Erschienen in: Computers in Human Behavior ; 49 (2015). - S. 608-614
https://dx.doi.org/10.1016/j.chb.2015.03.044

Psychometric properties of measurements obtained with the Marlowe–Crowne Social Desirability Scale in an Icelandic probability based Internet sample

Vaka Vésteinsdóttir [a,*], Ulf-Dietrich Reips [b], Adam Joinson [c], Fanney Thorsdottir [d]

[a] *University of Iceland, Department of Psychology, Aragata 14, 101 Reykjavik, Iceland*
[b] *University of Konstanz, Department of Psychology, Fach 31, 78457 Konstanz, Germany*
[c] *Behavioural Research Lab, Faculty of Business and Law, University of the West of England, Bristol, United Kingdom*
[d] *University of Iceland, Department of Psychology, Sturlugata 3, 101 Reykjavík, Iceland*

ARTICLE INFO

Keywords:
Marlowe–Crowne Social Desirability Scale
Internet administration
Psychometric properties
Short forms
Unidimensionality

ABSTRACT

Internet surveys have become a very popular research tool. Relatively little attention has, however, been devoted to the possible changes in psychometric properties when measurements are obtained with Internet surveys. The Marlowe–Crowne Social Desirability Scale (MCSDS) is the most widely used instrument for measuring the tendency to respond in a socially desirable way and is often used to validate other measures. The purpose of the current research is to evaluate the dimensionality and reliability of measurements obtained with the MCSDS and short forms of the scale in an Internet sample of the general public in Iceland. An e-mail invitation was sent to a sample of 1200 panel members drawn from the Social Science Research Institute (SSRI) probability based panel, of those 536 participants completed all items on the MCSDS. Reliability estimates were in line with results from previous studies ($\alpha = .81$ for the MCSDS data and α ranging from .59 to .75 for short forms). Using confirmatory factor analysis, a good fit was obtained for a one-factor model of measurements obtained with the MCSDS and its short forms (apart from significant chi square values in all cases but one), which generally supports the assumption of unidimensionality.

1. Introduction

Measures of social desirability (SD) have been used to validate psychological measures in paper and pencil format for over half a century. The Marlowe-Crowne Social Desirability Scale (MCSDS; Crowne & Marlowe, 1960) is the most widely used instrument for measuring social desirability response style (SDRS), and the psychometric properties of the paper and pencil format of the scale have been evaluated in numerous studies (Barger, 2002; Beretvas, Meyers, & Leite, 2002; Fischer & Fick, 1993; Loo & Loewen, 2004; Loo & Thorpe, 2000; Sârbescu, Costea, & Rusu, 2012; Ventimiglia & MacDonald, 2012). It has, however, been shown that psychometric properties observed in a paper and pencil mode of administration are not necessarily retained when transferred to Internet administered measures (Buchanan, 2002). Because of such mode effects on measurements, the American Educational Research Association (AERA), American Psychological Association (APA) and National Council on Measurement in Education (NCME) have advised test users to provide evidence that changes in mode of administration do not compromise the validity and reliability of the measurements obtained (U.S., 1999). Given the number of studies using the Internet to collect data, it is imperative that a validated and reliable measure of SDRS is available to researchers. Otherwise, researchers risk having their findings compromised by being unable to control for, or measure, socially desirable responding in their samples. However, to our knowledge, no published research has evaluated the validity and reliability of the Internet version of the MCSDS. If the practice of using the MCSDS to validate psychological measures is to be continued over the Internet, it is important to have a formal Internet version of the scale and to know that it produces psychometrically sound data. The purpose of the present study is to evaluate the validity and reliability of measurements obtained with the Internet version of the MCSDS in a representative sample of the Icelandic population. The focus is both on the psychometric properties of the original 33 item scale, and also of the various short forms that been proposed (Ballard, 1992; Ramanaiah, Schill, & Leung, 1977; Reynolds, 1982; Strahan & Gerbasi, 1972).

* Corresponding author. Tel.: +354 8472315.
E-mail addresses: vakav@hi.is (V. Vésteinsdóttir), reips@uni-konstanz.de (U.-D. Reips), adam.joinson@uwe.ac.uk (A. Joinson), fanneyt@hi.is (F. Thorsdottir).

