires as in study 1 (see METHOD 1 section) were employed with the exception that each participant was randomly assigned to one of the three instructions. Thus, study 2 was based on a three-group between-subjects design. The online questionnaire was again implemented using SoSci Survey (Leiner, 2019) and made available to participants at www.soscisurvey.de. The conduct of study 2 complied with the ethical standards of the responsible committee (The Ethics Committee of the Faculty of Empirical Human and Economic Sciences of Saarland University). Written informed consent was obtained from all subjects before the study.

Analytic strategies

The descriptive values were calculated analogously to the procedure in study 1. For diet and physical activity, two MANOVAs were conducted to detect significant differences overall concerning the two constructs. Individual comparisons on each facet of the constructs as well as planned contrasts were used to specify the results. An ANOVA was applied to analyze group differences in the reported frequency of smoking. As the base rate of participants reporting any smoking was very small ($n = 31$), differences in the reported number of cigarettes per day were not analyzed. Similarly, for alcohol consumption, the frequency and quantity of alcohol consumption were compared across the three groups through two ANOVAs.

Results 2

To examine differences in the reported dietary habits, a MANOVA was used. The three groups differed significantly in their reported diet, *Wilks Lambda* = .59, $F(14, 238) = 5.22$, $p < .001$, $\eta_p^2 = .24$. Furthermore, most facets showed similar differences. For the reported amount of vegetables, $F(2, 122) = 22.29$, $p < .001$, $\eta_p^2 = .26$, fruit, $F(2, 122) = 14.12$, $p < .001$, $\eta_p^2 = .18$, grains, $F(2, 122) = 3.98$, $p = .021$, $\eta_p^2 = .07$, meat, $F(2, 122) = 20.62$, $p < .001$, $\eta_p^2 = .25$, eggs, $F(2, 122) = 3.54$, $p = .032$, $\eta_p^2 = .05$, and drinking, $F(2, 122) = 8.03$, $p < .001$, $\eta_p^2 = .12$, the three groups differed significantly, except for dairy products, $F(2, 122) = .20$, $p = .819$, $\eta_p^2 = .00$, and fish, $F(2, 122) = 0.74$, $p = .477$, $\eta_p^2 = .02$, where the reported portions did not differ significantly between the three groups. The descriptive measures and the group contrasts are displayed in [Table 2](#). Across all variables, the fake good group and the honest group differed significantly from the fake bad group, and the fake bad group differed significantly from the honest group. Yet, the fake good group did not differ significantly from the honest group, thus the hypotheses were only partly confirmed (H1). For example, the fake good group and the honest group reported eating significantly more fruit and vegetables than the fake bad group, and the fake bad group reported eating significantly less fruit and vegetables per day than the honest group. But the honest group did not differ significantly from the fake good group in the reported amount of fruit and vegetables eaten.

A MANOVA showed that there were significant differences between the three groups concerning their reported levels of physical activity, *Wilks Lambda* = .64, $F(10, 242) = 6.15$, $p < .001$, $\eta_p^2 = .20$. With the exception of moderate physical activity, $F(2, 125) = 1.30$, $p = .138$, $\eta_p^2 = .02$, all variables showed the hypothesized effects. The reported level of physical activity differed concerning vigorous physical activity, $F(2, 125) = 16.11$, $p < .001$, $\eta_p^2 = .21$, walking, $F(2, 125) = 5.61$, $p = .022$, $\eta_p^2 = .08$, and time spent sitting, $F(2, 125) = 6.56$, $p = .002$, $\eta_p^2 = .10$. Again, the descriptive values and the contrasts between the groups are displayed in [Table 2](#). Consistent with our hypotheses, the honest group differed from the fake bad group and the fake good group differed from the fake bad group (H2). Thus, the fake bad group reported less time spent on physical activity and more time spent sitting than the honest group and the fake good group. Yet, contrary to the hypotheses, the fake good group did mostly not differ from the honest group, except for the reported time spent sitting. Thus, the fake good group reported spending less time sitting but did not report spending significantly more time with physical activity than the honest group.

