Social desirability is the same in offline, online, and paper surveys
A meta-analysis

Dodou, D.; de Winter, J.C.F.

DOI
10.1016/j.chb.2014.04.005

Publication date
2014

Document Version
Final published version

Published in
Computers in Human Behavior

Citation (APA)
Dodou, D., & de Winter, J. C. F. (2014). Social desirability is the same in offline, online, and paper surveys: A meta-analysis. *Computers in Human Behavior*, *36*(July), 487-495. https://doi.org/10.1016/j.chb.2014.04.005

Computers in Human Behavior 36 (2014) 487–495

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Review

Social desirability is the same in offline, online, and paper surveys: A meta-analysis

D. Dodou *, J.C.F. de Winter

Department of BioMechanical Engineering, Delft University of Technology, Delft, The Netherlands

ARTICLE INFO

Article history:
Available online 13 May 2014

Keywords:
Social desirability
Meta-analysis
Self-disclosure
Decline effect

ABSTRACT

The aim of this meta-analysis was to compare social desirability scores between paper and computer surveys. Subgroup analyses were conducted with Internet connectivity, level of anonymity, individual or group test setting, possibility of skipping items, possibility of backtracking previous items, inclusion of questions of sensitive nature, and social desirability scale type as moderators. Subgroup analyses were also conducted for study characteristics, namely the randomisation of participants, sample type (students vs. other), and study design (between- vs. within-subjects). Social desirability scores between the two administration modes were compared for 51 studies that included 62 independent samples and 16,700 unique participants. The overall effect of administration mode was close to zero (Cohen's $d = 0.00$ for fixed-effect and $d = -0.01$ for random-effects meta-analysis). The majority of the effect sizes in the subgroup analyses were not significantly different from zero either. The effect sizes were close to zero for both Internet and offline surveys. In conclusion, the totality of evidence indicates that there is no difference in social desirability between paper-and-pencil surveys and computer surveys. Publication year and sample size were positively correlated ($\rho = .64$), which suggests that certain of the large effects that have been found in the past may have been due to sampling error.

Contents

1. Introduction

Computer-based surveys are widely used in social and behavioural research (Buchanan & Smith, 1999; Gosling, Vazire, Srivastava, & John, 2004). Computers enable large-scale cost-effective data collection, automatic response time monitoring, and error-free transcription (Birnbaum, 2000; Cook, Heath, Thompson, & Thompson, 2001; Fox & Schwartz, 2002; Griffis, Goldsby, & Cooper, 2003). Despite the popularity of computers in survey studies, researchers often question whether computer surveys are as valid as their paper counterparts. A particularly common question is whether computer and paper-and-pencil surveys evoke the same amount of social desirability.

* Corresponding author. Address: Department of BioMechanical Engineering, Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands. Tel: +31 15 2784221.
E-mail address: d.dodou@tudelft.nl (D. Dodou).

http://dx.doi.org/10.1016/j.chb.2014.04.005
0747-5632/

Social desirability is inversely related to the degree of privacy and anonymity that a person experiences (Ben-Ze'ev, 2003; Buchanan, 2000; Davis, 1999; Fisher, 1993; Joinson, 1999; Pasveer & Ellard, 1998; Smith & Leigh, 1997). It has been argued that early personal computers offered limited social context, making the respondent feel anonymous and self-absorbed (Sproull & Kiesler, 1986). As early as 1963, Smith argued that individuals are more honest when "confessing to a machine". Evan and Miller (1969) found that people who completed a computer version of a survey scored lower on a lie scale than did people who completed the paper version of the same survey. This finding led the authors to conclude that "a computer ... guaranteed ... a greater sense of privacy and anonymity than the conventional situation wherein questionnaires or psychological tests are filled out by hand and then scrutinized, scored, and interpreted directly by other human beings" (p. 216).

A large number of studies have been published since the seminal works conducted in the 1960s. Some studies have found that, compared to paper surveys, computer surveys yield lower social desirability (Joinson, 1999; Kiesler & Sproull, 1986; Martin & Nagao, 1989), but other studies have reported higher social desirability (Davis & Cowles, 1989; Lautenschlager & Flaherty, 1990; Rosenfeld, Booth-Kewley, Edwards, & Thomas, 1996; Whitener & Klein, 1995), and some have found no statistically significant effects of administration mode (e.g., Booth-Kewley, Edwards, & Rosenfeld, 1992; Finegan & Allen, 1995; Fox & Schwartz, 2002; Hancock & Flowers, 2001; Wilkerson, Nagao, & Martin, 2002).

In some computer surveys, primarily early ones, items were presented one-by-one, and there was no possibility to backtrack, which is an intrinsic characteristic of paper-and-pencil surveys (King & Miles, 1995). It is not clear how backtracking affects social desirability scores. Some authors have suggested that the possibility of backtracking *increases* socially desirable responses because the respondents are able to 'manipulate' their previous answers (Fox & Schwartz, 2002), whereas others have argued that backtracking creates a sense of trust and therefore *decreases* social desirability (Whitener & Klein, 1995).

Four meta-analyses have previously summarised comparisons of social desirability scores between computer and paper-and-pencil surveys. Weisband and Kiesler (1996) meta-analysed 30 studies that were published between 1967 and 1994 and found that computers evoked greater self-disclosure than did paper surveys (Cohen's $d = 0.20$), with the difference between administration modes reducing over time. Richman, Kiesler, Weisband, and Drasgow (1999) extended Weisband and Kiesler's work by meta-analysing 42 studies published between 1967 and 1997, and found a near-zero effect ($d = 0.05$) between the scores obtained with paper-and-pencil and computer surveys. Note that only one-third of the studies in Richman et al. (1999) used social desirability scales. In the remaining two-thirds of the studies, social desirability was inferred from the mean scores of the surveys. For example, a lower score in a symptom checklist was interpreted to indicate higher social desirability. When meta-analysing only the studies that used social desirability scales, a near-zero effect size ($d = 0.01$) between the scores obtained with paper-and-pencil surveys and those obtained from computer surveys was found. Subgroup analyses further revealed that computer surveys yielded lower social desirability than did paper surveys, when both conducted in an individual setting. When skipping/backtracking was allowed, computer surveys were associated with lower social desirability than were paper surveys, whereas when skipping/backtracking was not allowed in the computerised version, the difference between the computer and paper surveys was considerably smaller. Linear regression analysis using publication year as one of the independent variables revealed that there was a tendency to respond with greater honesty in computerised surveys in the older studies.

In a meta-analysis of 24 studies that were published between 1969 and 1997, Dwight and Feigelson (2000) distinguished between impression management (attempting to convince others about inflated engagement in socially desirable behaviours and concealing from others the engagement in socially undesirable behaviours) and self-deception (convincing oneself about inflated engagement in socially desirable behaviours and concealing from oneself the engagement in socially undesirable behaviours). Dwight and Feigelson found that impression management was slightly but significantly lower ($d = -0.08$) for computer surveys than for paper surveys, and that, in line with the results of Richman et al. (1999), the strength of this effect was diminishing over time (correlation between publication year and effect size = .44), with older studies reporting lower impression management for computer surveys than for paper surveys. Self-deception did not differ between the two administration modes ($d = 0.04$ based on 25 effect sizes).