1.1. Background

The use of Internet surveys has grown rapidly over the past decade, with researchers in various areas of the academia, government and the private sector making increasing use of Internet surveys to collect self-report data. The type of information (e.g. people's thoughts, behavior and experience) gathered in self reports is often very difficult to obtain or even inaccessible through other means of data collection. One of the main concerns with self-reports is the reduced accuracy of responses due to SDRS, a tendency to systematically respond in a manner likely to be approved of by others (Nunnally & Bernstein, 1994; Paulhus, 1991). This produces an association between the likelihood of a response and the desirability of that response which results in added systematic error variance in target measurements. It follows that if a question is neutral in regards to SD, responses to that question are not expected to be influenced by this tendency. Thus, in order to identify systematic error variance due to SDRS in psychological measurements, the correlation between an SDRS measure and the instrument of interest is calculated. Also, the responses to an SDRS measure and another scale can be concurrently factor analyzed to see whether the two scales load on distinct factors (Beretvas et al., 2002). For this to be viable it is essential to obtain valid and reliable measurements of SDRS.

A number of scales have been developed to measure SDRS (see Paulhus, 1991, for an overview) but the most commonly used scale is the Marlowe–Crowne Social Desirability Scale (MCSDS) (Beretvas et al., 2002).

1.2. The Marlowe–Crowne Social Desirability Scale

The MCSDS, designed by Crowne and Marlowe (1960), contains 33 items that were chosen based on their judged SD and item analysis. The items are thought to reflect, "behaviors which are culturally sanctioned and approved but which are improbable of occurrence" (Crowne & Marlowe, 1960, pp. 350). Answers in the keyed direction can, thus, be taken to indicate that the respondent is exaggerating desirable behavior i.e. showing SDRS. What all the items should have in common is purely their relation to the construct of SDRS. In other words, the MCSDS should be unidimensional (see Hattie, 1985 for a detailed discussion of unidimensionality).

The dimensionality of the paper and pencil format of the MCSDS has been addressed in a number of factor analytic studies. If all the items on MCSDS measure the same construct they would be expected to have a clear one factor structure. This has, however, not always been the case.

Confirmatory factor analyses have been conducted on the original English version of the MCSDS (in the USA: Barger, 2002; Leite & Beretvas, 2005, in Canada: Fischer & Fick, 1993; Loo & Loewen, 2004; Loo & Thorpe, 2000; Ventimiglia & MacDonald, 2012) and on a Romanian version of the scale (Sârbescu et al., 2012). None of these analyses have shown strong support for a one factor model of the full 33 item version. Very few studies have published results for factor loadings of individual items, but those that have suggest that some items need to be revised or removed from the scale due to low factor loadings (Reynolds, 1982; Ventimiglia & MacDonald, 2012). Reliability analyses have shown mixed results because internal consistency reliability estimates for measurements with the full 33 item scale have ranged from .72 (Loo & Thorpe, 2000) to .96 (Fischer & Fick, 1993).

Because of concerns about the scales dimensionality and item content, several short forms of the MCSDS have been developed, (Ballard, 1992; Hays, Hayashi, & Stewart, 1989; Reynolds, 1982; Rudmin, 1999; Sârbescu et al., 2012; Strahan & Gerbasi, 1972). The most frequently cited short forms in psychometric evaluation studies of MCSDS data are those developed by Strahan and Gerbasi (1972), Ramanaiah et al. (1977), Reynolds (1982) and Ballard (1992) and therefore we will limit our discussion to those short forms.

Strahan and Gerbasi (1972), Reynolds (1982) and Ballard (1992) all used principal component analyses to develop short forms. Strahan and Gerbasi (1972) reduce the number of items based on factor loadings on the first component. Factor loadings were used to form two distinct ten item short forms of the MCSDS, each with an equal number of attribution and denial items, and a 20 item short form comprised of the combined ten item scales (the ten item short forms will be referred to as X1 and X2 and the 20 item version as XX). Reynolds (1982), however, only selected items with factor loadings above .40 on the first component to form an 11 item short form, which he called 'short Form A'. In addition to Form A, he developed a 12 and a 13 item short form (Form B and C respectively) by adding items to the initial 11 item short form, based on item total correlations. Using the same method as Reynolds, Ballard (1992) developed three short forms of the MCSDS: Scale 1 (11 items), Scale 2 (12 items) and Scale 3 (13 items). Ballard also developed a 13 item short form (Composite scale) comprised of items that had been selected for one or more short forms in at least two of the three previous short form development studies (i.e. Ballard, 1992; Strahan & Gerbasi, 1972; Reynolds, 1982) (see Ballard, 1992, for an overview of which items belong to each short form).

Confirmatory factor analysis has previously been used to assess the unidimensionality assumption of the short forms. A better fit has been obtained for some short forms compared to the full 33 item scale but the studies have not agreed on the most valid short form, with Fischer and Fick (1993) recommending Strahan and Gerbasi's (1972) short forms X1 and X2, Loo and Thorpe (2000) obtaining the best fit for Reynolds' (1982) Form A and B, and Loo and Loewen (2004) recommending Ballard's Scale 1 or Composite scale. Lower reliability estimates have been obtained for measurements with the short forms than with the full scale, as can be expected. For the short forms, the highest internal consistency reliability estimates have been found for data obtained with the XX form but the lowest for data obtained with the X1 and X2 forms (Ballard, 1992; Barger, 2002; Fischer & Fick, 1993; Loo & Loewen, 2004; Loo & Thorpe, 2000; Reynolds, 1982; Strahan & Gerbasi, 1972).