Concerning the reported smoking behavior, an ANOVA was conducted. Significant differences between the three groups were found, $F(2, 125) = 21.51$, $p < .001$, $\eta_p^2 = .26$. Contrast tests were consistent with the previous findings: the fake bad group ($M = 2.31$, $SD = 0.93$) reported unhealthier smoking behavior than the honest group ($M = 1.35$, $SD = 0.69$), $t(125) = -5.86$, $p < .001$, $\eta_p^2 = .26$, and the fake good group ($M = 1.34$, $SD = 0.71$), $t(125) = -5.53$, $p < .001$, $\eta_p^2 = .26$. Yet, the fake good group again did not differ significantly from the honest group, $t(125) = -.03$, $p = .977$, $\eta_p^2 < .01$ (H3).

An ANOVA was also conducted to investigate whether the three groups differed in their reported frequency of alcohol consumption. In support of our hypotheses, the three groups differed significantly in their reported frequency of alcohol consumption, $F(2, 125) = 5.65$, $p = .004$, $\eta_p^2 = .08$. Participants of the fake good group reported a

Table 2. Means, standard deviations, and planned contrasts statistics for reported diet and physical activity in the between-subjects design.

Variable	Fake good (n = 38)		Honest (n = 49)		Fake bad (n = 41)		Fake good vs. Honest		Fake good vs. Fake bad		Fake bad vs. Honest	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t</i>	η_p^2	<i>t</i>	η_p^2	<i>t</i>	η_p^2
Diet ^f												
Vegetables ^a	2.10	0.97	1.71	1.13	0.77	0.45	1.93	.03	6.40***	.44	4.87***	.22
Fruit ^a	1.81	0.93	1.53	0.90	0.83	0.67	1.47	.02	5.07***	.27	3.92***	.16
Grains ^a	1.69	0.73	2.11	1.24	2.70	2.29	-1.20	.04	-2.79**	.08	-1.75	.03
Dairy Products ^a	1.46	1.56	1.80	3.63	1.75	1.60	-0.60	<.01	-0.49	<.01	0.09	<.01
Meat ^b	3.68	3.90	6.33	7.04	14.17	9.32	-1.46	.09	-6.02***	.35	-4.96***	.19
Fish ^b	2.04	2.53	1.22	2.11	1.69	2.81	-1.01	.03	0.02	<.01	1.07	.01
Eggs ^a	0.62	0.41	0.81	1.06	1.11	0.81	-1.04	.01	-2.62**	.13	-1.73	.02
Drinking ^a	2.66	0.82	2.84	1.24	2.03	0.72	-0.85	<.01	2.80**	.14	3.89***	.13
Physical Activity ^g												
Vigorous ^c	259.82	159.49	243.04	237.82	62.66	66.27	38.54***	<.01	5.00***	.40	-4.87***	.20
Moderate ^c	311.84	504.66	248.59	376.28	171.56	254.22	39.79***	<.01	1.61	.03	0.94	.01
Walking ^c	548.32	740.67	394.76	525.65	153.29	199.93	17.55***	.01	3.31***	.12	2.15*	.08
Sitting ^d	6.06	2.76	7.61	3.23	8.33	2.36	2.53**	.06	-3.55***	.17	-1.19	.02

Note. ^a portions per day. ^b portions per week. ^c minutes per week. ^d hours per day. ^f *df* = 122. ^g *df* = 125

p* < .05 *p* < .01 ****p* < .001.

healthier level of alcohol consumption ($M = 2.66$, $SD = .58$) than participants of the honest group ($M = 2.96$, $SD = .46$), $t(125) = -2.69$, $p = .008$, $\eta_p^2 = .08$, and differed significantly from participants of the fake bad group ($M = 3.02$, $SD = .52$), $t(125) = -3.15$, $p = .002$, $\eta_p^2 = .10$. The reported frequency of alcohol consumption did not differ significantly between the honest group and the fake bad group, $t(125) = .59$, $p = .553$, $\eta_p^2 = .01$ (H4), however. Thus, participants instructed to fake good reported consuming alcoholic beverages less frequent than participants instructed to report their consumption of alcoholic beverages honestly, but only the fake good group differed significantly from the fake bad group.

Concerning the amount of alcoholic drinks, the three groups also differed significantly, $F(2, 107) = 9.60$, $p < .001$, $\eta_p^2 = .15$. Participants of the fake good group ($M = 1.11$, $SD = 0.53$) reported less portions of alcoholic drinks than participants in the honest group ($M = 1.50$, $SD = 0.69$), $t(107) = -2.36$, $p = .020$, $\eta_p^2 = .09$. The fake good group differed as expected from the fake bad group ($M = 1.86$, $SD = 0.76$), $t(107) = -4.37$, $p < .001$, $\eta_p^2 = .26$. This also was the case for the honest group, $t(107) = -2.37$, $p = .02$, $\eta_p^2 = .08$. Thus, the three groups differed both in the reported frequency as well as the amount of consumption of alcoholic drinks.