Tourangeau and Yan (2007) estimated the mean effect size across 10 studies that compared computer- and paper-based administration of sensitive questionnaires. The computer mode included not only self-administered surveys but also interactive voice responses and audio computer-assisted self-interviews. Only 4 of the 10 studies included social desirability scales, and all 4 of these studies were included both in Richman et al. (1999) and Dwight and Feigelson (2000). No significant difference between the administration modes was found across these 4 studies ($d = -0.02$, where a negative effect indicates lower social desirability for the computer surveys compared to the paper surveys).

Since these previous meta-analyses, a large number of studies comparing the effect of administration mode on social desirability response have been published. None of the studies that were included in the previous meta-analyses included computer surveys that were completed via the Internet, which is logical because Internet surveys were uncommon before the late 1990s, when the previous meta-analyses were published. Social desirability in Internet surveys may be associated with negative experiences, such as online criminality (De Zwart, Lindsay, Henderson, & Phillips, 2011), the "Big Brother syndrome" (defined as the "growing and pervasive fear of computers monitoring and controlling people's lives", Rosenfeld et al., 1996, p. 265–266), and the increasing awareness that online data are monitored, stored, shared, and leaked.

The aim of this study was to provide an updated meta-analysis that compares the social desirability scores of paper and computer surveys. Subgroup analyses were conducted using Internet connectivity, level of anonymity, test setting, possibility of skipping items in the computer surveys, possibility of backtracking previous answers in the computer surveys, inclusion of questions of sensitive nature, and social desirability scale type as moderators. Subgroup analyses were also conducted for study characteristics, namely the randomisation of participants between administration modes, the sample type (students or other), and the study design (between- or within-subjects). Of these moderators, Internet connectivity, social desirability scale type, and the three moderators that are related to the study characteristics were not examined in the four previous meta-analyses.

2. Method

A literature search (last update: 28 January 2014) was conducted in Google Scholar and Web of Knowledge. Google Scholar was used because it enables full-text search and provides access to a large number of articles, reports, theses, and conference papers, and because it is a recommended tool for systematic reviews (Gehanno, Rollin, & Darmoni, 2013; Shariff et al., 2013).

Web of Knowledge was used because it covers early publications that are not available online and that Google Scholar may therefore have failed to retrieve (De Winter, Zadpoor, & Dodou, 2014).

Social desirability can be measured either by means of formal social desirability scales or indirectly by means of responses to sensitive questions, that is, via the underreporting of socially undesirable behaviours (e.g., illegal substance abuse, alcohol consumption, violation of the law) and characteristics (e.g., physical or mental illness) and the overreporting of socially desirable properties (e.g., participation in voting, financial wealth). Richman et al. (1999) noted that "instruments specifically designed to measure social desirability distortion... are more reliable and direct indicators of distortion than instruments used to measure diverse other traits, syndromes, attitudes, and behavior" (p. 757). In line with Richman et al.'s argument and to ensure a meaningful comparison across studies, this meta-analysis only considered social desirability that was measured by means of formal social desirability scales.

We conducted six separate search queries by combining the terms "social desirability", "computer", and "paper and pencil" with each of the social desirability and lie scales reviewed by Paulhus (1991) and Tan and Grace (2008), that is: (1) "Edwards Social Desirability Scale", (2) "Marlowe-Crowne" for Marlowe-Crowne Social Desirability Scale (MCSD), (3) "BIDR" for Balanced Inventory of Desirable Responding, (4) "RD-16", (5) "Children's Social Desirability Scale", (6) "MMPI" for the Minnesota Multiphasic Personality Inventory, (7) "Self-Deception Questionnaire" "Other-Deception Questionnaire", (8) "Short Social Desirability Scale", (9) "Basic Personality Inventory" "Denial", (10) "Social Acquiescence Scale", (11) "Sexual Social Desirability Scale", and (12) "Wiggins Social Desirability". We also conducted a Google Scholar search query for the combination of the keywords: "lie scale", "computer", "paper and pencil", and "Eysenck" to retrieve studies that included the lie scale of the Eysenck Personality Questionnaire.

The abstracts and/or full texts of the retrieved articles were read to identify whether the inclusion criteria were fulfilled. The reference lists of the meta-analyses by Richman et al. (1999) and Dwight and Feigelson (2000), of a meta-analysis on the equivalence of paper and computer versions of the MMPI (Finger & Ones, 1999), of a systematic review of 381 studies on survey research (Hood et al., 2012), and of each study that fulfilled the inclusion criteria were reviewed to retrieve additional studies. In the Web of Knowledge, the following Boolean syntax was used: Topic = ("social desirability") AND Topic = (computer) AND Topic = (paper). Additional extensive opportunistic searches were conducted with the general Google search engine.

Our inclusion criterion was whether the scores on a social desirability or lie scale were reported for both a paper and a computer version of the same survey. Studies that did not mention social desirability scores separately for each administration mode but did mention a correlation coefficient or *t*-value between the two modes were also eligible. Any type of publication (journal article, book, dissertation, etc.) in any language was eligible for inclusion. If multiple studies reported results from the same respondents, only the study that included the most usable data was included in the analysis. If the data were incomplete, the authors were contacted for further information.

The following data were collected for each study (when available): publication year, sample size, type of survey, type of social desirability scale, percentage of males in the sample, mean age, age range and/or standard deviation, study design (within vs. between subjects), sample type (students or other), group assignment method (randomised vs. not randomised, where not randomised indicates that either the participants chose the administration mode themselves or, in cases of student samples, that the participants in the paper and computer groups were recruited from different classes), Internet connectivity, anonymity, test setting (group vs. individual), possibility to skip items in the computer survey (either by allowing blanks or by offering the option of "cannot say"/"do not answer"), possibility to backtracking responses in the computer survey, and inclusion of sensitive questions (e.g., questions about drug use, sexual behaviour, or religious orientation; for definitions of sensitive questions see Krumpal, 2013; Tourangeau & Yan, 2007). Studies that did not report whether the computer surveys were completed over the Internet were coded as offline (this coding is likely accurate for the studies published before the late 1990s when the notion of the Internet was non-existent or not yet widespread). Studies in which the computers were connected only to a local network and not on Internet were also coded as offline.

To reduce the risk of data extraction error, for the studies that were also meta-analysed by Richman et al. (1999) or Dwight and Feigelson (2000), we compared our extracted data and effect sizes with the values reported in these two meta-analyses. Disagreements between our values and the values from one or both of these previous meta-analyses were resolved through discussion between the two authors.

Some studies reported results from multiple experiments (typically distinguished as "Experiment 1", "Experiment 2", or "Study 1", "Study 2", etc.). As long as the experiments were conducted with different participants, under different experimental conditions (e.g., half of the sample was tested in a group setting, and the other half was tested in an individual setting), and at different times, the samples of these experiments were treated as independent.

Some studies (Honaker, Harrell, & Buffaloe, 1988; Toppins, 1985; White, Clements, & Fowler, 1985) included four groups: one group was tested twice with a paper survey, one group was tested twice with a computer survey, and two groups were tested with both paper and computer surveys in alternate orders (i.e., one group was tested first with the computer survey and then with the paper survey, and the other was tested first with the paper survey and then with the computer survey). For these studies, we used the first take of the two groups that were tested twice with the same medium for between-subjects comparison of administration modes (one independent sample) and the other two groups for within-subjects comparisons (another independent sample, for which the effect size was calculated as the weighted average of the effect sizes of the two groups).

The mean standardised difference (Cohen's *d*) between the social desirability scores from the paper and the computer version of a survey was calculated for each sample as the difference between the mean scores of the two administration modes divided by their pooled standard deviation. If means but no standard deviations, *t*-values, or *F*-values were reported, the standard deviations were estimated based on the mean of the standard deviations reported in all other studies using the same scale, weighted by the sample size of each study.