Taken together, studies of the paper and pencil version of MCSDS suggest a need for improvement of the full 33 item scale. It is not clear, however, which of the short forms that have been proposed produces the most valid and reliable data. It is also important to note that previous studies have all based their findings on nonprobability samples – and almost exclusively on student samples (which causes important variables, such as participants age (Kaufmann & Reips, 2008), to be sampled over a very narrow range). It is therefore questionable to draw strong conclusions regarding the quality of the data obtained with the paper and pencil versions of the MCSDS merely on the basis of these studies.

1.3. Internet version of the MCSDS

With the growing number of studies using the Internet as the mode of administration, researchers have become increasingly concerned with the validation of measures for use online (Buchanan & Smith, 1999; for an example of an online validation study see Hearn, Ceschi, Brillon, Fürst, & Van der Linden, 2012). Thus, to further validation research on the Internet, especially in large-scale surveys, it is important to have an Internet version of the MCSDS, and to know that it produces valid and reliable data. Unfortunately, a psychometrically sound Internet version of the MCSDS cannot be presented based on the studies described above

for three main reasons. First, the studies did not agree in their results on the validity and reliability of various forms of the MCSDS. Secondly, due to methodological limitations, the results of these studies need to be viewed with caution. Finally, since there are reasons to believe that transferring a scale to an Internet mode may change the instrument (Buchanan, 2001; Buchanan, Johnson, & Goldberg, 2005), the practice of simply adapting a paper and pencil instrument for the Internet mode and assuming that the measurements are equivalent to the original mode is not recommended (Buchanan, 2002; Coles, Cook, & Blake, 2007; Hewson & Charlton, 2005; Noyes & Garland, 2008). This is especially relevant for the MCSDS since responses to SDRS scales have been shown to differ between modes of administration (Dodou & de Winter, 2014; Joinson, 1999; Kaufmann & Reips, 2008; Lautenschlager & Flaherty, 1990; Martin & Nagao, 1989). It is therefore important to evaluate the psychometric properties of the Internet version of the MCSDS before the scale is used in validation studies on the Internet.

No published study to date has explored the quality of data obtained with an Internet version of the MCSDS. A failure to establish adequate psychometric properties in a representative Internet sample would suggest that the scale is not suitable for use in online studies. Thus, the main purpose of the present research is to evaluate the unidimensionality, factor loadings and reliability of measurements obtained with the full MCSDS and various short forms in a probability based Internet sample of the general public in Iceland, which has the highest Internet penetration rates in Europe (Hagtíðindi, 2015). The aim is also to compare the findings to previous results obtained in paper and pencil mode.

The reminder of this paper is structured as follows: Section 2 describes the research method by providing information on the sample and technical information on the MCSDS as well a discussion of the sampling technique and statistical techniques used. In Section 3 we present and discuss the results and compare the results to previous studies. Finally, Section 4 concludes with a summary of the results, the limitations of the study, and suggestions for future research.

2. Method

2.1. Study sample and procedures

To maximize the representativeness of the Internet sample we used a probability based Internet panel. The advantages of surveys on probability pre-recruited panels of Internet users is that the sample is obtained from a sampling frame which covers the target population to a satisfactory degree (see Lozar Manfreda & Vehovar, 2008, for a taxonomy and terminology of web surveys). Although this approach is limited to the proportion of the population that has Internet access, Internet penetration rates are very high in Iceland and higher than in any other European country (Hagtíðindi, 2015). At the time of the Internet survey administration, 95% of people aged 16–74 were regular users of the Internet and 97% of Icelandic households had Internet access (Hagtíðindi, 2014). A probability panel of the general population in which those who do not have Internet access are given the equipment needed to participate in Internet surveys, would thus add little to the population of Internet users, and no such panel exists in Iceland.

A stratified random sample of 1200 panel members was drawn from the Social Science Research Institute (SSRI) Internet panel. The SSRI Internet panel consists of people aged 18 years and older, randomly drawn from Registers Iceland (Iceland's national registry), who consented over the telephone to be contacted for Internet surveys. The SSRI estimates that approximately 60–65% of those who respond to telephone surveys consent to participating in Internet surveys or a little over 40% of the original sample. The SSRI panel is under constant revision to ensure that its members are representative of the Icelandic nation and care is given to rebalancing when needed. Given the very high Internet penetration in Iceland and the probabilistic nature of the recruitment, the online panel is representative of the population in Iceland. The sample was stratified by gender, age and residence to reflect the composition of the Icelandic population as closely as possible.