Discussion 2

Study 2 also supports the assumption that participants are able to distort their responses in a health behavior self-report questionnaire when instructed to do so. On all four dimensions of health behavior, thus both concerning preventive behavior and risk behavior, significant differences were found under the different experimental instructions.

The effect sizes of study 2 comply with previous findings. Again, for most dimensions of health behavior, large effects for faking were found ($.08 \leq \eta_p^2 \leq .25$). Yet, although participants were generally able to distort their responses, the reported health behavior of the fake good group did not differ significantly from the one of the honest group concerning at least some of the health behavior dimensions. However, both the fake good group and the honest group differed significantly from the fake bad group except for very few facets.

The missing differences between the fake good group and the honest group might be a result of several processes. For example, participants in the honest group might have practiced response distortion, too. According to Mazar and Ariely (2006; Mazar, Amir, & Ariely, 2008), it is possible to behave dishonestly to a certain extent without challenging the self-concept of being an honest person. The Theory of Self-Concept Maintenance assumes that people can solve their motivational dilemma of profiting from dishonest behavior versus risking the extrinsic and intrinsic costs of dishonest behavior by balancing both elements. The theory claims that there is a range of dishonesty within which people can behave dishonestly enough to profit from it but do not endanger their positive self-view (Mazar, Amir, & Ariely, 2008). In the present study, this mechanism would allow for participants to report their health behavior slightly ameliorated without getting into conflict with the experimental instruction to report their health behavior honestly.

The missing differences between the two experimental groups might also be explained by an egocentric bias of all participants. Again, the above-average effect (Alicke et al., 1995; Taylor & Brown, 1988) might have led participants to believe that their health behavior is healthier than the health behavior of other people. Thus, when asked to

fake good and report a particularly positive health behavior, participants of this experimental group might have adjusted their reports insufficiently since they perceived their behavior to already be healthy above average. Correspondingly, they would adjust their responses very much in the fake bad group to report a health behavior that is even more unfavorable than that of most people.

A third explanation for the similarity of the reported health behavior of the fake good group and the honest group arises from design characteristics. A possible, although improbable inherent design flaw might be that the three groups differed from one another regarding their real health behavior. However, as participants were assigned randomly to the three groups, there should not have been significant differences in health behavior between the groups. As Rost (2013) state, all confounding variables should be present in all experimental groups equally if participants are assigned randomly to the groups and the experimental groups are sufficiently large, thus minimizing the probability of systematic a priori group differences. Yet, significant differences in the real health behavior of the participants of the three experimental groups cannot entirely be ruled out.

Comparison of main results of the two studies

To investigate first indicators of design inherent differences, the honest responses given in study 1 were compared to the honest responses in study 2. MANOVAs and ANOVAs indicate that the honest responses from the within-subjects design were significantly healthier than honest responses of the between-subjects design for diet, Wilks Lambda = .60, $F(8, 173) = 14.15$, $p < .001$, $\eta_p^2 = .40$, physical activity, Wilks Lambda = .94, $F(4, 178) = 2.96$, $p = .021$, $\eta_p^2 = .06$, smoking, $F(1, 181) = 23.95$, $p < .001$, $\eta_p^2 = .12$, and alcohol consumption, $F(1, 181) = 9.09$, $p = .003$, $\eta_p^2 = .05$.

To secure the internal validity of this comparison, potential differences in the demographic characteristics of the two samples were investigated. The mean age was comparable in the two samples, $t(249,284) = -.57$, $p = .568$, $\eta_p^2 < .01$, the mean age in the within-subjects design ($M = 25.49$, $SD = 11.39$) did not differ significantly from the mean age in the between-subjects design ($M = 26.16$, $SD = 7.71$). However, the gender ratio differed significantly between the two samples, $\chi^2(1) = 10.45$, $p < .001$. In the within-sample, there were significantly more female participants (74.3%) than in the between-sample (55.5%). Moreover, the education level differed slightly between the two samples. Participants of the between-subjects design tended to have a higher education ($M = 5.97$, $SD = 1.05$), than participants of the within-subjects design ($M = 5.40$, $SD = 0.91$), $t(268) = -4.76$, $p < .001$, $\eta_p^2 = .08$.