Summary effect sizes were calculated using both fixed-effect and random-effects meta-analysis. Although many have argued in favour of random-effects meta-analysis (Hunter & Schmidt, 2000), fixed-effect and random-effects meta-analyses are two different methods of summarising data and have different aims (conditional versus unconditional inference; Hedges & Vevea, 1998). For both types of meta-analysis, the methodology of Borenstein, Hedges, Higgins, and Rothstein (2011) was followed. Heterogeneity among studies was assessed via the I^2 statistic, which is defined as the percentage of the observed variances that was due to real differences in effect sizes (Higgins & Green 2011; Higgins, Thompson, Deeks, & Altman, 2003):

$$I^2 = \left(\frac{Q - df}{Q}\right) \times 100\%,$$

where $df = k - 1$ is the degrees of freedom, k is the number of studies, and Q is the observed weighted sum of squares:

$$Q = \sum_{i=1}^{k} W_i Y_i^2 - \frac{\left(\sum_{i=1}^{k} W_i Y_i\right)^2}{\sum_{i=1}^{k} W_i},$$

where W_i is the weight of the study i (i.e., the inverse variance, $1/V_i$), and Y_i is the effect size of study i (Borenstein et al., 2011).

Subgroup analyses were conducted with the following moderators: Internet connectivity, anonymity, test setting, possibility of scanning in the computer surveys, possibility of backtracking in the computer surveys, inclusion of questions of sensitive nature in the surveys, social desirability scale type, sample type, study design, and randomisation of participants between administration modes. In our analysis, positive effect sizes represent higher social desirability in the computer survey as compared to the paper survey. Sample-size weighted (i.e., fixed-effect) linear regression and sample-size weighted Spearman correlations between the effect sizes (one effect size per sample) and publication year were calculated to evaluate longitudinal trends.

All analyses were conducted in MATLAB (Version R2012b, The MathWorks, Inc., Natick, MA). The employed script is provided in the Supplementary material.

3. Results

Fig. 1 shows the flow diagram of the study selection. The literature searches yielded 1473 publications. After removing 362 duplicates between and within searches, 1111 unique publications were reviewed. Of these, 644 were excluded as irrelevant based on their title or because they were reviews, overviews, or meta-analyses or because they did not provide a comparison between computer and paper surveys. The abstracts of the 467 remaining publications were reviewed, and 222 of these were excluded as irrelevant, or because they were reviews, or because of not providing a comparison of social desirability scale scores between administration modes or computer vs. paper-and-pencil comparisons. The full texts of the remaining 245 publications were then reviewed.

Based on the inclusion criteria described in the methods section, 29 studies were eligible for inclusion, and 22 more studies were retrieved from the reference lists of these eligible studies and the meta-analyses mentioned in the method section, leading

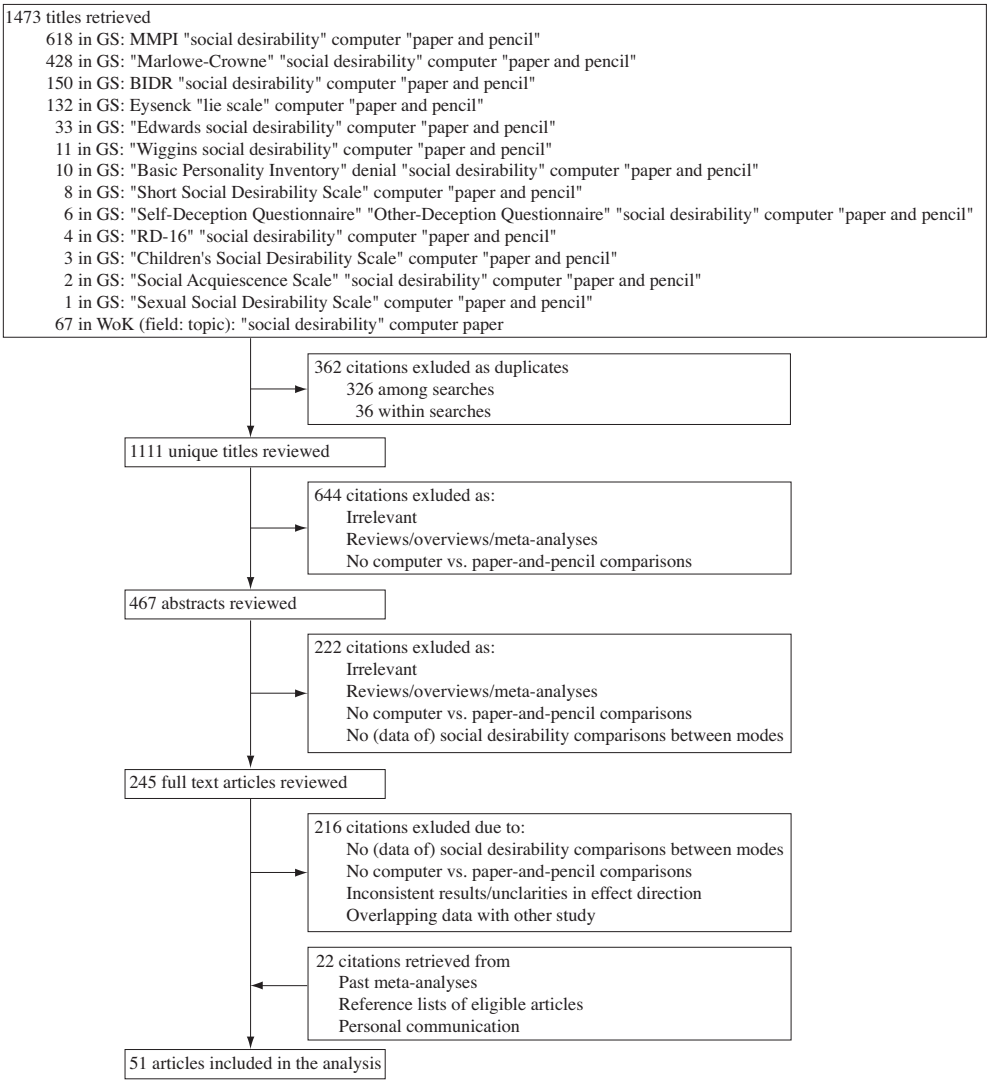

Fig. 1. Flow diagram of study selection. GS: Google Scholar. WoK: Web of Knowledge.

to a total of 51 studies being included in the meta-analysis. Of these, 15 were included in the meta-analyses of both Richman et al. (1999) and Dwight and Feigelson (2000), 4 were included only in Richman et al. (1999), 6 were included only in Dwight and Feigelson (2000), 19 were published after 1997 (year up to which the literature searches of both aforementioned meta-analyses were conducted), and 7 were published before 1997 but were not mentioned in the previous meta-analyses. The 51 studies included 62 independent samples. We were not able to retrieve the full text of three studies (Baydoun & Emperada, 1995; Dahl, 1992; Mitchell, 1993); for these studies, the sample sizes for the computer and paper surveys and the effect sizes (*ds*) reported in Dwight and Feigelson were used in the analysis. An overview of the studies included in the meta-analysis is provided in the Supplementary material.