The survey, containing all 33 items of the MCSDS in Icelandic, was sent out in July 2013 via e-mail, to 1200 potential participants. Duration of data collection was 2 weeks, with three reminders being sent out within the first 12 days after the original invitation was sent. Six hundred and thirty-nine participants gave a response to at least one of the items on the MCSDS, with 536 participants (44.7% of the original sample) completing all items on the MCSDS. Their responses were used for psychometric evaluation of data obtained with the translated MCSDS. The final sample comprised 272 women (50.7%) and 264 men (49.3%), aged between 20 and 81 years (mean age being 49 years), with educational levels varying from elementary school education to a post-graduate university degree.

2.2. Instrument

The *Marlowe–Crowne Social Desirability Scale* (Crowne & Marlowe, 1960) consists of 33 true/false items, 18 keyed in the true direction (attribution items) and 15 in the false direction (denial items). Responses in the keyed direction are coded as one and responses not in the keyed direction as zero. The highest possible score on the MCSDS is therefore 33 and the lowest is zero, with higher scores indicating more SD in responses.

An existing translation of the MCSDS was revised. The original translation was done by two researchers in the field of psychology. Each researcher translated the MCSDS from English into Icelandic, the translations were then compared and any differences in translation were discussed until an agreement was reached (see in Ragnarsson, 2011). This translation method has several limitations (see Behling & Law, 2000 for more on translations of instruments). However, since a pilot study of the psychometric properties of the MCSDS in Internet format had shown the original translation to be adequate (Vésteinsdóttir, Reips, Joinson, & Thorsdottir, 2012), it was decided to revise the existing translation instead of retranslating the MCSDS.

The revision of the MCSDS relied on intensive individual interviews (Karlsdóttir, 2013) and item response theory analysis (Vésteinsdóttir, Reips, Joinson, & Thorsdottir, 2013). All items with poor psychometric properties were revised in accordance to the interviewees' understanding of the items. Items with multiple possibilities of interpretation and/or indications of information lost in the translation process were also revised. The aim of the revision process was solely to improve the translation of the MCSDS not to change the content of the items.

2.3. Analysis

Responses from participants who omitted one or more items on the MCSDS were excluded from the analysis. Internal consistency was estimated with Cronbach's alpha coefficient. The criterion for a good reliability was set at $\alpha = .80$ or higher (DeVellis, 2003).

Confirmatory factor analysis (CFA) was used to evaluate the unidimensionality assumption of the MCSDS. The MCSDS is comprised of multiple indicators of SDRS. A fundamental assumption of validity and reliability testing is that the items measure the same construct. Measurements obtained using the MCSDS can be thought of as unidimensional if just one latent variable – SDRS – underlies the data. The implication of this is that if the items only share SDRS as the latent variable, then a keyed response to an item

can be taken as evidence of SDRS (see Hattie, 1985 for a more detailed explanation).

Unidimensionality of measurements obtained with the MCSDS can best be assessed with CFA. The advantage of CFA to other methods is its ability to test both the internal and external consistency criteria of unidimensionality, which means that: What each item pair has in common is equal to what the pair has in common with the observed factor (internal consistency) and that the observed correlations are consistent with the estimated parameters of the hypothesized factor structure, represented by a reproduced correlation matrix (external consistency). If the criteria for internal and external consistency are sufficiently met, the difference between the observed correlation matrix and the reproduced correlation matrix (indicated by goodness-of-fit indices) will be small and unidimensionality can be assumed (see Gerbing & Anderson, 1988 for a more detailed explanation). If, however, the assumption of unidimensionality does not hold for measurements obtained with the MCSDS, responses to items cannot be taken to reflect variations in SDRS, which renders all calculations (e.g. sum scores) and assumptions based on the MCSDS invalid.

Table 1
Descriptive statistics for measures obtained with the MCSDS in full length.

	n	Mode	Sample	Language	*M*	SD
Crowne and Marlowe (1960)	120	p&p	Students	English	13.7	5.8
Crowne and Marlowe (1964)	300	p&p	Students	English	15.5	4.4
Strahan and Gerbasi (1972)	272	p&p	Students	English	14.5	5.4
Reynolds (1982)	608	p&p	Students	English	15.0	5.9
Ballard (1992)	399	p&p	Students	English	14.6	5.1
Loo and Thorpe (2000)	232	p&p	Students	English	15.8	4.9
Barger (2002)	466	p&p	Students	English	17.2	5.0
	401	p&p	Students	English	16.2	5.0
Collazo (2005)	348	p&p	Students	Spanish	17.3	5.5
Sârbescu, Costea and Rusu (2012)	215	p&p	Students, friends and relatives	Romanian	15.5	5.1
Internet sample	536	Internet	Internet panel	Icelandic	17.7	5.6

Note. p&p = paper and pencil.