General discussion

The current studies aimed at exploring people's ability to fake self-reported health behavior both concerning preventive health behavior and health risk behavior. Both the conventional within-subjects design and the between-subjects design yielded evidence for people's ability to practice response distortion and fake reports of their health behavior. The effect sizes of faking in self-report measures of health behavior indicate that the phenomenon should not be underestimated. The results thus comply with and expand previous findings.

Design-related differences

Comparing the results of both studies, the patterns of the differences in both studies between the instructions to fake good, be honest, or fake bad look similar. However, there are repeatedly larger and more pronounced differences between the instructions in the within-subjects design. The observation that the responses to the different instructions are more comparable and closer together in the between-subjects design is also backed by the insignificant differences between the fake good group and the honest group on most facets. The more distinct response pattern in the within-subjects design might be an indicator for a participant \times treatment interaction (Viswesvaran & Ones, 1999). It seems plausible that there might be interindividual differences in faking. That is, participants with broader knowledge or a higher need for approval from others might fake more than others (Levashina & Campion, 2006; Rzewnicki, Auweele, & De Bourdeaudhuij, 2003). Alternatively, the more pronounced effects might be a design characteristic. For example, demand-effects might have caused participants to answer rather extremely in the two faking instructions, as participants tend to alter their responses as a result of direct contrasts of the conditions (Charness et al., 2012).

The more pronounced effects in the within-subjects design might also be a result of comparison effects. The study design might have caused participants of the within-subjects design to alter their responses carefully according to the instructions and previously given answers. Participants of the within-subjects design might not have reported their actual health behavior but somehow calculated figures in the honest condition. A first indicator of such referencing is that participants of the honest condition of the within-subjects design reported significantly healthier behavior than participants of the honest group of the between-subjects design. Yet, as stated above, the samples of the two studies differed concerning some characteristics. In study 1, there were significantly more female participants than in study 2. As women usually have a better health behavior than men, this also might explain why participants of the within-subjects design showed healthier honest reports than those of the between-subjects design (Dehghan, Akhtar-Danesh, & Merchant, 2011; Wardle et al., 2004). In addition, the education level differed slightly between the two samples. Participants of the between-subjects design tended to have a higher education than participants of the within-subjects design. Higher education is usually positively correlated with better health behavior (Cowell, 2006). Similarly, higher education is positively correlated with the ability to fake (Levashina & Campion, 2006). Thus, it seems rather surprising that participants of the within-subjects design reported healthier behavior in the honest condition. Nevertheless, most participants of both studies had at least graduated from high school and it might be possible that the variance of educational levels was quite restricted to highly educated participants in the present studies. The effects of those potential differences concerning the interpretation of the different findings between the two studies are not clear. The most important comparison between the two samples would doubtlessly be their self-reported health behavior in a context that does not involve any faking instruction. This would probably call for an innovative research design that future studies might employ. Ideally, this design would not only allow excluding the possibility of a priori group differences in the health behavior of the participants of the two study designs but would also shed light on the processes of response distortion in the respective design characteristics.

Implications for research and practice

Although a few issues remain to be resolved, the current research contributes to the body of knowledge with important insights. For future research, it seems crucial to beware of design-related differences in studies concerning response distortion. Whereas quite a few previous studies concluded that the within-subjects design might be better suited to investigate faking, it seems important to bring into consideration the ecological validity of both designs. As noted previously, the evoked mental processes might differ in within- and between-subjects designs (Hsee & Zhang, 2004). Thus, it would be important to access the cognitive processes that prevail when faking is done. Therefore, future studies should include qualitative methods like the Thinking Aloud Method, in which participants verbalize their thoughts while faking a health behavior questionnaire to get closer to which design seems to be more naturally aligned with the ongoing mental processes in practice (Eccles & Arsal, 2017).