3.1. Main effect and moderator analyses

Table 1 shows the results of the meta-analysis for the overall effect and for each of the moderator variables. The overall effect of administration mode on social desirability was nearly zero (fixed-effect $d = 0.00$ and random-effects $d = -0.01$). The effects yielded by the moderator analyses were also small; the summary effects ranged from −0.15 to 0.10 for the fixed-effect meta-analysis and from −0.17 to 0.09 for the random-effects meta-analysis, and only few were significantly different from zero (5 and 4 of the 30 effect sizes for the fixed-effect and random-effects meta-analyses, respectively).

The effect sizes were nearly zero for both the Internet and offline surveys (fixed-effect $d = -0.01$ and 0.01, respectively and random-effects $d = -0.03$ and 0.00, respectively). An additional exploratory meta-analysis comparing Internet surveys and paper surveys with anonymity as moderator also yielded no significant effect sizes (anonymous Internet vs. paper surveys: $d = -0.01$, 95% CIs: [−0.12, 0.10], $p = .822$ based on 13 effect sizes and identifiable Internet vs. paper surveys: $d = -0.17$, 95% CIs: [−0.63, 0.28], $p = .455$ based on 5 effect sizes). The effect size of the Internet surveys that were conducted from any place that the participants desired (i.e., the test setting was the participants' choice) vs. paper surveys was $d = 0.01$ (95% CIs: [−0.09, 0.11], $p = .876$ based on 9 effect sizes).

No significant effect of administration mode was found when analysing randomised studies alone ($d = 0.03$ based on random-effects analysis), whereas significantly lower social desirability was found for the computer surveys than for the paper surveys for the non-randomised studies ($d = -0.15$). The heterogeneity in the non-randomised studies was twice as large as the heterogeneity in the randomised studies (83.7% vs. 42.8%, respectively).

The summary effect across the studies that used sensitive questions indicated a significantly higher social desirability for the

Table 1
Meta-analysis results.

		Effect sizes	Participants	Fixed-effect		Random-effects		
				d [95% CI]	*p*-value	*d* [95% CI]	*p*-value	I^2 (%)
	All	62	16,700	0.00 [−0.03, 0.03]	.787	−0.01 [−0.07, 0.04]	.605	63.5
Survey characteristics								
Internet connectivity[a]	Internet	17	9057	−0.01 [−0.05, 0.04]	.696	−0.03 [−0.14, 0.07]	.539	78.4
	Offline[b]	46	7693	0.01 [−0.02, 0.05]	.464	0.00 [−0.07, 0.06]	.936	51.5
Identity	Anonymous	29	10,564	0.01 [−0.03, 0.05]	.755	−0.01 [−0.08, 0.07]	.897	64.8
	Identifiable	24	3014	−0.02 [−0.09, 0.05]	.619	−0.07 [−0.21, 0.07]	.317	71.8
	Unknown	17	3122	0.01 [−0.05, 0.07]	.647	0.01 [−0.07, 0.09]	.758	30.0
Test setting	Individual/Own choice[c]	34	8929	−0.02 [−0.06, 0.02]	.311	−0.03 [−0.10, 0.05]	.448	52.5
	Group	24	6285	0.02 [−0.03, 0.07]	.459	−0.01 [−0.11, 0.10]	.888	76.3
	Unknown	6	1486	0.06 [−0.02, 0.14]	.153	–	–	–
Skipping[a]	Allowed	31	11,362	0.00 [−0.03, 0.04]	.820	0.00 [−0.06, 0.05]	.873	47.2
	Not allowed	13	1297	0.10 [0.00, 0.20]	.052	0.04 [−0.20, 0.28]	.749	81.3
	Unknown	20	4121	−0.03 [−0.09, 0.03]	.303	−0.06 [−0.16, 0.04]	.234	57.0
Backtracking[a]	Allowed	30	9444	0.00 [−0.04, 0.04]	.976	0.00 [−0.07, 0.08]	.905	61.7
	Not allowed	15	2602	0.06 [−0.01, 0.13]	.093	0.04 [−0.10, 0.17]	.601	62.6
	Unknown	21	4898	−0.02 [−0.07, 0.04]	.509	−0.07 [−0.17, 0.03]	.174	62.7
Nature of questions	Sensitive	4	1639	0.10 [0.01, 0.20]	**.032**	–	–	–
	Nonsensitive	52	13,645	0.00 [−0.03, 0.03]	.908	−0.01 [−0.07, 0.05]	.799	64.8
	Unknown	6	1416	−0.09 [−0.20, 0.01]	.077	−0.15 [−0.33, 0.03]	.103	62.3
Scale types	BIDR-IM	21	5253	0.05 [0.00, 0.09]	.055	0.09 [0.01, 0.17]	**.037**	61.6
	BIDR-SD	16	4309	0.05 [0.00, 0.10]	**.039**	0.06 [0.00, 0.12]	**.033**	12.1
	MMPI-L	17	1563	0.02 [−0.05, 0.10]	.534	0.02 [−0.06, 0.09]	.689	3.3
	MMPI-K	17	1525	−0.00 [−0.08, 0.07]	.972	−0.08 [−0.22, 0.06]	.241	57.7
	ELS	7	638	0.05 [−0.07, 0.18]	.399	0.04 [−0.12, 0.20]	.616	35.8
	MCDS	13	4053	−0.15 [−0.22, −0.07]	**.000**	−0.17 [−0.29, −0.04]	**.010**	55.7
	Other	11	4041	0.02 (−0.04, 0.07)	.589	−0.07 (−0.21, 0.08)	.388	83.8
Study characteristics								
Sample type	Students	42	10,440	0.02 [−0.01, 0.06]	.188	0.01 [−0.06, 0.08]	.790	66.6
	Other	20	6260	−0.04 [−0.10, 0.01]	.120	−0.07 [−0.17, 0.02]	.142	52.9
Study design	Within	18	2893	0.03 [−0.02, 0.08]	.289	0.02 [−0.06, 0.09]	.614	35.0
	Between	44	13,807	−0.01 [−0.04, 0.03]	.679	−0.03 [−0.10, 0.04]	.432	69.2
Participant assignment	Randomised/Within-subjects	49	8915	0.04 [0.00, 0.07]	**.041**	0.03 [−0.03, 0.08]	.288	42.8
	Non-randomised	13	7785	−0.06 [−0.11, −0.01]	**.020**	−0.15 [−0.29, −0.01]	**.031**	83.7

Note: A positive effect size implies higher social desirability in the computer survey as compared to the paper survey. The total number of effect sizes is larger than 62 for some moderators because, in some of the studies, different portions of the sample were tested under different conditions (e.g., half of the sample was tested in a group setting, and the other half was tested in an individual setting).
p-values < .05 are in boldface.
[a] Moderator for computer administration only.
[b] Computers linked in a local network were also coded as offline.
[c] Own choice refers to Internet surveys that the participants completed at their convenience from any place of their choice.

computer surveys than for the paper surveys. Note that this summary effect relied on 4 studies only, and only one of these (Booth-Kewley, Larson, & Miyoshi, 2007) reported significantly higher social desirability for the computer surveys than for the paper surveys (specifically, for self-deception; for impression management no statistically significant administration effect was found).

The BIDR-IM and BIDR-SD measured significantly *lower* social desirability for paper surveys than for computer surveys (BIDR-IM: $d = 0.05$ and $d = 0.09$ for fixed-effect and random-effects, respectively; BIDR-SD: $d = 0.05$ and $d = 0.06$ for fixed-effect and random-effects, respectively), whereas the MCSD measured *higher* social desirability for paper surveys than for computer surveys ($d = -0.15$ and -0.17, for fixed-effect and random-effects, respectively). The MMPI-L and ELS showed near-zero effects.