A one factor model was set up in LISREL 8.72 using DWLS (diagonally weighted least squares) with asymptotic variance and covariance matrices as input matrices. Factor loadings and random errors were estimated freely but the factor variance was set to one to enable identification of the model. Error variances were not allowed to co-vary.

It is usually recommended to use more than one fit index when evaluating the fit of a CFA model. A chi-square fit statistic and three fit indices were used to evaluate the fit of the models estimated in the present study: Satorra Bentler χ^2, CFI (comparative fit index), NNFI (nonnormed fit index) and RMSEA (root mean square error of approximation). CFI and NNFI indicate whether the model fits the data better than a model that does not assume any relation between measured variables, while adjusting for sample size. Both of these indices range from zero to one. The cutoff criterion for an adequate fit was set a .96 for CFI and .95 for NNFI. The RMSEA index indicates the degree of discrepancy between the model and the data. Values of RMSEA range from zero to one, with a lower value signifying a better fit. The cutoff criterion for an adequate fit for RMSEA was set at .05. Fit indices and their cutoff criteria were chosen based on how they have performed when DWLS was used in medium-large samples (Yu, 2002).

3. Results and discussion

3.1. Descriptive statistics

Means and standard deviations of measurements obtained with the MCSDS in full length are shown in Table 1 for the Icelandic Internet sample and comparison samples. The mean of 17.7 for the Icelandic Internet data is somewhat higher than the means obtained in prior studies, which have ranged from 13.7 to 17.3, with higher means more often seen in more recent studies. However, when comparing means from the studies listed in Table 1, it should be kept in mind that psychometric invariance between measurements obtained in different languages and/or cultures has not been established, nor can it be taken for granted that mean scores obtained in student samples are psychometrically compatible to other samples. It would therefore be premature to infer something about differences in SDRS between these groups.

Mean scores did not significantly differ between men ($M = 17.8$, $SD = 5.8$) and women ($M = 17.6$, $SD = 5.4$) in the Internet sample, $t(534) = .421$, $p > 0.05$, and in all but one of the above studies no gender differences were found in means. The only study showing

Table 2
Internal reliability estimates for measures obtained with the MCSDS in full length and short versions.

	N of Items	Crowne and Marlowe (1960)	Strahan and Gerbasi (1972)	Reynolds (1982)	Ballard (1992)	Fischer and Fick (1993)	Loo and Thorpe (2000)	Loo and Loewen (2004)	Barger (2002)	Internet sample (2013)
Full scale	33	.88	.73–.87	.82	.75	.96	.72	.75	.73–.74	.81
X1[a]	10		.59–.70	.63	.50	.88	.52	.51	.53–.56	.59
X2[a]	10		.49–.75	.66	.54	.88	.42	.46	.50–.56	.59
XX[a]	20		.73–.87	.79	.71	.94	.68	.68	.69–.72	.75
Form A[b]	11			.74	.64	.86	.59	.61	.60–.63	.70
Form B[b]	12			.75	.67	.88	.61	.66	.62–.64	.71
Form C[b]	13			.76	.68	.89	.62	.66	.62–.64	.72
Attribution[c]	18					.88	.56	.62	.59–.61	.67
Denial[c]	15					.88	.63	.65	.64–.65	.75
Scale 1[d]	11				.69		.60	.60		.69
Scale 2[d]	12				.69		.60	.64		.68
Scale 3[d]	13				.70		.61	.62		.71
Composite scale[d]	13				.70		.62	.64		.73

[a] Strahan and Gerbasi (1972).
[b] Reynolds (1982).
[c] Ramanaiah et al. (1977).
[d] Ballard (1992).

Table 3
Confirmatory factor analysis of measures obtained with the Internet version of the MCSDS.