Moreover, it is important to emphasize the meaning of the current research for previous and future research as well as for practitioners. Data on health behavior based on self-report measures have to be interpreted cautiously, as there is a very real possibility that the reports have suffered from faking. For example, nationwide assessments of dietary habits or physical activity are often realized based on phone-based interviews. It seems plausible that these results are prone to faking. In their large-scale study on assessing the level of physical activity worldwide, Guthold et al. (2018) acknowledge the possibility of faked responses in self-reports. They attempted to correct for it by applying a correcting factor which resulted of a comparison of the results of the IPAQ to another self-report questionnaire, the Global Physical Activity Questionnaire (Armstrong & Bull, 2006). Assuming that people have a desire to create a favorable image of themselves, it is quite conceivable that participants would have augmented their reported levels of physical activity in both questionnaires. If that was the case, the real extent of insufficient physical activity would still be underestimated. The current research leads to the assumption that this underestimation might be a huge problem, as faking self-reported levels of physical activity was shown to be easily executed. The occurrence of faking in self-report questionnaires might also lead to faulty interventions, either because a need for an intervention is not recognized or because interventions are implemented that are not optimally tailored to the need.

A useful strategy would be to counteract faking behavior in the first place. Therefore, the willingness to fake should be targeted. As Levashina and Campion (2006) claim, the willingness to modify the information given to create the desired impression relies largely on the potential benefits of the desired impression. The prevention of negative feelings as shame and guilt, as well as fear of judgment, have been identified as major reasons for dishonesty in practice (Levy et al., 2018). Thus, researchers and practitioners should attempt to create an environment of trust and acceptance to minimize the initial willingness to fake. Also, highlighting the benefits of honest responses might decrease the willingness to fake (Law, Bourdage, & O'Neill, 2016).

Researchers and practitioners would probably also benefit from a critical pluralism of methods when assessing subjective constructs like mental and physical health status and behavior. For example, Rzewnicki et al. (2003) previously showed that interviewing

participants that previously filled out a written self-report questionnaire on their physical activity level led participants to correct their reports. Thus, it might be a simple solution to compare different self-report measures. As noted earlier, this method was applied in research, for example by Guthold et al. (2018). Yet, assuming that all self-report measures might be prone to faking, the additional use of more objective measurement methods might be plausible, for example, the use of wearable activity trackers to assess physical activity (Wong, Mentis, & Kuber, 2018). Although these indicators have their faults and weaknesses, in combination with self-report measures, the resulting findings might be less prone to errors due to faking and response distortion.

Limitations

An important limitation is that the current studies investigated solely whether successful faking of self-reported health behavior is theoretically possible. As customary in directed-faking studies, we specifically instructed participants to distort their responses. Levashina and Campion (2006) claim that the capacity, the willingness, and the opportunity to control the information given are essential for faking. By controlling the willingness and the opportunity to fake, we simplified the faking process artificially and solely investigated people's capacity to fake their responses. The results do not answer the important question of whether in practice, people tend to fake their self-reported health behavior. Yet, the lack of differences between the instructions to be honest and to fake good in the between-subjects design might be indicative of participants practicing response distortion unsolicited, as there were clear differences in the within-subjects design when a comparison between the responses to the different instructions was possible. As previous research indicates a high probability for the presence of such faking in practice (DePaulo et al., 1996; Levy et al., 2018) and the current studies hint at faking being a substantial threat to the assessment of health behavior, future studies should target the presence of faking in self-reports concerning health behavior in practice.

Conclusion

This research yields evidence for people's ability to practice response distortion and fake their reported health behavior both concerning preventive health behavior and health risk behavior. As faking is linked with important considerations about an adequate study design, a major benefit of this research is the analysis of the research question on behalf of two research designs to investigate the robustness of the results and profit from the scientific insights of each design. The findings of the two studies call for caution when interpreting health behavior data based on self-report measures. It is highly recommended to consider faking in future research to clarify its impact on the interpretation of self-reported health behavior in research and practice.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported there is no funding associated with the work featured in this article.