The funnel plot of the effect sizes looked symmetric, with 49 of the 62 effect sizes falling within the fixed-effect 95% confidence interval of $d = 0.00$. Of the remaining 13 effects, 8 (2 of which corresponded to randomised studies) were lower than the lower 95% confidence bound, and 5 (3 of which corresponded to randomised studies) were higher than the upper 95% confidence bound (Fig. 2).

3.2. Longitudinal trends

The sample-size weighted Spearman correlation coefficient between the publication year and effect size was $\rho = .24$ ($p = .060$) based on 62 effect sizes, and $\rho = .18$ ($p = .204$) based on the 49 effect sizes from the randomised and within-subjects samples alone (Fig. 3; linear regression: $y = 0.006 * x - 11.604$, where x is the publication year and y is the effect size when all 62 effect sizes were included and $y = 0.003 * x - 6.924$ when only the 49 effect sizes corresponding to randomised and within-subjects samples were included). The newer studies had larger sample sizes than did the older studies ($\rho = .64$, $p = 2.3 * 10^{-8}$, 62 effect sizes), and relied more on Internet than on offline computer surveys ($\rho = .65$, $p = 8.1 * 10^{-9}$, 62 effect sizes).

Exploratory correlational analyses were conducted between the publication year and each of the moderators. The newer studies were more likely to have been conducted in a group setting (unweighted Spearman correlation: $\rho = .24$, $p = .077$, 56 effect sizes) and to have allowed backtracking in the computer version ($\rho = .38$, $p = .015$, 41 effect sizes). The newer studies also tended to be anonymous ($\rho = .27$, $p = .069$, 45 effect sizes) and randomised/within-subjects ($\rho = .17$, $p = .194$, 62 effect sizes).

3.3. Validity of the social desirability scales

To investigate whether the social desirability scales were valid and sensitive, for the studies in which the level of anonymity was a moderator variable, we conducted an additional meta-analysis with the level of anonymity as the main effect (i.e., social desirability in the surveys with anonymous vs. identifiable participants across both administration modes). To protect the quality of the dataset, non-randomised studies were excluded from this analysis. The results revealed that the anonymous surveys generated lower desirability than did the non-anonymous surveys ($d = -0.23$, 95% CIs [−0.34, −0.13], $p = 2.0 * 10^{-5}$ based on 7 effect sizes from 7 studies, and 1343 unique participants).

4. Discussion

This article provided an updated meta-analysis on social desirability between computer-based and paper-and-pencil surveys based on studies published over the last 45 years. Social desirability scores between the two administration modes were compared across 51 studies that included 62 independent samples and 16,700 unique participants —about 4 times greater than the 4205 unique participants included in the meta-analysis by Dwight and Feigelson (2000; based on 24 studies) and 7 times greater than the 2213 unique participants included in the meta-analysis by Richman et al. (1999; taking into account only the 16 studies with social desirability scales). We also conducted a number of subgroup analyses with all moderators included in the previous meta-analyses as well as for Internet connectivity, social desirability scale, study design, and randomisation, which have not previously been tested.

The overall effect of administration mode was very close to zero (fixed-effect $d = 0.00$, 95% CIs [−0.03, 0.03], and random-effects $d = -0.01$, 95% CIs [−0.07, 0.04]). Most of the effect sizes found in the moderator analyses were also not significantly different from zero. Specifically, the effect sizes were close to zero among the anonymous studies as well as the non-anonymous studies, the studies in which participants were tested individually as well as the studies in which the participants were tested in groups, the studies in which the surveys allowed skipping or backtracking

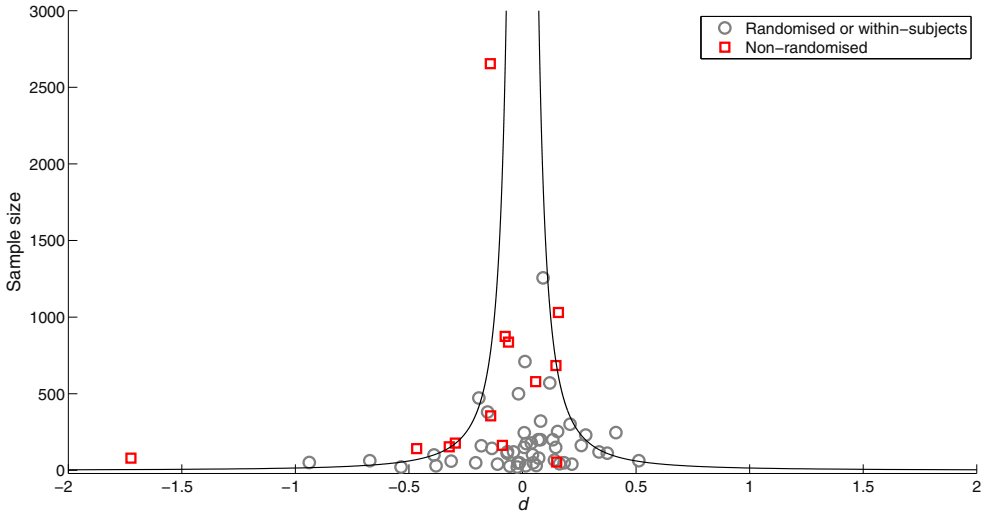

Fig. 2. Funnel plot of the effect sizes versus sample sizes. A positive effect size indicates higher social desirability in the computer survey as compared to the paper survey. A 95% confidence interval (CI) is drawn around $d = 0.00$, calculated as $CI = \pm 1.96 * V^{0.5}$, where V is the within-study variance calculated as: $V = 4/N$.

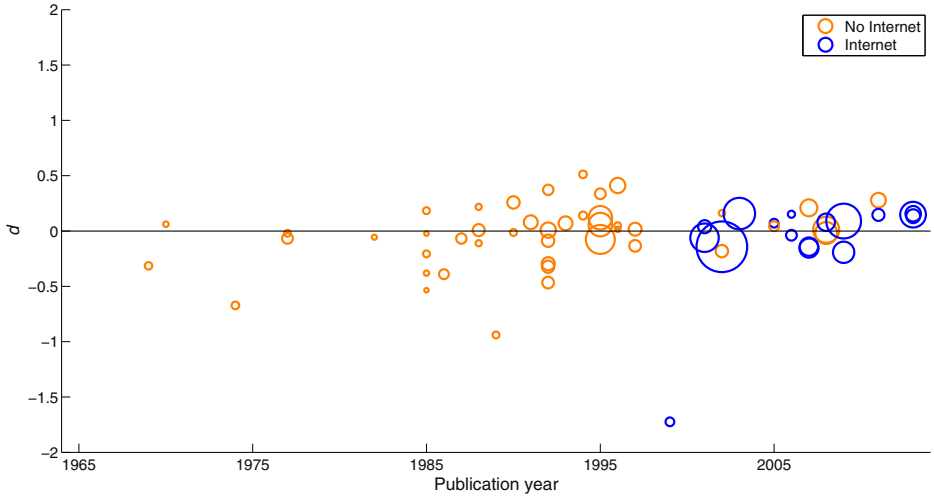

Fig. 3. Scatter plot of the 62 effect sizes included in the analysis as a function of publication year. A positive effect size indicates higher social desirability in the computer survey as compared to the paper survey. The area of the markers corresponds linearly to the sample size. The markers in the legend correspond to a sample size of 200. The horizontal line at $d = 0.00$ is presented for reference purposes.

and the studies in which skipping or backtracking was not allowed, the studies that utilised Internet surveys and the studies with offline computer surveys, the studies with student populations and the studies with nonstudent populations, and the studies that were conducted within-subjects as well as the studies that were conducted between-subjects. Hence, our results suggest that it does not matter whether one uses a computers survey or a paper-and-pencil survey, as far as social desirability is concerned.

