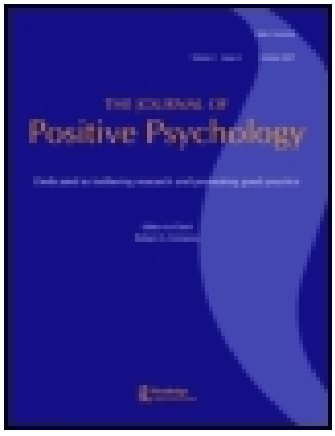

The Journal of Positive Psychology: Dedicated to furthering research and promoting good practice
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/rpos20

Revisiting desirable response bias in well-being reports

Samantha J. Heintzelman[a], Jason Trent[b] & Laura A. King[a]
[a] Department of Psychological Sciences, University of Missouri, 210 McAlester Hall, Columbia, MO 65201, USA
[b] Department of Psychology, Hood College, Frederick, MD, USA
Published online: 23 Jun 2014.

To cite this article: Samantha J. Heintzelman, Jason Trent & Laura A. King (2014): Revisiting desirable response bias in well-being reports, The Journal of Positive Psychology: Dedicated to furthering research and promoting good practice, DOI: 10.1080/17439760.2014.927903

To link to this article: http://dx.doi.org/10.1080/17439760.2014.927903

The Journal of Positive Psychology, 2014
http://dx.doi.org/10.1080/17439760.2014.927903

Revisiting desirable response bias in well-being reports

Samantha J. Heintzelman[a]*, Jason Trent[b] and Laura A. King[a]

[a]*Department of Psychological Sciences, University of Missouri, 210 McAlester Hall, Columbia, MO 65201, USA;* [b]*Department of Psychology, Hood College, Frederick, MD, USA*

(*Received 6 May 2013; accepted 5 May 2014*)

Growing popular interest in positive psychology may have important implications for the measurement of well-being. Five studies tested the prediction that well-being ratings are influenced by desirability bias. In Study 1, participants ($N = 176$) instructed to fake good endorsed higher well-being; those instructed to fake bad endorsed lower well-being, compared to controls. In Studies 2 and 3 (N's = 111, 121), control participants endorsed higher levels of well-being compared to those attached to a bogus pipeline. These differences were mediated by desirability bias. In Study 4 ($N = 417$), instruction manipulations did not affect well-being levels, but presenting a desirability measure prior to well-being measures attenuated the correlations between them. In Study 5 ($N = 391$), however, this order effect did not replicate. We discuss the importance of continued vigilance for desirability bias in well-being research as a ready solution to this clear problem remains elusive.

Keywords: well-being; response bias; self-report measures; bogus pipeline; meaning in life; life satisfaction

Over a dozen years since the advent of positive psychology (Seligman & Csikszentmihalyi, 2000), it may be hard to believe that there was a time when well-being was not considered particularly socially desirable. However, in their initial validation studies, both the Satisfaction with Life Scale (SWLS; Diener, Emmons, Larsen, & Griffin, 1985) and the presence of meaning subscale of the Meaning in Life Questionnaire (MLQ-P; Steger, Frazier, Oishi, & Kaler, 2006) were unrelated to the Marlowe-Crowne Social Desirability Scale (MCSD, Crowne & Marlowe, 1964; r's = 0.02, −0.08 for the SWLS and MLQ-P, respectively). Such results were taken to indicate (justifiably) that the contamination of these scales with social desirability was not a concern and subsequent research using these scales has not routinely addressed the issue. For other components of well-being, positive and negative affect (PA and NA), it is, perhaps, informative that the initial (very thorough) validation of the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988) did not address social desirability at all, suggesting that, at the time, such bias was considered irrelevant to mood reports.

Times have changed. In recent years, positive psychology research has been disseminated widely (Hart & Sasso, 2011; Seligman, Steen, Park, & Peterson, 2005). An ever-growing number of books promote happiness, and well-being is well-represented in popular media. Importantly, positive psychology research has highlighted the benefits of positive psychological states for a host of outcomes (e.g. Lyubomirsky, King, & Diener, 2005; Pressman & Cohen, 2005), emphasizing that well-being is not only a byproduct of positive life experiences, but a *precursor* to these. In turn, these ideas have been applied, energetically, to education, manufacturing, and the military. For instance, aspects of psychological well-being (e.g. optimism, resilience, and hope) have been presented as valuable resources in the work domain, under the revealing label, *Psychological Capital* (*PsyCaps*; Luthans, Youssef, & Avolio, 2007; Schaubroeck, Riolli, Peng, & Spain, 2011). Research shows that *PsyCaps* predict numerous workplace outcomes, including increased employee performance (Luthans, Avolio, Avey, & Norman, 2007) and decreased work absences (Avey, Patera, & West, 2006). Further, policy-makers have taken notice. The United Nations General Assembly passed a resolution emphasizing the enhancement of well-being worldwide (U.N. News Centre, July 19, 2011). Among other things, the resolution called upon nations to use measures of well-being to guide national policies.

Clearly, the science of well-being has infiltrated public awareness successfully. A potential unintended consequence of this success is that well-being has changed from an evaluatively neutral construct to one that is quite socially desirable. If it was once acceptable to report one's psychological well-being as low, we might ask whether it still is: Given the well-publicized value of well-being, is it still okay to report oneself as unhappy? Potentially worrisome evidence shows associations between well-being and social desirability are higher

*Corresponding author. Email: sjhgh2@mail.missouri.edu

than previously reported, at least differing from 0. Soubelet and Salthouse (2011) found significant correlations between the MCSD and the SWLS (0.26), PANAS-PA (0.30), and PANAS-NA (−0.22).

In this article, we present five studies examining the influence of desirability bias on reports of meaning in life, life satisfaction, and mood. Study 1 employed fake-good/fake-bad instructions. Studies 2 and 3 employed bogus pipeline procedures. Studies 4 and 5 explored whether simple strategies (instruction manipulations, presenting a social desirability scale prior to well-being measures) might mitigate desirability bias. Further, in all but Study 2, participants completed the Balanced Inventory of Desirable Responding (BIDR; Paulhus & Reid, 1991) to examine the relationships between desirability bias and well-being measures.