ORCID

Viktoria S. Egele  <http://orcid.org/0000-0003-4354-6751>

References

- Alicke, M. D., Klotz, M. L., Breitenbecher, D. L., Yurak, T. J., & Vredenburg, D. S. (1995). Personal contact, individuation, and the better-than-average effect. *Journal of Personality and Social Psychology*, 68(5), 804–825. doi:10.1037/0022-3514.68.5.804
- Armstrong, T., & Bull, F. (2006). Development of the world health organization global physical activity questionnaire (GPAQ). *Journal of Public Health*, 14(2), 66–70. doi:10.1007/s10389-006-0024-x
- Atienza, A., & King, A. (2005). Comparing self-reported versus objectively measured physical activity behavior: A preliminary investigation of older Filipino American women. *Research Quarterly for Exercise and Sport*, 76(3), 358–362.
- Becker, G. S. (1968). Crime and punishment: An economic approach. In N. G. Fielding, A. Clarke, & R. Witt (Eds.), *The economic dimensions of crime* (pp. 13–68). London: Palgrave Macmillan UK.
- 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(4), 317–335. doi:10.1111/j.1468-2389.2006.00354.x
- Booth, M. L., Owen, N., Bauman, A. E., & Gore, C. J. (1996). Retest reliability of recall measures of leisure-time physical activity in Australian adults. *International Journal of Epidemiology*, 25(1), 153–159. doi:10.1093/ije/25.1.153
- Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, 81(1), 1–8. doi:10.1016/j.jebo.2011.08.009
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cook, M. (2004). *Personnel selection: Adding value through people*. New York, NY: Wiley.
- Cowell, A. J. (2006). The relationship between education and health behavior: Some empirical evidence. *Health Economics*, 15(2), 125–146. doi:10.1002/hec.1019
- Craig, C. L., Marshall, A. L., Sjöström, M., Bauman, A. E., Booth, M. L., Ainsworth, B. E., ... Oja, P. (2003). International physical activity questionnaire: 12-country reliability and validity. *Medicine & Science in Sports & Exercise*, 35(8), 1381–1395. doi:10.1249/01.mss.0000078924.61453.fb
- Crowne, D., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24(4), 349–354. doi:10.1037/h0047358.
- Dehghan, M., Akhtar-Danesh, N., & Merchant, A. T. (2011). Factors associated with fruit and vegetable consumption among adults. *Journal of Human Nutrition and Dietetics*, 24(2), 128–134. doi:10.1111/j.1365-277X.2010.01142.x
- DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology*, 70(5), 979–995. doi:10.1037/0022-3514.70.5.979
- Deutsche Gesellschaft für Ernährung. (2010). 10 Regeln der DGE. Retrieved May 3, 2020, from <https://www.dge.de/ernaehrungspraxis/vollwertige-ernaehrung/10-regeln-der-dge/>
- Eccles, D. W., & Aarsal, G. (2017). The think aloud method: What is it and how do I use it? *Qualitative Research in Sport, Exercise and Health*, 9(4), 514–531. doi:10.1080/2159676x.2017.1331501