Note that the near-zero effects were not due to low validity of the social desirability scales; the social desirability scales were valid and sensitive, as the anonymous surveys yielded significantly lower social desirability scores than did the non-anonymous surveys ($d = -0.23$). Richman et al. (1999) and Dwight and Feigelson (2000) also provided comparisons between computer surveys and face-to-face interviews and found that the former led to significantly lower social desirability for computers surveys compared to face-to-face interviews ($d = -0.19$ and -0.14, respectively).

Subgroup analyses showed lower social desirability for computer than for paper surveys when measured by MCSD and higher when measured by BIDR. This result is not theoretically interpretable, as MCSD and BIDR (the impression management dimension in particular) are supposed to represent similar social desirability constructs and are known to share positive intercorrelations of about $r = .4$ (Paulhus, 1984). It is possible that this observed contradictory effect is the result of sampling error or an unidentified moderation.

In line with the previous meta-analyses, we found a small positive correlation between publication year and effect size, with older studies reporting lower social desirability scores for computer surveys than for paper surveys ($\rho = .24$, $p = .060$). This result may be due to software and hardware improvements that have resulted in paper and computer surveys becoming increasingly similar (Richman et al., 1999). Another plausible explanation is that the early studies were of poorer quality in terms of randomisation, sample size, and similarity between computer and paper surveys. After excluding the non-randomised studies, the correlation between publication year and effect size decreased to $\rho = .18$ ($p = .204$). Also, publication year and sample size were positively correlated ($\rho = .64$, $p = 2.3 * 10^{-8}$). Moreover, older studies tended to not allow backtracking in the computer version ($\rho = .38$, $p = .015$). As can be seen from Table 1, the difference between the computer and paper surveys was larger when backtracking was not allowed than when it was allowed, and the computer surveys without backtracking possibility were associated with higher social desirability than the corresponding paper surveys. In summary, the observed correlation between publication year and effect size was small and not robust. The correlation between effect size and publication year may well be an artefact that is akin to the commonly known 'decline effect' (Schooler, 2011), with older studies being conducted with non-randomised samples, smaller sample sizes, and with surveys that exhibited intrinsic differences between the computer and paper versions, thereby producing larger effect sizes than did the recent studies.

4.1. Limitations

Of the 62 samples, 42 consisted of university students. Students are likely more familiar with computers than the average population and generally deviant from the general population (Henrich, Heine, & Norenzayan, 2010). Moreover, 46 samples were from the United States, 8 were from Europe, 4 were from Canada, 1 was from Australia, 1 was from Israel, 1 was from China, and 1 was from Japan. Social desirability responses between administration modes may have been different if the studies had been conducted with more diverse populations. For further discussion on cross-cultural differences in social desirability, see De Beuckelaer and Lievens (2009), Fang, Wen, and Prybutok (2013), and Furner (2011).

In some of the studies, social desirability scores were not estimated as the average of all responses but as the sum of all responses (e.g., the majority of the BIDR scales in the included studies). In studies in which item skipping was not allowed in the computer surveys, missing items in the paper surveys (in which it is in principle always possible to skip items) may have led to lower social desirability scores than those observed in the computer survey scores. In other words, the fact that the effect size between the paper and computer surveys with no skipping possibility tended to be larger ($d = 0.04$) than the effect size between the paper and computer surveys that allowed skipping ($d = 0.00$) may have partially been caused by missing items in the paper surveys.

Only a few of the studies included in our meta-analysis (Evan & Miller, 1969; Fang, Wen, & Prybutok, 2014; Kiesler & Sproull, 1986; Pettit, 2002; Rossiter, 2009; Whelan, 2008) mentioned acquiescence response bias (the tendency to agree with all statements) or extremity bias (selecting the first option provided; the so-called 'primacy effect'; Krosnick, 1991, p. 216). It is unclear whether

differential effects of acquiescence bias and extreme response bias between administration modes were at play. Sproull (1986) reported that responses to computer surveys are more extreme than responses to paper surveys, whereas others have found no significance differences in extreme response bias between the two administration modes (e.g., Helgeson & Ursic, 1989).

In most of the included studies, the surveys were conducted in a controlled manner (e.g., during a university class), and therefore, response rates were close to 100%. For the surveys that were completed in settings chosen by the participants (primarily Internet surveys and paper surveys that were communicated via the post), the response rates ranged from 22% to 68%. Attrition is a well-known source of bias, and the response rate differences between administration modes may have undermined the randomised nature of these samples. Some authors have argued that Web surveys are associated with lower response rates than mailed surveys (Crawford, Couper, & Lamias, 2001; Cronk & West, 2002), whereas others have reported no difference (Kaplowitz, Hadlock, & Levine, 2004). Of the studies included in the meta-analysis, Hancock and Flowers (2001), Rossiter (2009), Uriell and Dudley (2009), and Weigold, Weigold, and Russell (2013) reported no significant differences between the response rates of Internet and paper respondents. In the remainder of the included studies in which Internet surveys were completed in settings chosen by the participants (Campbell, Cumming, & Hughes, 2006; Fang et al., 2014; Pettit, 2002), the Internet samples were self-selected (e.g., by registering the website of the survey in search engines), which inhibits the estimation of a response rate. In any case, as mentioned in the results section, the effect size of Internet versus paper surveys conducted from the participants' places of preference was near zero.

5. Conclusions

This meta-analysis has shown that social desirability in offline, online, and paper surveys is practically the same. All effect sizes in the moderator analyses were also small, and the majority were not statistically significant. The few significant effects were neither robust nor theoretically interpretable and can be considered negligible. The longitudinal decrease in the social desirability difference between computer and paper surveys does not seem to be robust and may be a manifestation of the so-called 'decline effect' instead.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.chb.2014.04.005.

References

References marked with an asterisk indicate studies included in the meta analysis

∗Baydoun, R. B., & Emperada, T. K. (1995). Equivalence of computer vs. paper-and-pencil administered attitude instruments. *Paper presented at the conference of the society for industrial and organizational psychology, Orlando, FL.*

Ben-Ze'ev, A. (2003). Privacy, emotional closeness, and openness in cyberspace. *Computers in Human Behavior, 19*, 451–467.

Birnbaum, M. H. (Ed.). (2000). *Psychological experiments on the Internet*. Elsevier.

∗Booth-Kewley, S., Edwards, J. E., & Rosenfeld, P. (1992). Impression management, social desirability, and computer administration of attitude questionnaires: Does the computer make a difference? *Journal of Applied Psychology, 77*, 562–566.

∗Booth-Kewley, S., Larson, G. E., & Miyoshi, D. K. (2007). Social desirability effects on computerized and paper-and-pencil questionnaires. *Computers in Human Behavior, 23*, 463–477.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. West Sussex, UK: John Wiley & Sons.

Buchanan, T. (2000). Potential of the Internet for personality research. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 121–140). San Diego, CA: Academic Press.

Buchanan, T., & Smith, J. L. (1999). Using the Internet for psychological research: Personality testing on the World Wide Web. *British Journal of Psychology, 90*, 125–144.

∗Campbell, A. J., Cumming, S. R., & Hughes, I. (2006). Internet use by the socially fearful: Addiction or therapy? *CyberPsychology & Behavior, 9*, 69–81.