Study 1

Participants completed online measures of life satisfaction, meaning in life, mood, and desirable responding with instructions to fake good, fake bad, or respond honestly. Such manipulations have been used in the application of personality testing to hiring and other workplace decision-making (Viswesvaran & Ones, 1999). Discrepancies between honest and fake instructions indicate the degree to which individuals can alter their scores on a scale if motivated to do so (Viswesvaran & Ones, 1999). This approach allowed us to gage awareness of the desirable direction of responses to these scales. We predicted life satisfaction, meaning in life, PA and the desirability measure to be higher (lower) in the fake good (bad) condition, relative to the honest condition. We predicted NA to show the opposite pattern. We also tested whether, within condition, well-being measures related to desirability bias, even in this anonymous sample.

Method

Participants

One hundred seventy-six Amazon Mechanical Turk (MTurk) workers (106 men, 1 not reporting, and paid $1) completed an online survey. Represented ethnicities included, 82.2% White/European American, 2.3% Black/African American, 5.7% Latino, 8% Asian, with the remaining participants indicating 'other' or not reporting. Ages ranged from 18 to 69 ($M = 28.16$, SD = 9.78).

Procedures and measures

We randomly assigned participants to instructions to fake good by making 'the best impression possible,' fake bad by making 'the worst impression possible,' or respond honestly (the control condition). Then, participants completed the BIDR and well-being measures.[1] All items were rated from 1 to 7.

The 40-item BIDR contains two subscales: self-deceptive enhancement, the tendency to give favorable self-descriptions that one truly believes (e.g. 'It would be hard for me to break any of my bad habits' recoded) and impression management, the tendency to describe oneself favorably to others (e.g. 'I never cover up my mistakes') (Paulhus & Reid, 1991). Scores are computed by counting the number of extreme scores (6's or 7's, after reverse coding appropriate items). Because the two subscales were strongly correlated ($r = 0.70$) and did not differ substantially in their relations with the dependent measures, we report, in all studies, results for the total score, which represents general positive response bias.

Participants completed the 5-item MLQ-P (Steger et al., 2006; sample item: 'I understand my life's meaning'); the 5-item SWLS (Diener et al., 1985; sample item 'I am satisfied with my life'); and rated a list of mood adjectives assessing current PA (happy, pleased, cheerful, and fun/enjoyment) and NA (worried, anxious, frustrated, and nervous; Diener & Emmons, 1984; Diener, Smith, & Fujita, 1995).

Results and discussion

Results for the effects of condition on all measures are presented in Table 1. Planned contrasts (fake bad = −1; control condition = 0; fake good = 1; with fake condition codes flipped for NA) were significant (and large) for all dependent measures. The fake bad and fake good conditions both differed significantly from the control group on all measures. Effect sizes indicate larger deviations from controls for those in the fake bad condition compared to the fake good condition, consistent with previous research showing that most people rate their lives as pretty meaningful (Heintzelman & King, in press) and themselves as quite happy (Diener & Diener, 1996).

Next, we examined the correlations among the variables, focusing on the relationships between the BIDR and well-being. As Table 2 shows, these were uniformly significant. Although there were no significant condition X BIDR interactions, the correlations within the control condition most appropriately reflect typical responding. Among anonymous participants instructed to answer honestly, the BIDR and well-being measures were significantly (and substantially) related.

Study 1 indicates that participants were both able to alter their responses and aware of the desirable response direction: Reporting high in well-being makes a good impression. Now, one might recognize the desirable direction of a response without necessarily engaging in that bias. However, the relationships between the BIDR and well-being in the control group indicate that desirability bias shares significant variance with 'honest'

Table 1. Descriptive statistics and results, Study 1.

	BIDR	MLQ-P	SWLS	PA	NA
Grand mean (SD)	12.19 (9.33)	3.99 (2.03)	3.93 (1.97)	3.67 (2.01)	3.03 (2.23)
Condition effects					
Fake bad (n = 58)	9.50 (6.68)	2.74 (1.88)	2.73 (1.89)	2.41 (1.73)	4.76 (2.45)
Honest (n = 60)	11.38 (7.47)	4.17 (1.68)	4.07 (1.70)	3.80 (1.58)	2.59 (1.61)
Fake good (n = 58)	15.72 (12.0)	5.04 (1.87)	4.97 (1.66)	4.78 (2.00)	1.79 (1.37)
Planned contrasts, all groups					
t(172)	3.72**	6.80**	6.84**	7.16**	8.54**
d	0.57	1.04	1.04	1.09	1.30
Fake bad vs. honest					
t(115)	1.44	4.35**	4.04**	4.57**	5.69**
d	0.27	0.80	0.75	0.84	1.05
Fake good vs. honest					
t(116)	2.37*	2.65**	2.89**	2.94**	2.89**
d	0.43	0.49	0.54	0.54	0.54

Notes: N = 176; BIDR = Balanced Inventory of Desirable Responding; MLQ-P = Meaning in Life Questionnaire presence subscale; SWLS = Satisfaction with Life Scale; PA = Positive Affect; NA = Negative Affect.
*p = 0.02; **p < 0.01.

well-being reports. To further examine whether typical well-being reports are inflated by social desirability, Studies 2 and 3 employed bogus pipeline procedures.

The bogus pipeline

The bogus pipeline (Jones & Sigall, 1971) involves connecting participants to a device that ostensibly detects deception. Once convinced of the authenticity of the device, participants' responses to questions are compared to those given by a control group. The procedure is based on the assumption that if they believe deception can be detected, participants will provide more honest responses than they might otherwise.

When reporting on socially undesirable characteristics, the bogus pipeline leads to greater accuracy and more endorsement of undesirable attitudes and behaviors (Roese & Jamieson, 1993). It also *decreases* reports of desirable characteristics (Mummendey & Bolten, 1981). If well-being reports are inflated by desirability bias, they should be higher in the control condition compared to the bogus pipeline condition.