	N of Items	S-B χ^2	*p*-Value	df	CFI	NNFI	RMSEA	RMSEA 90% CI
Full scale	33	965.37	<0.001	495	0.97	0.96	0.042	0.038; 0.046
X1[a]	10	37.45	0.36	35	1.00	1.00	0.011	0.000; 0.034
X2[a]	10	63.09	0.0025	35	0.98	0.98	0.039	0.023; 0.054
XX[a]	20	312.30	<0.001	170	0.98	0.97	0.040	0.033; 0.046
Form A[b]	11	111.89	<0.001	44	0.98	0.97	0.054	0.041; 0.066
Form B[b]	12	129.84	<0.001	54	0.98	0.97	0.051	0.040; 0.063
Form C[b]	13	154.45	<0.001	65	0.98	0.97	0.051	0.040; 0.061
Attribution[c]	18	234.67	<0.001	135	0.97	0.97	0.037	0.029; 0.045
Denial[c]	15	250.66	<0.001	90	0.97	0.96	0.058	0.049; 0.066
Scale 1[d]	11	92.21	<0.001	44	0.98	0.98	0.045	0.032; 0.058
Scale 2[d]	12	108.69	<0.001	54	0.98	0.97	0.043	0.032; 0.055
Scale 3[d]	13	124.42	<0.001	65	0.98	0.98	0.041	0.030; 0.052
Composite scale[d]	13	143.42	<0.001	65	0.98	0.97	0.047	0.037; 0.058

[a] Strahan and Gerbasi (1972).
[b] Reynolds (1982).
[c] Ramanaiah et al. (1977).
[d] Ballard (1992).

a gender difference was Barger (2002), with lower mean score for men than for women in one of the two samples.

3.2. Reliability

Table 2 shows the internal reliability estimates from previous studies and for the Internet sample. Reliability estimates were made for measurements obtained with the full scale and the short forms developed by Strahan and Gerbasi (1972), Reynolds (1982) and Ballard (1992), as well as the Denial and Attribution scales (Ramanaiah et al., 1977). Internal reliability for the full Internet scale data was good ($\alpha = .81$), with lower reliability estimates for the short form data (as can be expected with internal reliability being in part a function of the number of items in a scale), ranging from .59 to .75. The reliability estimates for the Internet scale data were, for the most part, in line with those shown in previous studies, although somewhat higher than estimates from the most recent studies (Barger, 2002; Loo & Thorpe, 2000). Noticeably different from other studies are the reliability estimates obtained by Fischer and Fick (1993), whose estimates are substantially higher than what others have shown before and since.

3.3. Confirmatory factor analysis

Information on fit for a one factor model of the full Internet scale and short form data are presented in Table 3. For the full 33 item scale, all the fit indices, except the chi-square test indicated an excellent fit of the unidimensional model, in sharp contrast to previous studies, which did not support the unidimensionality assumption and obtained a poor fit for the one factor model in most cases (Barger, 2002; Fischer & Fick, 1993; Loo & Loewen, 2004; Loo & Thorpe, 2000; Sârbescu et al., 2012; Ventimiglia & MacDonald, 2012).

Factor loadings of the full 33 item MCSDS from Reynolds (1982), Ventimiglia and MacDonald (2012) and the Internet sample are listed in Table 4. Most of the items had acceptable factor loadings in the Internet sample. The factor loadings ranged from .22 to .69 with an average factor loading of .45. Six of the items had factor loadings below .30 on the full 33 item scale.

Taken together, the CFA results for the full 33-item scale support the unidimensionality assumption. The main cause for concern about the scale is the apparently low factor loadings of some of the items. Low factor loadings complicate the interpretation of what is being measured by the scale. They either indicate that the items measure SDRS insufficiently or that the items are representative of SDRS, but measure different aspects of a broader and more complex construct (see Nunnally, 1978 for discussion). Items 1, 7, 17, 18, 24 and 29, all had factor loadings below .30 on the full scale. These items also have low factor loadings in Reynolds (1982) and Ventimiglia and MacDonald (2012), which are the only two studies that report factor loadings for all items. Factor loadings of these items are between .15 and .34 in Reynolds study and between .07 and .32 in Ventimiglia and MacDonald's study. Three of the items, items 7, 18 and 29, have factor loadings below .30 in all three studies. Items with low factor loadings could be dealt with as insufficient and simply deleted from the scale, which is in essence what has been done in short form development.

Table 4
Factor loadings obtained with the MCSDS in Reynolds (1982), Ventimiglia and MacDonald (2012) and the Internet sample.

Item number	Reynolds 1982	Ventimiglia and MacDonald 2012	Internet sample 2013
1.	.31	.17	.22
2.	.33	.33	.31
[a]3.	.40	.26	.43
4.	.38	.41	.47
[a]5.	.33	.18	.40
[a]6.	.54	.47	.69
7.	.25	.18	.29
8.	.32	.30	.34
[a]9.	.22	.32	.54
[a]10.	.39	.28	.48
[a]11.	.35	.32	.50
[a]12.	.39	.35	.39
13.	.40	.38	.54
[a]14.	.32	.26	.54
[a]15.	.49	.48	.62
16.	.46	.46	.54
17.	.30	.32	.28
18.	.15	.16	.25
[a]19.	.48	.50	.62
20.	.34	.41	.46
21.	.44	.43	.46
[a]22.	.23	.31	.37
[a]23.	.38	.30	.57
24.	.34	.32	.27
25.	.31	.21	.30
26.	.41	.37	.43
27.	.21	.27	.42
[a]28.	.53	.34	.66
29.	.22	.07	.28
[a]30.	.50	.42	.55
31.	.33	.30	.55
[a]32.	.37	.26	.46
33.	.42	.49	.52

[a] Items keyed in the opposite direction.