Cook, C., Heath, F., Thompson, R. L., & Thompson, B. (2001). Score reliability in Web-or-Internet-based surveys: Unnumbered graphic rating scales versus Likert-type scales. *Educational and Psychological Measurement, 61*, 697–706.

Crawford, S. D., Couper, M. P., & Lamias, M. J. (2001). Web surveys perceptions of burden. *Social Science Computer Review, 19*, 146–162.

Cronk, B. C., & West, J. L. (2002). Personality research on the Internet: A comparison of Web-based and traditional instruments in take-home and in-class settings. *Behavior Research Methods, Instruments, & Computers, 34*, 177–180.

∗Dahl, R. E. (1992). *Response bias, interactive effects of interview method, social desirability, evaluation apprehension of self-reported data*. Unpublished doctoral dissertation, Georgia State University.

∗Davis, R. N. (1999). Web-based administration of a personality questionnaire: Comparison with traditional methods. *Behavior Research Methods, Instruments, & Computers, 31*, 572–577.

Davis, C., & Cowles, M. (1989). Automated psychological testing: Method of administration, need for approval, and measures of anxiety. *Educational and Psychological Measurement, 49*, 311–320.

De Beuckelaer, A., & Lievens, F. (2009). Measurement equivalence of paper-and-pencil and Internet organisational surveys: A large scale examination in 16 countries. *Applied Psychology, 58*, 336–361.

De Winter, J. C. F., Zadpoor, A. A., & Dodou, D. (2014). The expansion of Google Scholar versus Web of Science: A longitudinal study. *Scientometrics, 98*, 1547–1565.

De Zwart, M., Lindsay, D., Henderson, M., & Phillips, M. (2011). *Teenagers, legal risks and social networking sites*. Faculty of Education: Monash University.

Dwight, S. A., & Feigelson, M. E. (2000). A quantitative review of the effect of computerized testing on the measurement of social desirability. *Educational and Psychological Measurement, 60*, 340–360.

∗Evan, W. M., & Miller, J. R. (1969). Differential effects on response bias of computer vs. conventional administration of a social science questionnaire: An exploratory methodological experiment. *Behavioral Science, 14*, 216–227.

Fang, J., Wen, C., & Prybutok, V. (2013). The equivalence of Internet versus paper-based surveys in IT/IS adoption research in collectivistic cultures: The impact of satisficing. *Behaviour & Information Technology, 32*, 480–490.

∗Fang, J., Wen, C., & Prybutok, V. (2014). An assessment of equivalence between paper and social media surveys: The role of social desirability and satisficing. *Computers in Human Behavior, 30*, 335–343.

∗Finegan, J. E., & Allen, N. J. (1995). Computerized and written questionnaires: Are they equivalent? *Computers in Human Behavior, 10*, 483–496.

Finger, M. S., & Ones, D. S. (1999). Psychometric equivalence of the computer and booklet forms of the MMPI: A meta-analysis. *Psychological Assessment, 11*, 58–66.

Fisher, R. J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of Consumer Research*, 303–315.

∗Fox, S., & Schwartz, D. (2002). Social desirability and controllability in computerized and paper-and-pencil personality questionnaires. *Computers in Human Behavior, 18*, 389–410.

Furner, C. P. (2011). Socially desirable reporting in international computer based data collection: A review and research agenda. *Journal of International Management Studies, 6*, 185–201.

Gehanno, J.-F., Rollin, L., & Darmoni, S. (2013). Is the coverage of Google Scholar enough to be used alone for systematic reviews. *BMC Medical Informatics and Decision Making, 13*, 7.

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist, 59*, 93–104.

Griffis, S. E., Goldsby, T. J., & Cooper, M. (2003). Web-based and mail surveys: A comparison of response, data, and cost. *Journal of Business Logistics, 24*, 237–258.

∗Hancock, D. R., & Flowers, C. P. (2001). Comparing social desirability responding on World Wide Web and paper-administered surveys. *Educational Technology Research and Development, 49*, 5–13.

Hedges, L. V., & Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological Methods, 3*, 486–504.

Helgeson, J. G., & Ursic, M. L. (1989). The decision process equivalency of electronic versus pencil-and-paper data collection methods. *Social Science Computer Review, 7*, 296–310.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*, 61–83.

Higgins, J. P. T., & Green, S. (2011). *Cochrane handbook for systematic reviews of interventions version 5.1. 0* [updated March 2011]. The Cochrane Collaboration.

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ. British Medical Journal, 327*, 557–560.

∗Honaker, L. M., Harrell, T. H., & Buffaloe, J. D. (1988). Equivalency of microtest computer MMPI administration for standard and special scales. *Computers in Human Behavior, 4*, 323–337.

Hood, K., Robling, M., Ingledew, D., Gillespie, D., Greene, G., Ivins, R., et al. (2012). Mode of data elicitation, acquisition and response to surveys: A systematic review. *Health Technology Assessment, 16*.