Study 2

The influence of a bogus pipeline on well-being reports has never been addressed. Study 2 represents an exploratory attempt to do so. Participants were randomly assigned to complete well-being measures while connected to a purported lie detector (or not). Participants had dummy sensors attached to their faces. Those in the bogus pipeline condition were led to believe the sensors detected deception and the validity of this claim was demonstrated using a rigged guilty knowledge test. During the manipulation phase, all participants completed well-being measures. We predicted that the bogus pipeline would lead to lower well-being reports compared to controls.

In addition, participants completed measures of PA, NA, and need satisfaction prior to the manipulation. These measures served as covariates and were used to examine whether the manipulation moderated their relevance to well-being.

Method

Participants

One hundred eleven undergraduates (81 women) completed this study in partial fulfillment of a research participation requirement. The sample was 87.4% White/European American, 9.9% Black/African American, 1.8% Latino, and 0.9% Asian (M_{age} = 18.50; SD = 0.71).

Procedure

First, participants completed measures of basic need satisfaction and mood. All items were rated on 1–7 scales. The 21-item Basic Need Satisfaction Scale (Gagné, 2003) measures autonomy ('I feel like I am free to decide for myself how to live my life'), social relatedness ('I really like the people I interact with'), and competence ('Most days I feel a sense of accomplishment from what I do'). The composite was used as an overall measure of basic need satisfaction (M = 5.32, SD = 0.53, α = 0.72). Participants rated three PA (happy, cheerful, and pleased; M = 4.60, SD = 1.11, α = 0.86) and three NA (sad, angry, and worried; M = 2.12, SD = 1.07, α = 0.67) items.

Table 2. Correlations, Study 1.

	BIDR	MLQ-P	SWLS	PA	NA
Full sample					
BIDR	0.93	0.58	0.53	0.53	−0.39
MLQ-P		0.95	0.84	0.76	−0.60
SWLS			0.94	0.78	−0.61
PA				0.97	−0.65
NA					0.96
Honest (n = 60)					
BIDR	0.87	0.54	0.36	0.28	−0.39
MLQ-P		0.94	0.68	0.52	−0.34
SWLS			0.93	0.69	−0.48
PA				0.95	−0.29
NA					0.91
Fake Good (n = 58)					
BIDR	0.96	0.38	0.66	0.65	−0.33
MLQ-P		0.97	0.87	0.76	−0.48
SWLS			0.93	0.74	−0.37
PA				0.97	−0.49
NA					0.90
Fake Bad (n = 57)					
BIDR	0.88	0.37	0.39	0.34	−0.36
MLQ-P		0.90	0.83	0.77	−0.58
SWLS			0.93	0.71	−0.54
PA				0.96	−0.78
NA					0.98

Notes: $N = 176$; All correlations are significant, $p < 0.03$. BIDR = Balanced Inventory of Desirable Responding; MLQ-P = Meaning in Life Questionnaire presence subscale; SWLS = Satisfaction with Life Scale; PA = Positive Affect; NA = Negative Affect. Coefficients on the diagonals are α reliabilities.

Participants were then randomly assigned to one of two conditions. Bogus pipeline participants ($n = 55$) were told:

> In this study, we use a facial electromyogram, also known as an EMG, to measure facial muscle activity. This particular EMG device detects micro-pulses in the facial muscles that involuntarily react to a person's increased psychological arousal, which has been shown to be associated with deception. You can think of it as a kind of lie-detector test. In order to calibrate the EMG to your muscle activity, we have to conduct what is called a guilty knowledge test. I'll start up the program and it will show the numbers 1 through 5 on the screen. For each number, I'll ask you, 'Is this the number 4?' You should say 'no' to every number on the screen, even if it is the number 4. By doing this test, we can establish the differences in activity between an honest and dishonest reply.

After completing this task, participants were shown the purported results, a sound wave image showing a small spike 4/5ths of the way across. The experimenter then said, 'We've established the baseline. As you can see, there is significantly increased activity around the block where you stated that the number 4 was not a 4. This will be applied to the rest of the measures.'

Control participants ($n = 56$) were told that the purpose of the EMG was to assess 'facial muscle activity … while a person answers self-report questions.' Participants were shown the same numbered screens and simply identified the numbers to 'calibrate' the instrument. Then, they were shown the same results display and were told that 'random variance in activity is quite normal' if they asked about the small spike.

Participants then completed the MLQ-P and SWLS in counterbalanced order followed by additional positive (curious, happy, and joyful) and negative (nervous, sad, and angry) mood descriptors.

Results

Pre-test measures showed no condition differences, all t's $(109) < 0.63$, all p's > 0.53, suggesting successful random assignment. As Table 3 shows, as predicted, meaning in life and life satisfaction were significantly higher in the control condition compared to the bogus pipeline condition. Condition did not influence mood reports. Controlling for the covariates (R^2 change = 0.25, $p < 0.001$), condition (dummy coded: 0 = control, 1 = bogus pipeline) significantly predicted meaning in life ($\beta = -0.21$, R^2 change = 0.04, $p < 0.05$), and life satisfaction (for the covariates R^2 change = 0.41, $p < 0.001$; for condition $\beta = -0.23$, R^2 change = 0.05, $p < 0.01$). Hierarchical regressions tested whether condition moderated the relations of the covariates to meaning in life and life satisfaction. No interactions approached significance. The manipulation did not alter the relevance of these variables to well-being suggesting the manipulation changed the levels of well-being espoused, not their underlying meaning.

Study 3

Given the novelty of applying the bogus pipeline to well-being measures, we sought to replicate Study 2, while also addressing three limitations. First, the null effects of the manipulation on mood may have been due to the presence of mood pre-measures. Participants may have been motivated to provide post-test measures consistent with their previous responses (e.g. Carlsmith, Collins, & Helmreich, 1966; Festinger, 1957). Thus, Study 3 included no pre-measures. Second, Study 2 did not include a suspicion check. One was added to Study 3. Most importantly, Study 2 did not assess the proposed mediator of the condition effect, desirability bias. Study 3 included the BIDR for this purpose. We expected that condition differences would be explained by desirability bias.