For the short forms, the fit of a one-factor model was always above the cut-off criterion for an adequate fit in terms of CFI and NNFI. RMSEA indicated a good fit (<.05) for all short form data except Reynolds (1982) (Form A, B and C) and the Denial scale. Reynolds Form A, B and C were slightly above the cut-off criterion for a good fit and the Denial scale received the highest RMSEA index of all forms tested. It is however worth noting that the upper limit of the 90% confidence interval for RMSEA exceeded the cut off criterion of .05 in most cases. The chi-square values were also significant in all cases except one, but that can be expected in a sample of this size.

Out of the short forms tested here, Strahan and Gerbasi's (1972), X1 performed the best with a non-significant Satorra-Bentler chi-square (in spite of a sample size of more than 500 participants). Both CFI and NNFI exceeded the cut off criterion for a good fit and RMSEA was well within the .05 criterion for a good fit (although the X1 contains item 17, which only had a loading of .24 on the X1 factor and .28 on the one factor model for the full scale data). This is in line with Fischer and Fick's (1993) assessment of the short forms which lead them to recommend X1 as one of the best short forms, but different from Loo and Thorpe (2000) who obtained the best fit for Reynolds' (1982) Form A and B.

4. Conclusion

Overall the results support the psychometric quality of measurements obtained with the Internet version of the 33 item MCSDS in a probability based sample of the general public in Iceland. The reliability estimate is acceptable and the items generally seem to measure the same construct (though some factor loadings fall below .30, the loadings for most items are higher). The results suggest that the Internet version of the full MCSDS is in no way inferior to the paper and pencil version of the scale (Barger, 2002; Fischer & Fick, 1993; Loo & Loewen, 2004; Loo & Thorpe, 2000; Sârbescu et al., 2012; Ventimiglia & MacDonald, 2012), and might even obtain more psychometrically sound measurements. However, because the comparison is not only based on a different mode of data collection but also different sample types (population representative versus non-probability), samples, and methods of analysis, such a claim would be premature. In future research, more attention needs to be directed toward items with low factor loadings to learn more about the nature of these items to understand why they measure the construct insufficiently and whether they could be revised or should simply be deleted from the scale. Intensive individual interviews with a focus on item content could be a useful tool in this process to gather information on the quality, clarity and relevance of each item.

Based on the CFA analysis, X1 would be the recommended short form for use on Internet samples, at least in Iceland. The unidimensionality assumption was strongly supported and the factor loadings for all items except one were high. A major concern for X1 however is the low reliability estimate. The short form with the highest reliability estimate was XX. Moreover, the unidimensionality assumption was also supported for that short form. However, the reliability estimate was below the .80 criterion and some of the factor loadings were low.

It must be emphasized that findings for the short forms should be interpreted with caution because the short forms were extracted from the full scale and not administered as short forms, ignoring the possible effects of such alterations. In fact no study, to the authors' knowledge, has examined whether the psychometric properties of the short forms extracted from the full scale and short forms administrated individually are equivalent when administered to fully compatible samples drawn from the same population. This is a limitation of the current study, and previous studies aimed at both short form development and the testing of the psychometric properties of the short forms.

The present research was conducted in Iceland, which sets an obvious limitation to the generalizability of the results. This is a problem frequently encountered in European research due to the fact that most psychometric instruments were developed in English but the majority of Europeans are not native English speakers. The question of measurement equivalence between translated versions of a scale is thus an ever present issue in much of European research, but one that was beyond the scope of this research. It would therefore be informative to investigate the psychometric properties of the MCSDS in Internet samples in other languages and cultures in a probability-based panel or a sample of similar quality.

In sum, the overall findings for the Internet version of the full MCSDS are positive and support the use of the MCSDS in validation research on the Internet. Measurements obtained with the MCSDS are sufficiently reliable and unidimensional. In stating that, it must be emphasized that "[s]aying that a test is unidimensional does not identify that dimension. . ." (Hattie, 1985, pp. 143). A good measure of SDRS that could be recommended for use in Internet studies would need to contain only items that are general enough not to be understood differently between different groups, but specific to the construct of SDRS. More work needs to be done before this can be fully achieved, but this research is a crucial first step in the process because if unidimensionality cannot be assumed, it renders all use of the MCSDS – all calculations and assumptions based on the scale – invalid. Thus, our results provide the necessary basis for further psychometric research on the MCSDS and are very promising for researchers who need to control for or measure SDRS on the Internet.