Hunter, J. E., & Schmidt, F. L. (2000). Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Assessment, 8*, 275–292.
*Joinson, A. (1999). Social desirability, anonymity, and Internet-based questionnaires. *Behavior Research Methods, Instruments, & Computers, 31*, 433–438.
Kaplowitz, M. D., Hadlock, T. D., & Levine, R. (2004). A comparison of web and mail survey response rates. *Public Opinion Quarterly, 68*, 94–101.
*Kiesler, S., & Sproull, L. S. (1986). Response effects in the electronic survey. *Public Opinion Quarterly, 50*, 402–413.
*King, W. C., & Miles, E. W. (1995). A quasi-experimental assessment of the effect of computerizing noncognitive paper-and-pencil measurements: A test of measurement equivalence. *Journal of Applied Psychology, 80*, 643–651.
Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*, 213–236.
Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: A literature review. *Quality & Quantity, 47*, 2025–2047.
*Lautenschlager, G. J., & Flaherty, V. L. (1990). Computer administration of questions: More desirable or more social desirability? *Journal of Applied Psychology, 75*, 310–314.
*Martin, C. L., & Nagao, D. H. (1989). Some effects of computerized interviewing on job applicant responses. *Journal of Applied Psychology, 74*, 72–80.
*Mitchell, D. (1993). *A multivariate analysis of the effects of gender and computer vs. paper/pencil modes of administration on survey results*. Unpublished doctoral dissertation, Louisiana Tech University.
Pasveer, K. A., & Ellard, J. H. (1998). The making of a personality inventory: Help from the WWW. *Behavior Research Methods, Instruments, & Computers, 30*, 309–313.
Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46*, 598–609.
Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17–59). New York: Academic Press.
*Pettit, F. A. (2002). A comparison of World-Wide Web and paper-and-pencil personality questionnaires. *Behavior Research Methods, Instruments, & Computers, 34*, 50–54.
Richman, W. L., Kiesler, S., Weisband, S., & Drasgow, F. (1999). A meta-analytic study of social desirability distortion in computer-administration questionnaires: Traditional questionnaires and interviews. *Journal of Applied Psychology, 84*, 754–775.
*Rosenfeld, P., Booth-Kewley, S., Edwards, J. E., & Thomas, M. D. (1996). Responses on computer surveys: Impression management, social desirability, and the big brother syndrome. *Computers in Human Behavior, 12*, 263–274.
*Rossiter, J. C. (2009). *A comparison of social desirability bias among four widely used methods of data collection as measured by the impression management subscale of the balance inventory of desirable responding*. Doctoral dissertation, Kent State University.
Schooler, J. (2011). Unpublished results hide the decline effect. *Nature, 470*, 437.
Shariff, S. Z., Bejaimal, S. A., Sontrop, J. M., Iansavichus, A. V., Haynes, R. B., Weir, M. A., et al. (2013). Retrieving clinical evidence: A comparison of PubMed and Google Scholar for quick clinical searches. *Journal of Medical Internet Research, 15*.
Smith, R. E. (1963). Examination by computer. *Behavioral Science, 8*, 76–79.
Smith, M. A., & Leigh, B. (1997). Virtual subjects: Using the Internet as an alternative source of subjects and research environment. *Behavior Research Methods, Instruments, & Computers, 29*, 496–505.
Sproull, L. S. (1986). Using electronic mail for data collection in organizational research. *Academy of Management Journal, 29*, 159–169.
Sproull, L., & Kiesler, S. (1986). Reducing social context cues: Electronic mail in organizational communication. *Management Science, 32*, 1492–1512.
Tan, L., & Grace, R. C. (2008). Social desirability and sexual offenders: A review. *Sexual Abuse: A Journal of Research and Treatment, 20*, 61–87.
*Toppins, J. D. (1985). *Automated administration of the MMPI with a psychiatric population, equivalence validity considerations*. Unpublished doctoral dissertation, University of Kansas.
Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133*, 859–883.
*Uriell, Z. A., & Dudley, C. M. (2009). Sensitive topics: Are there modal differences? *Computers in Human Behavior, 25*, 76–87.
*Weigold, A., Weigold, I. K., & Russell, E. J. (2013). Examination of the equivalence of self-report survey-based paper-and-pencil and internet data collection methods. *Psychological Methods, 18*, 53–70.
Weisband, S., & Kiesler, S. (1996). Self disclosure on computer forms: Meta-analysis and implications. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 3–10).
*Whelan, T. J. (2008). *Effects of survey modality and access controls on perceived anonymity and socially desirable responding*. North Carolina State University.
*White, D. M., Clements, C. B., & Fowler, R. D. (1985). A comparison of computer administration with standard administration of the MMPI. *Computers in Human Behavior, 1*, 153–162.
*Whitener, E. M., & Klein, H. J. (1995). Equivalence of computerized and traditional research methods: The roles of scanning, social environment, and social desirability. *Computers in Human Behavior, 11*, 65–75.
*Wilkerson, J. M., Nagao, D. H., & Martin, C. L. (2002). Socially desirable responding in computerized questionnaires: When questionnaire purpose matters more than the mode. *Journal of Applied Social Psychology, 32*, 544–559.

Further reading

*Bigot, O. (2011). *Passations papier/crayon et informatisées: quelle influence sur les stratégies d'autoprésentation?* Doctoral dissertation, Université Rennes 2.
*Biskin, B. H., & Kolotkin, R. L. (1977). Effects of computerized administration on scores on the Minnesota Multiphasic Personality Inventory. *Applied Psychological Measurement, 1*, 543–549.
*Clough, S. J. (2008). *Computerized versus paper-and-pencil assessment of socially desirable responding: Score congruence, completion time, and respondent preferences*. Doctoral dissertation, University of Iowa.
*Fastić-Pajk, I. (2005). *Socijalno poželjno odgovaranje u različitim modalitetima prezentacije čestica i njegova povezanost s osobinama ličnosti*. Ocjenski rad. Filozofski fakultet u Zagrebu, Department of Psychology.
*Hart, R. R., & Goldstein, M. A. (1986). Computer-assisted psychological assessment. *Computers in Human Services, 1*, 69–75.
*Hays, S., & McCallum, R. S. (2005). A comparison of the pencil-and-paper and computer-administered Minnesota Multiphasic Personality Inventory-Adolescent. *Psychology in the Schools, 42*, 605–613.
*Koson, D., Kitchen, C., Kochen, M., & Stodolosky, D. (1970). Psychological testing by computer: Effect on response bias. *Educational and Psychological Measurement, 30*, 803–810.
*Lambert, M. E., Andrews, R. H., Rylee, K., & Skinner, J. R. (1987). Equivalence of computerized and traditional MMPI administration with substance abusers. *Computers in Human Behavior, 3*, 139–143.
*Lushene, R. E., O'Neil, H. F., Jr., & Dunn, T. (1974). Equivalent validity of a completely computerized MMPI. *Journal of Personality Assessment, 38*, 353–361.
*Meade, A. W., Michels, L. C., & Lautenschlager, G. J. (2007). Are Internet and paper-and-pencil personality tests truly comparable? An experimental design measurement invariance study. *Organizational Research Methods, 10*, 322–345.
*Merten, T., & Siebert, K. (1997). A comparison of computerized and conventional administration of the EPQ-R and the CRS: Further data on the Merten and Ruch (1996) study. *Personality and Individual Differences, 22*, 283–286.
*Pinsoneault, T. B. (1996). Equivalency of computer-assisted and paper-and-pencil administered versions of the Minnesota Multiphasic Personality Inventory-2. *Computers in Human Behavior, 12*, 291–300.
*Potosky, D., & Bobko, P. (1997). Computer versus paper-and-pencil administration mode and response distortion in noncognitive selection tests. *Journal of Applied Psychology, 82*, 293–299.
*Rezmovic, V. (1977). The effects of computerizing experimentation on response bias variance. *Research Methods and Instrumentation, 9*, 144–147.
*Ridgway, J., MacCulloch, M. J., & Mills, H. E. (1982). Some experiences in administering a psychometric test with a light pen and microcomputer. *International Journal of Man-Machine Studies, 17*, 265–278.
*Risko, E. F., Quilty, L. C., & Oakman, J. M. (2006). Socially desirable responding on the web: Investigating the candor hypothesis. *Journal of Personality Assessment, 87*, 269–276.
*Roper, B. L., Ben-Porath, Y. S., & Butcher, J. N. (1995). Comparability and validity of computerized adaptive testing with the MMPI-2. *Journal of Personality Assessment, 65*, 358–371.
*Sanitioso, R., & Reynolds, J. H. (1992). Comparability of standard and computerized administration of two personality questionnaires. *Personality and Individual Differences, 13*, 899–907.
*Schuldberg, D. (1988). The MMPI is less sensitive to the automated testing format than it is to repeated testing: Item and scale effects. *Computers in Human Behavior, 4*, 285–298.
*Sloan, H. J. (1991). *Respondent self-completion on computers in marketing research: Computers compared with alternative collection methods*. Doctoral dissertation, The Ohio State University.
*Sukigara, M. (1996). Equivalence between computer and booklet administrations of the new Japanese version of the MMPI. *Educational and Psychological Measurement, 56*, 570–584.
*Watson, C. G., Manifold, V., Klett, W. G., Brown, J., Thomas, D., & Anderson, D. (1990). Comparability of computer-and booklet-administered Minnesota Multiphasic Personality Inventories among primarily chemically dependent patients. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 2*, 276–280.
*Wicherts, J. M. (2002). *Meetequivalentie tussen pen & papier en computerafname van vragenlijsten: Alexithymie, sociale wenselijkheid en angst* [Measurement equivalence of paper-and pencil and computerized administration of questionnaires: Alexithymia, social desirability, and anxiety]. Unpublished master's thesis, University of Amsterdam.
*Wicherts, J. M., & Bakker, M. (2012). Publish (your data) or (let the data) perish! Why not publish your data too? *Intelligence, 40*, 73–76.