Table 3. Descriptive statistics, correlations, and results, Studies 2 and 3.

	Study 2 ($N = 111$)				Study 3 ($N = 120$)				
	MLQ-P	SWLS	PA	NA	BIDR	MLQ-P	SWLS	PA	NA
Full sample									
BIDR					0.77	0.41**	0.31**	0.31**	−0.31**
MLQ-P	0.91	0.67**	0.34**	−0.37**		0.84	0.49**	0.39**	−0.25**
SWLS		0.83	0.50**	−0.47**			0.82	0.57**	−0.36**
PA			0.56	−0.45**				0.77	−0.23**
NA				0.66					0.77
Control									
BIDR						0.26*	0.20	0.19	−0.17
MLQ-P		0.57**	0.34**	−0.33*			0.37**	0.12	−0.04
SWLS			0.43**	−0.45**				0.49**	−0.27*
PA				−0.26*					−0.20
Bogus pipeline									
BIDR						0.51**	0.34**	0.32*	−0.38**
MLQ-P		0.73**	0.35**	−0.43**			0.56**	0.57***	−0.36**
SWLS			0.57**	−0.53**				0.58**	−0.39**
PA				−0.60**					−0.19
Grand M (SD)	4.60 (1.37)	4.92 (1.04)	4.96 (1.33)	2.20 (1.08)	8.43 (4.99)	4.63 (1.17)	4.64 (1.20)	4.43 (1.10)	2.92 (1.10)
Condition									
Control	4.87 (1.34)	5.40 (.90)	4.96 (1.22)	2.15 (1.22)	9.97 (5.10)	4.88 (1.15)	4.91 (1.15)	4.72 (1.03)	2.72 (1.04)
Pipeline	4.33 (1.36)	4.68 (1.13)	4.95 (1.45)	2.30 (1.15)	6.90 (4.41)	4.37 (1.14)	4.37 (1.19)	4.13 (1.10)	3.11 (1.13)
t	2.12*	2.39*	0.04	−0.75	3.52**	2.43*	2.53*	3.08**	−1.94+
d	0.41	0.46	0.01	0.14	0.64	0.45	0.46	0.56	0.36

Notes: Coefficients on the diagonal are α reliabilities. BIDR = Balanced Inventory of Desirable Responding; MLQ-P: Meaning in Life Questionnaire presence subscale, SWLS: Satisfaction with Life Scale, PA: Positive Affect, NA: Negative Affect. For Study 2 df = 109; for Study 3, df = 118.
*$p < 0.05$; **$p < 0.01$; +$p = 0.05$.

Method

Participants

One hundred twenty-one (88 women) undergraduates completed the study in partial fulfillment of a research participation requirement. Ages ranged from 18 to 22 ($M = 18.5$, $SD = 0.88$); ethnicities included White/European American (72%), Black/African American (14%), Asian (9%), Latino (3%), and other (2%).

Procedure

Participants were assigned randomly to control ($n = 60$) or bogus pipeline conditions ($n = 61$). Study 2 procedures were altered slightly: For both conditions, one of the two sensors was repositioned from the face to the left hand, to conform more closely to media depictions of lie detector tests. Instructions were modified to accommodate this change. After the 'guilty knowledge test' (or control procedures), participants completed the well-being measures from Study 2 and the BIDR.

Finally, participants were asked three suspicion probe questions. In response to the first, 'What were the sensors used to detect in this study?' one control participant indicated that the sensors were used to detect the truthfulness of responses. Of the participants in the bogus pipeline condition, 36 mentioned the detection of lying directly. Of the 25 in this group *not* explicitly stating that the purpose of the sensors was to detect deception, 23 described the process at a more detailed level. For example, 'They were used to detect the muscle usage in your face and hand to detect reactions to the questions.' This left only two participants in the bogus pipeline condition, one who responded 'no clue' and the other stating 'nothing, it was fake to make you think you had to tell the truth.' This participant's data were discarded.

Next, participants were asked, 'Do you think the sensors could actually do this?' Note that this question might have alerted participants to the possibility that the bogus pipeline was fake. Still, only 19 of the 121 participants (15.7%) responded 'No.' Of these, 12 were from the *control* condition and seven were from the bogus pipeline condition. For the association between condition and suspicion, $\chi^2(1) = 1.66$, $p = 0.20$.

Finally, participants were asked, 'Despite what you were told about the sensors, what do you think they were for?' In response to this question, four participants in the *control* condition indicated that they believed the sensors were used as a lie detector. The manipulation was, then, generally believable.[2]

Results

Table 3 shows that all well-being measures were significantly affected by the manipulation, as predicted. The bogus pipeline also led to significantly lower scores on the BIDR (replicating Alexander & Fisher, 2003).

Next, we examined the BIDR as a mediator of condition effects. Results of regression equations and mediational models (Preacher & Hayes, 2004) are shown in Table 4. As can be seen, for all but PA, the condition effects (dummy coded as in Study 2) were fully mediated by the BIDR. PA showed partial mediation. For all dependent measures, the indirect effect of condition through the BIDR was significant.

Brief discussion of Studies 2 and 3

Well-being ratings were inflated in the control relative to the bogus pipeline condition with effect sizes comparable to the overall effect size found in a bogus pipeline meta-analysis ($d = 0.41$; Roese & Jamieson, 1993). Moreover, Study 3 showed that condition differences were fully attributable to desirable responding, for all but PA. The reduction of PA in the bogus pipeline condition might reflect not only changes in desirability bias but perhaps the influence of the manipulation on affect, per se. Of course, it seems impractical that researchers will be able

Table 4. Results of mediational analyses, Study 3.