- Edwards, A. (1957). *The social desirability variable in personality assessment and research*. New York: Dryden Press.
- Emanuel, A. S., McCully, S. N., Gallagher, K. M., & Updegraff, J. A. (2012). Theory of planned behavior explains gender difference in fruit and vegetable consumption. *Appetite*, 59(3), 693–697. doi:10.1016/j.appet.2012.08.007
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. doi:10.3758/BF03193146
- Fekken, C., Holden, R., McNeill, B., & Wong, J. (2012, July 22-27). *Faking self reported health behaviour* [Poster presentation]. International Congress of Psychology, Cape Town, South Africa.
- Fell, C. B., & König, C. J. (2016). Cross-cultural differences in applicant faking on personality tests: A 43-nation study. *Applied Psychology: An International Review*, 65(4), 671-717. doi:10.1111/apps.12078
- Foa, E. B., Cashman, L., Jaycox, L., & Perry, K. (1997). The validation of a self-report measure of posttraumatic stress disorder: The posttraumatic diagnostic scale. *Psychological Assessment*, 9(4), 445–451. doi:10.1037/1040-3590.9.4.445
- Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, 2(2), 156–168. doi:10.1177/2515245919847202
- Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and Individual Differences*, 7(3), 385–400. doi:10.1016/0191-8869(86)90014-0
- Furnham, A. (1990). Faking personality questionnaires: Fabricating different profiles for different purposes. *Current Psychology*, 9(1), 46–55. doi:10.1007/BF02686767
- Gochman, D. S. (Ed.). (1997). *Handbook of health behavior research I: Personal and social determinants*. New York, NY: Plenum Press.
- Gorber, S., Schofield-Hurwitz, S., Hardt, J., Levasseur, G., & Tremblay, M. (2009). The accuracy of self-reported smoking: A systematic review of the relationship between self-reported and cotinine-assessed smoking status. *Nicotine & Tobacco Research*, 11(1), 12–24. doi:10.1093/ntr/ntn010
- Griffith, R. L., Chmielowski, T., & Yoshita, Y. (2007). Do applicants fake? An examination of the frequency of applicant faking behavior. *Personnel Review*, 36(3), 341–355. doi:10.1108/00483480710731310
- Guthold, R., Stevens, G. A., Riley, L. M., & Bull, F. C. (2018). Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1.9 million participants. *The Lancet Global Health*, 6(10), e1077–e1086. [https://doi.org/10.1016/S2214-109X\(18\)30357-7](https://doi.org/10.1016/S2214-109X(18)30357-7)
- Hagströmer, M., Oja, P., & Sjörström, M. (2006). The International physical activity questionnaire (IPAQ): A study of concurrent and construct validity. *Public Health Nutrition*, 9(6), 755–762. doi:10.1079/phn2005898
- Helmerhorst, H., Brage, S., Warren, J., Besson, H., & Ekelund, U. (2012). A systematic review of reliability and objective criterion-related validity of physical activity questionnaires. *International Journal of Behavioral Nutrition and Physical Activity*, 9(1), 1–55. doi:10.1186/1479-5868-9-103
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., ... McElreath, R. (2001). In search of homo economicus: Behavioral experiments in 15 small-scale societies. *The American Economic Review*, 91(2), 73–78. doi:10.1257/aer.91.2.73.
- Hsee, C. K., & Zhang, J. (2004). Distinction bias: Misprediction and mischoice due to joint evaluation. *Journal of Personality and Social Psychology*, 86(5), 680–695. doi:10.1037/0022-3514.86.5.680
- Hwang, J., Kim, J., Lee, D., Jung, H., & Park, S. (2018). Underestimation of self-reported smoking prevalence in Korean adolescents: Evidence from gold standard by combined method. *International Journal of Environmental Research and Public Health*, 15(4), 689. doi:10.3390/ijerph15040689

- Kasl, S. V., & Cobb, S. (1966). Health behavior, illness behavior and sick role behavior. *Archives of Environmental Health: An International Journal*, 12(2), 246–266. doi:10.1080/00039896.1966.10664365
- Komar, S., Brown, D. J., Komar, J. A., & Robie, C. (2008). Faking and the validity of conscientiousness: A monte carlo investigation. *Journal of Applied Psychology*, 93(1), 140–154. doi:10.1037/0021-9010.93.1.140.
- Law, S. J., Bourdage, J., & O'Neill, T. A. (2016). To fake or not to fake: Antecedents to interview faking, warning instructions, and its impact on applicant reactions. *Frontiers in Psychology*, 7, 1–13. <https://doi.org/10.3389/fpsyg.2016.01771>
- Lechner, L., Brug, J., & De Vries, H. (1997). Misconceptions of fruit and vegetable consumption: Differences between objective and subjective estimation of intake. *Journal of Nutrition Education*, 29(6), 313–320. doi:10.1016/S0022-3182(97)70245-0
- Leiner, D. J. (2019). *SoSci Survey* (Version 3.1.06) [Computer software]. <https://www.sosicurvey.de>
- Levashina, J., & Campion, M. A. (2006). A model of faking likelihood in the employment interview. *International Journal of Selection and Assessment*, 14(4), 299–316. doi:10.1111/j.1468-2389.2006.00353.x
- Levy, A. G., Scherer, A. M., Zikmund-Fisher, B. J., Larkin, K., Barnes, G. D., & Fagerlin, A. (2018). Prevalence of and factors associated with patient nondisclosure of medically relevant information to clinicians. *JAMA Network Open*, 1(7), e185293–e185293. doi:10.1001/jamanetworkopen.2018.5293.
- Lewis, S., Cherry, N., Mcl Niven, R., Barber, P., Wilde, K., & Povey, A. (2003). Cotinine levels and self-reported smoking status in patients attending a bronchoscopy clinic. *Biomarkers*, 8(3-4), 218–228. doi:10.1080/1354750031000120125
- Littlefield, A., Brown, J., DiClemente, R., Safonova, P., Sales, J. M., Rose, E. S., ... Rassokhin, V. (2017). Phosphatidylethanol (PEth) as a biomarker of alcohol consumption in HIV-infected young Russian women: Comparison to self-report assessments of alcohol use. *AIDS and Behavior*, 21(7), 1938–1949. doi:10.1007/s10461-017-1769-7
- Locander, W., Sudman, S., & Bradburn, N. (1976). An investigation of interview method, threat and response distortion. *Journal of the American Statistical Association*, 71(354), 269–275. doi:10.1080/01621459.1976.10480332
- Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, 45(6), 633–644. doi:10.1509/jmkr.45.6.633
- Mazar, N., & Ariely, D. (2006). Dishonesty in everyday life and its policy implications. *Journal of Public Policy and Marketing*, 25(1), 117–126. doi:10.1509/jppm.25.1.117
- McCrae, R. R., & Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology*, 51(6), 882–888. doi:10.1037/0022-006X.51.6.882.
- McFarland, L., & Ryan, A. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology*, 85(5), 812–821. doi:10.1037/0021-9010.85.5.812.
- Mensch, B., & Kandel, D. (1988). Underreporting of substance use in a national longitudinal youth cohort: Individual and interviewer effects. *Public Opinion Quarterly*, 52(1), 100–124.
- Morgan, M. S. (2006). Economic man as model man: Ideal types, idealization and caricatures. *Journal of the History of Economic Thought*, 28, 1–27. doi:10.1080/10427710500509763
- Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology*, 15(3), 263–280. doi:10.1002/ejsp.2420150303
- Nelson, M., Taylor, K., & Vella, C. (2019). Comparison of self-reported and objectively measured sedentary behavior and physical activity in undergraduate students. *Measurement in Physical Education and Exercise Science*, 23(3), 237–248. doi:10.1080/1091367X.2019.1610765
- Norman, W. T. (1967). On estimating psychological relationships: Social desirability and self-report. *Psychological Bulletin*, 67(4), 273–293. doi:10.1037/h0024414
- Paulhus, D. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46(3), 598–609. doi:10.1037/0022-3514.46.3.598.
- Rost, J. (2013). *Interpretation und bewertung pädagogisch-psychologischer studien [interpretation and evaluation of pedagogical-psychological studies]* (Vol. 3). Weinheim: Beltz.