	Regression results		Bootstrapping results		
Dependent variable	β for condition	β for condition after controlling for BIDR	B (SE) for condition → DV	B (SE) condition → DV controlling for BIDR	Indirect effect (z)
MLQ-P	−0.22*	−0.10	−0.51 (0.21)*	−0.24 (0.21)	2.69*
SWLS	−0.23*	−0.15	−0.54 (0.21)*	−0.35 (0.22)	2.18*
Positive affect	−0.27*	−0.20*	−0.60 (0.19)*	−0.43 (0.20)*	2.12*
Negative affect	0.18$^+$	0.09	0.39 (0.20)*	0.20 (0.20)	−2.24*

Notes: Condition was coded 0 = control, 1 = bogus pipeline. BIDR = Balanced Inventory of Desirable Responding; MLQ-P = Meaning in Life Questionnaire presence subscale; SWLS = Satisfaction With Life Questionnaire. For bootstrapping, 3000 resamplings were used.
*$p < 0.05$; $^+p = 0.054$.

to incorporate laboratory bogus pipeline manipulations routinely in well-being research. As such, Studies 4 and 5 tested whether more feasible steps might remove such biases.

Study 4

Bogus pipeline effects can emerge even in the absence of an apparatus: Cohen, Jussim, Harber, and Bhasin (2009) found that simply warning participants that psychologists have ways of detecting deception led to higher reports of anti-Semitism. In addition, research on ethical behavior has shown that committing to honest responding reduces subsequent deceptive responding (Shu, Mazar, Gino, Ariely, & Bazerman, 2012). In Study 4, we employed each of these tacks to investigate whether participants would temper their well-being reports if they were led to believe that we could detect deception or if they committed to honest responding beforehand. We predicted that well-being ratings would be higher in the control than in the experimental conditions.

Method

Participants

Four hundred seventeen MTurk workers (229 men, 12 not reporting; paid $0.50) participated. Ages ranged from 18 to 69 ($M = 30.66$, $SD = 10.75$); ethnicities included White/European American (78.9%), Black/African American (5.5%), Latino (3.8%), Asian (7%), and other or not reporting (4.8%).

Procedure

Study 4 was initially conceived as two studies. In the first study, the control group ($n = 99$) was instructed to 'Please respond to the following scales honestly' and the experimental group ($n = 106$) was told, 'Please respond to the following scales honestly. Any deception on your part can be detected by sophisticated methods developed by psychologists. Responses to some of the following questions have been shown to reliably indicate dishonest responses.'

In the second study, the control group ($n = 106$) was instructed to 'Please respond to the following items using the scales provided,' while the experimental group ($n = 107$) read:

> The research in which you are about to participate relies on your honesty in answering each of the questions. Because of the importance of honest responding, there are items included in the questionnaires that are meant to detect deception. However, we know that even these can be faked by skilled liars. Thus, it remains crucial that each participant provides honest and accurate responses throughout the entire survey.

These participants then were asked to indicate their intention. All participants selected the option, 'On my honor, I certify that the answers I am about to give are honest to the best of my knowledge.'

Participants then completed the MLQ-P, SWLS, PA, and NA measures. Additionally, the BIDR was completed either before or after the well-being measures.

Results

Counter to predictions, comparisons of the control, and experimental groups in each study showed no effects of condition on well-being or the BIDR. For the first study, for well-being measures, all $t(202)$'s < 1.0. For the BIDR, $t(202) = 1.60$, $p = 0.11$. For the second study, all $t(211)$'s < 1.0.

A 2 (study) X 2 (condition, control vs. special instructions) X 2 (order, BIDR first vs. last) multivariate analysis of variance on the dependent measures showed no significant 3- or 2-way interactions and no main effects of study or condition, all F's $(5, 400) < 1.24$, all p's > 0.29. However, the multivariate main effect of the order of administration of the BIDR was significant, $F(5, 400) = 2.60$, $p = 0.025$. As such, analyses focused on this variable, pooling the two data-sets. In terms of mean levels of well-being, order affected only meaning in life, with marginal significance, $t(411) = 1.70$, $p = 0.09$, $d = 0.17$. Those who completed the BIDR first reported lower meaning in life, $M = 4.32$ (1.59), compared to those who completed it last, $M = 4.58$ (1.54).

We next shifted focus to the relationships between the BIDR and well-being reports. Table 5 shows substantial correlations between the well-being measures and the BIDR, overall, and in each instruction condition. We examined whether order of administration influenced the relationship between desirable responding and well-being, testing whether completing the BIDR first led to well-being ratings that were relatively less contaminated by desirability bias. In hierarchical regressions, main effects of centered BIDR scores and order (dummy coded: 0 = BIDR first, 1 = BIDR last) were entered on the first step followed by the product of these for each dependent measure. For meaning in life, in the absence of a main effect for order, a main effect for the BIDR, $\beta = 0.30$, $p < 0.001$, was qualified by a marginal BIDR X order interaction, $\beta = 0.11$, $p = 0.095$. For life satisfaction, in the absence of a main effect for order, a main effect of BIDR, $\beta = 0.23$, $p = 0.001$, was qualified by a BIDR X order interaction, $\beta = 0.14$, $p = 0.05$. Similarly, for NA, a main effect for BIDR, $\beta = -0.14$, $p = 0.05$, was qualified by a significant interaction, $\beta = -0.21$, $p = 0.003$. Finally, for PA, only a significant main effect for the BIDR emerged, $\beta = 0.23$, $p = 0.002$. To probe the interactions, we computed the correlations between BIDR and well-being measures in each order. Table 5 shows that

Table 5. Descriptive statistics and correlations among measures, Study 4.

	BIDR	MLQ-P	SWLS	PA	NA
BIDR	0.80	0.39	0.33	0.29	−0.30
MLQ-P		0.94	0.68	0.56	−0.32
SWLS			0.92	0.71	−0.35
PA				0.94	−0.36
NA					0.89
Grand M (SD)	9.71 (5.99)	4.45 (1.57)	4.10 (1.55)	4.16 (1.50)	2.81 (1.53)
Correlations with the BIDR by instruction condition					
Control (n = 205)		0.36**	0.38**	0.33**	−0.33**
Experimental (n = 212)		0.42**	0.29**	0.25*	−0.27*
Correlations with the BIDR by order					
BIDR last (n = 209)		0.49**	0.43**	0.35**	−0.44**
BIDR first (n = 204)		0.28**	0.22**	0.21*	−0.13
z		2.51*	2.38*	1.51	−3.44**

Notes: $N = 417$; All correlations within the full sample are significant, $p < 0.001$. BIDR = Balanced Inventory of Desirable Responding; MLQ-P = Meaning in Life Questionnaire; PA = Positive Affect, NA = Negative Affect. Coefficients on the diagonal are α reliabilities. For the correlations within conditions and orders and z tests. For z tests, p values are one-tailed. The interaction between order and BIDR was marginal or significant for all measures except positive affect.
*$p < 0.01$; **$p < 0.001$.

when the BIDR was administered first, well-being ratings were less strongly related to this measure. These results suggest that completing a desirability measure prior to completing well-being measures led to less contamination in well-being ratings though these changes did not alter mean level responses.

Study 5

Given the *post hoc* nature of the order effects on the correlations in Study 4, we sought to test the effects of order directly by administering the BIDR and the well-being measures in a counterbalanced order in the absence of other manipulations.

Method

Participants and procedures

Three hundred ninety-one MTurk workers (254 men, 8 not reporting; paid $0.50) participated.[3] Ages ranged from 18 to 70 ($M = 32.73$, $SD = 11.78$); ethnicities included White/European American (76.2%), Black/African American (6.1%), Latino, (5.6%), Asian (7.7%), and other or not reporting (4.3%). Participants completed the Study 4 measures with the BIDR administered either prior to ($n = 188$) or after ($n = 203$) the well-being measures.

Results

Table 6 shows the correlations and descriptive statistics for all measures. As can be seen, overall, the BIDR was significantly correlated with all well-being measures. We computed a series of equations, regressing each well-being measure on the main effects of BIDR and dummy-coded order (as in Study 4) and their interaction. In the presence of main effects of BIDR on each of the well-being measures, β's $\geq |0.22|$, p's ≤ 0.004, there were no order main effects, β's $\leq |0.06|$, p's ≥ 0.24, and no interactions, β's $\leq |0.10|$, p's ≥ 0.17. Finally, we computed the correlations between the BIDR and the well-being measures within each order. As Table 6 shows, order did not affect the magnitude of the correlations between the BIDR and well-being measures, z's ≤ 0.96, p's ≥ 0.34. Thus, unfortunately, the simple strategy of administering a measure of desirability bias prior to measuring well-being did not influence the relevance of that bias to well-being reports.

General discussion

These studies examined the role of desirability bias in self-reports of well-being. Study 1 showed that participants inflated well-being reports to make a good impression (and lowered them to make a bad impression). Studies 2 and 3 showed that when led to believe that deception could be detected, participants reported lower well-being, a shift that supports the notion that these ratings are exaggerated under typical measurement conditions. Further, Study 3 showed that these differences were generally explained by changes in desirability bias. In Studies 4 and 5 attempts to reduce this bias with instruction or order manipulations were unsuccessful. In sum, the current studies support the conclusion that

Table 6. Descriptive statistics and correlations among measures, Study 5.

	BIDR	MLQ-P	SWLS	PA	NA
BIDR	0.83	0.29**	0.20**	0.20***	−0.25**
MLQ-P		0.94	0.62**	0.55**	−0.27**
SWLS			0.92	0.69**	−0.37**
PA				0.92	−0.38**
NA					0.89
Correlations with the BIDR by order					
BIDR last (n = 203)		0.24**	0.19*	0.20*	−0.28**
BIDR first (n = 188)		0.33**	0.20*	0.20*	−0.21*
Grand M (SD)	10.17 (6.16)	4.23 (1.64)	3.92 (1.57)	4.25 (1.46)	2.54 (1.42)
Means by order					
BIDR last (n = 203)	9.90 (6.40)	4.15 (1. 58)	3.88 (1.49)	4.22 (1.48)	2.43 (1.43)
BIDR first (n = 188)	10.46 (5.89)	4.32 (1.70)	3.96 (1.65)	4.28 (1.44)	2. 63 (1.41)
t(389)	0.90	1.04	0.52	−0.44	−1.36

Notes: BIDR = Balanced Inventory of Desirable Responding; MLQ-P = Meaning in Life Questionnaire presence subscale; SWLS = Satisfaction with Life Scale; PA = Positive Affect, NA = Negative Affect. Coefficients on the diagonal are α reliabilities.
N = 391; All correlations are significant.
*p < 0.05; **p < 0.001.

portraying one's life as satisfying, meaningful, and filled with happiness is socially desirable and reports on these constructs can be inflated when the motivation to respond in a valued manner remains greater than the motivation to respond honestly.

Correlational data (from Studies 1, 3–5) show consistent relationships between well-being measures and desirability bias. Table 7 shows the confidence intervals and sample size weighted average correlations between the BIDR and each well-being variable across the studies. These associations are not trivial, ranging from 0.26 for PA to 0.36 for meaning in life.

Generally, anonymity is considered an effective means of reducing social desirability concerns. Yet, Studies 1, 4, and 5 were conducted in anonymous online formats. Furthermore, the subscales of the BIDR consistently showed parallel relationships with well-being. These findings may indicate that bias in well-being measures is not exclusively social in nature. The desirability of well-being may be internalized such that participants are hesitant to admit to themselves, without added incentive, that their lives fall short of some optimum. Further research is required to test the potential effects of the salience of one's anonymity for the goal of accurately assessing well-being as well as addressing potential differences in social and personal desirability bias in well-being responses.

Although a number of alternate methodological strategies such as peer reports, bio-medical markers, and statistical controls (e.g. Podsakoff, MacKenzie, Lee, & Podsakoff, 2003), have been developed to circumvent or account for biases associated with self-report responses for many constructs, the inherent subjectivity of well-being makes these options unsatisfactory. Friends, blood-tests, or statistical formulas are all unlikely to have a better understanding of one's well-being than the person himself or herself. The present results suggest that when using self-report to assess these variables, researchers should be aware of potential desirability biases and take steps to reduce inaccurate responding.