- Rzewnicki, R., Auweele, Y. V., & De Bourdeaudhuij, I. (2003). Addressing overreporting on the International Physical Activity Questionnaire (IPAQ) telephone survey with a population sample. *Public Health Nutrition*, 6(3), 299–305. doi:10.1079/PHN2002427
- Sattler, M. C., Ainsworth, B. E., Andersen, L. B., Foster, C., Hagströmer, M., Jaunig, J., ... Poppel, M. N. M. v. (2021). Physical activity self-reports: Past or future? *British Journal of Sports Medicine*, 55(16), 889–891.
- Straßburg, A., Eisinger-Watzl, M., Krems, C., Roth, A., & Hoffmann, I. (2019). Comparison of food consumption and nutrient intake assessed with three dietary assessment methods: Results of the German National Nutrition Survey II. *European Journal of Nutrition*, 58(1), 193–210. doi:10.1007/s00394-017-1583-z
- Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, 103(2), 193–210. doi:10.1037/0033-2909.103.2.193
- Van Hoof, E. A., & Born, M. P. (2012). Intentional response distortion on personality tests: Using eye-tracking to understand response processes when faking. *Journal of Applied Psychology*, 97(2), 301–316. doi:10.1037/a0025711
- Vickers, R. R., Conway, T. L., & Hervig, L. K. (1990). Demonstration of replicable dimensions of health behaviors. *Preventive Medicine*, 19(4), 377–401. doi:10.1016/0091-7435(90)90037-k
- Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement*, 59(2), 197–210. doi:10.1177/00131649921969802
- Wardle, J., Haase, A. M., Steptoe, A., Nillapun, M., Jonwutiwes, K., & Bellis, F. (2004). Gender differences in food choice: The contribution of health beliefs and dieting. *Annals of Behavioral Medicine*, 27(2), 107–116.
- Wong, C. K., Mentis, H. M., & Kuber, R. (2018). The bit doesn't fit: Evaluation of a commercial activity-tracker at slower walking speeds. *Gait & Posture*, 59, 177–181. doi:10.1016/j.gaitpost.2017.10.010
- World Health Organization. (2013). *Global action plan for the prevention and control of non-communicable diseases 2013-20*. Retrieved from <https://www.who.int/publications/i/item/9789241506236>