Commonly, correcting for desirable responding in self-report ratings involves controlling for responses on social desirability scales; however, this practice may actually reduce, rather than improve, the validity of some measures (McCrae & Costa, 1983). The present data cannot address whether well-being reports that are relatively more (or less) contaminated by desirability bias differ in their correlates or consequences. It might be that reporting oneself as high on well-being reflects awareness of and sensitivity to the social value of these characteristics, tendencies that, themselves might be related to success (cf. van der Linden, te Nijenhuis, & Bakker, 2010).

Do desirable response patterns matter in well-being research?

Though exaggerated responding to well-being measures may seem harmless, painting an unrealistically rosy picture of one's life can have implications. For instance, falsely espousing happiness, life satisfaction, and meaning in life, whether through unconscious self-deception or public faking, could diminish the motivation for change and decrease the chances that others will offer aid when needed.

One might question whether the mean level differences shown in Studies 2 and 3 are substantively important. After all, Study 2 showed that the associations

Table 7. Summary of desirability bias in well-being measures.

	N	*r*	95% CI	Estimated raw score difference attributable to desirability bias
MLQ-P				
Study 1	60 (control)	0.54	0.37 to 0.69	
Study 2	111			0.55
Study 3	120	0.41	0.27 to 0.54	0.52
Study 4	417	0.39	0.31 to 0.46	
Study 5	391	0.29	0.19 to 0.37	
Sample size weighted average	1099	0.36	0.33 to 0.39	
SWLS				
Study 1	60 (control)	0.36	0.11 to 0.54	
Study 2	111			0.51
Study 3	120	0.31	0.07 to 0.54	0.54
Study 4	417	0.33	0.25 to 0.42	
Study 5	391	0.20	0.10 to 0.29	
Sample size weighted average	1099	0.28	0.25 to 0.31	
PA				
Study 1	60 (control)	0.28	0.14 to 0.62	
Study 2	111			0.01
Study 3	120	0.31	0.13 to 0.47	0.60
Study 4	417	0.29	0.20 to 0.38	
Study 5	391	0.20	0.11 to 0.30	
Sample size weighted average	1099	0.26	0.29 to 0.23	
NA				
Study 1	60 (control)	−0.39	−0.32 to −0.61	
Study 2	111			−0.18
Study 3	120	−0.31	−0.13 to −0.47	−0.39
Study 4	417	−0.30	−0.22 to −0.37	
Study 5	391	−0.25	−0.17 to −0.34	
Sample size weighted average	1099	−0.29	−0.26 to −0.32	

Notes: All correlations are statistically significant, $p < 0.001$. Bootstrapped confidence intervals based on 1000 re-samplings. For the confidence interval (+/−1 *SE*) of the weighted average correlation, the *r* was transformed to z, the standard error was calculated and then the *z*'s were transformed back to *r*'s. MLQ-P = Meaning in Life Questionnaire presence subscale; SWLS = Satisfaction with Life Scale; PA = Positive Affect; NA = Negative Affect. Estimated difference attributable to bias for Studies 2 and 3 was calculated by multiplying the group mean weighted standard deviation by the effect size.

between pre-test measures of mood and basic need satisfaction with well-being were not changed by the bogus pipeline manipulation. The final column of Table 7 shows the average difference in the means for each of the well-being measures in the bogus pipeline. These differences were not enormous, generally about a half a point in raw scores. If desirability bias only inflates ratings by this small constant, does it matter?

The answer to this question depends on how well-being reports are used. The current results suggest that contamination of well-being by desirability may not be critical to studies examining the relationships among well-being and other variables. However, if those other variables are measured using self-report and are desirable themselves, this shared desirability may inflate the associations and is best controlled for, statistically.

The effects of desirability are likely to be more important when mean levels of well-being are of interest, for instance, in comparisons between individuals (e.g. in hiring decisions; Luthans, Youssef et al., 2007) and across groups (e.g. by policy-makers in their quest to improve standards of living; UN News Centre, 2011). Even relatively small mean differences in well-being produced by desirability bias would seem to be particularly important in cross-group comparisons, using large samples. Systematic group differences in the social desirability of well-being may inflate (or mask) mean differences in well-being across groups.

In cross-cultural studies examining the potential underlying mechanisms of the difference in well-being reports of East Asians vs. Americans, Diener, Suh, Smith, and Shao (1995) found that Americans reported higher well-being and higher cultural valuing of positive affective states than did East Asians. Beliefs about how much one's culture values emotional experiences were (sometimes strongly) correlated with self-reports of those experiences (*r*'s ranging from 0.30 to 0.72; Diener, Suh et al., 1995). Cross-cultural (and cross-group) comparisons of ratings on well-being measures, then, might be problematic since these responses are influenced by cultural (and other group) norms.

It is important, then, for future research to assess the desirability of well-being within groups prior to comparing well-being across groups. The bogus pipeline

procedure offers a means by which to establish within group desirability estimates as it can provide information about the nuances of the direction of group norms. For instance, compared to control conditions, the bogus pipeline led to reports of *fewer* sex partners by men, but *more* partners by women (Alexander & Fischer, 2003). Similarly, this procedure might be utilized to identify cultural and group differences in norms for well-being. With established estimates of desirability bias within a group, it might then be appropriate to correct for within-group bias before making between-group comparisons. Research focusing on group differences in desirability influences in well-being measures is crucial if these measures are to be used in wide-scale comparisons across groups, countries, and cultures.

The recent promotion of well-being seems to have had important consequences for the scientific measurement of these constructs that were previously only of trivial concern. Our findings are perhaps troubling, suggesting that desirability bias is prevalent in common well-being measures. Our goal is not to question the importance of well-being research but to suggest vigilance to the potential effects of bias moving forward. The increasing value placed on subjective well-being coupled with the increasingly global nature of well-being research (Diener, Ng, Harter, & Arora, 2010) creates a challenge for accurate measurement of these variables. As well-being becomes more important to individuals, employers, and governments, desirability bias will likely play a larger role in these measures.

Notes

1. In Studies 1 and 3, half of the participants completed the BIDR first and the other half last. Analyses for all studies showed no main effects of order and no order by condition interactions so results collapse across order. We examine order effects explicitly in Studies 4 and 5.
2. Analyses were repeated dropping the six additional participants indicating suspicion in their responses to questions 1 and 3. With the exception of NA, which dropped to marginal significance $t(115) = -1.75$, $p = 0.08$, responses to all well-being scales and the BIDR remained significantly different by condition.
3. Data from 20 participants were discarded prior to analyses based on an initial data screening for duplicate IP addresses.

References

Alexander, M. G., & Fisher, T. D. (2003). Truth and consequences: Using the bogus pipeline to examine sex differences in self-reported sexuality. *Journal of Sex Research, 40*, 27–35.

Avey, J. B., Patera, J. L., & West, B. J. (2006). The implications of positive psychological capital on employee absenteeism. *Journal of Leadership and Organizational Studies, 13*, 42–60.

Carlsmith, J. M., Collins, B. E., & Helmreich, R. L. (1966). Studies in forced compliance: I. The effect of pressure for compliance on attitude change produced by face-to-face role playing and anonymous essay writing. *Journal of Personality and Social Psychology, 4*, 1–13.

Cohen, F., Jussim, L., Harber, K. D., & Bhasin, G. (2009). Modern anti-semitism and anti-Israeli attitudes. *Journal of Personality and Social Psychology, 97*, 290–306.

Crowne, D. P., & Marlowe, D. A. (1964). *The approval motive*. New York, NY: Wiley.

Diener, E., & Diener, C. (1996). Most people are happy. *Psychological Science, 7*, 181–185.

Diener, E., & Emmons, R. A. (1984). The independence of positive and negative affect. *Journal of Personality and Social Psychology, 47*, 1105–1117.

Diener, E., Emmons, R., Larsen, R., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment, 49*, 71–75.

Diener, E., Ng, W., Harter, J., & Arora, R. (2010). Wealth and happiness across the world: Material prosperity predicts life evaluation, whereas psychosocial prosperity predicts positive feeling. *Journal of Personality and Social Psychology, 99*, 52–61.

Diener, E., Smith, H., & Fujita, F. (1995). The personality structure of affect. *Journal of Personality and Social Psychology, 69*, 130–141.

Diener, E., Suh, E. M., Smith, H., & Shao, L. (1995). National differences in reported subjective well-being: Why do they occur? *Social Indicators Research, 34*, 7–32.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.

Gagné, M. (2003). The role of autonomy support and autonomy orientation in prosocial behavior engagement. *Motivation and Emotion, 27*, 199–223.

Hart, K. E., & Sasso, T. (2011). Mapping the contours of contemporary positive psychology. *Canadian Psychology, 52*, 82–92.

Heintzelman, S. J., & King, L. A. (in press). Life is pretty meaningful. *American Psychologist*.

Jones, E., & Sigall, H. (1971). The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin, 76*, 349–364.

Luthans, F., Avolio, B. J., Avey, J. B., & Norman, S. M. (2007). Positive psychological capital: Measurement and relationship with performance and satisfaction. *Personnel Psychology, 60*, 541–572.

Luthans, F., Youssef, C. M., & Avolio, B. J. (2007). *Psychological capital: Developing the human competitive edge*. New York, NY: Oxford University Press.

Lyubomirsky, S., King, L. A., & Diener, E. (2005). The benefits of frequent positive affect: Does happiness lead to success? *Psychological Bulletin, 131*, 803–855.

McCrae, R. R., & Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology, 51*, 882–888.

Mummendey, H. D., & Bolten, H. (1981). Die veränderung von social-desirability-antworten bei erwarteter warheitskontrolle (bogus-pipeline-paradigma) [Modification of social desirability responses when control of truthful responding is anticipated: The bogus pipeline paradigm]. *Zeitschrift für Differentielle und Diagnostische Psychologie, 2*, 151–156.

Paulhus, D. L., & Reid, D. B. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology, 60*, 307–317.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*, 879–903.

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, and Computers, 36*, 717–731.

Pressman, S. D., & Cohen, S. (2005). Does positive affect influence health? *Psychological Bulletin, 131*, 925–971.

Roese, N. J., & Jamieson, D. W. (1993). Twenty years of bogus pipeline research: A critical review and meta-analysis. *Psychological Bulletin, 114*, 363–375.

Schaubroeck, J. M., Riolli, L. T., Peng, A. C., & Spain, E. S. (2011). Resilience to traumatic exposure among soldiers deployed in combat. *Journal of Occupational Health Psychology, 16*, 18–37.

Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist, 55*, 5–14.

Seligman, M. E. P., Steen, T. A., Park, N., & Peterson, C. (2005). Positive psychology progress: Empirical validation of interventions. *American Psychologist, 60*, 410–421.

Shu, L. L., Mazar, N., Gino, F., Ariely, D., & Bazerman, M. H. (2012). Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end. *Proceedings of the National Academy of Sciences, 109*, 15197–15200.

Soubelet, A., & Salthouse, T. A. (2011). Influence of social desirability on age differences in self-reports of mood and personality. *Journal of Personality, 79*, 741–762.

Steger, M. F., Frazier, P., Oishi, S., & Kaler, M. (2006). The meaning in life questionnaire: Assessing the presence of and search for meaning in life. *Journal of Counseling Psychology, 53*, 80–93.

U.N. News Centre. (2011, July 19). *Happiness should have greater role in development policy – UN Member States*. Retrieved from http://www.un.org/apps/news/story.asp?NewsID=39084

van der Linden, D., te Nijenhuis, J., & Bakker, A. B. (2010). The general factor of personality: A meta-analysis of Big Five intercorrelations and a criterion-related validity study. *Journal of Research in Personality, 44*, 315–327.

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

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*, 1063–1070.