Acknowledgements

The authors have no conflict of interest to disclose. This research was partly funded by The Eimskip Fund of The University of Iceland (Háskólasjóður Eimskipafélags Íslands). The funding source had no role in study design; in the collection, analysis and interpretation of data; in the writing of the report; nor in the decision to submit the article for publication.

The authors would like to acknowledge networking support by the COST Action IS1004. www.webdatanet.eu.

The authors would also like to thank Dr. Freyr Halldórsson for providing us with an Icelandic translation of the Marlowe–Crowne Social Desirability Scale.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, Joint Committee on Standards for Educational and Psychological Testing (U.S.) (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Ballard, R. (1992). Short forms of the Marlowe–Crowne social desirability scale. *Psychological Reports, 71*, 1155–1160. http://dx.doi.org/10.2466/PR0.71.8.1155-1160.

Barger, S. D. (2002). The Marlowe–Crowne affair: Short forms, psychometric structure, and social desirability. *Journal of Personality Assessment, 79*, 286–305. http://dx.doi.org/10.1207/S15327752JPA7902_11.

Behling, O., & Law, K. S. (2000). *Translating questionnaires and other research instruments: Problems and solutions*. London: Sage Publications.

Beretvas, S. N., Meyers, J. L., & Leite, W. L. (2002). A reliability generalization study of the Marlowe–Crowne social desirability scale. *Educational and Psychological Measurement, 62*, 570–589. http://dx.doi.org/10.1177/0013164402062004003.

Buchanan, T. (2002). Online assessment: Desirable or dangerous? *Professional Psychology: Research and Practice, 33*, 148–154. http://dx.doi.org/10.1037/0735-7028.33.2.148.

Buchanan, T., Johnson, J. A., & Goldberg, L. R. (2005). Implementing a five-factor personality inventory for use on the Internet. *Journal of Psychological Assessment, 21*(2), 115–127. http://dx.doi.org/10.1027/1015-5759.21.2.115.

Buchanan, T. (2001). Online personality assessment. In U.-D. Reips & M. Bosnjak (Eds.), *Dimensions of internet science* (pp. 57–74). Lengerich, Germany: Pabst.

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. http://dx.doi.org/10.1348/000712699161189.

Coles, M. E., Cook, L. M., & Blake, T. R. (2007). Assessing obsessive compulsive symptoms and cognition on the internet: Evidence for the comparability of paper and internet administration. *Behaviour Research and Therapy, 45*, 2232–2240. http://dx.doi.org/10.1016/j.brat.2006.12.009.

Collazo, A. A. (2005). Translation of the Marlowe-Crowne Social Desirability Scale into an equivalent Spanish version. *Educational and Psychological Measurement, 65*, 780–806. http://dx.doi.org/10.1177/0013164405275660.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 349–354. http://dx.doi.org/10.1037/h0047358.

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

DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). London: Sage Publications.

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*, 487–495. http://dx.doi.org/10.1016/j.chb.2014.04.005.

Fischer, D. G., & Fick, C. (1993). Measuring social desirability: Short forms of the Marlowe–Crowne social desirability scale. *Educational and Psychological Measurement, 53*, 417–424. http://dx.doi.org/10.1177/0013164493053002011.

Gerbing, D. W., & Anderson, J. C. (1988). An updated paradigm for scale development incorporating unidimensionality and its assessment. *Journal of Marketing Research, 25*, 186–192. http://dx.doi.org/10.2307/3172650.

Hagtíðindi (2014). Tölvu- og netnotkun á Íslandi og í öðrum Evrópulöndum 2013 [*Computer and Internet usage in Iceland and other European countries 2013*]. Reykjavík: Hagstofa Íslands [Statistics Iceland]. For an English summary see: <http://www.statice.is/?pageid=452&itemid=9302f0f0-a6d0-49fe-9f7e-1e5d9beb5c15>.

Hagtíðindi (2015). Tölvu- og netnotkun á Íslandi og í öðrum Evrópulöndum 2014 [*Computer and Internet usage in Iceland and other European countries 2014*]. Reykjavík: Hagstofa Íslands [Statistics Iceland]. For an English summary see: <http://www.statice.is/Pages/452?itemid=f8bfd5a8-1c7c-4cec-9ac8-4d1d2492087b>.

Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement, 9*, 139–164. http://dx.doi.org/10.1177/014662168500900204.

Hays, R. D., Hayashi, T., & Stewart, A. L. (1989). A five-item measure of socially desirable response set. *Educational and Psychological Measurement, 49*, 629–636. http://dx.doi.org/10.1177/001316448904900315.

Hearn, M., Ceschi, G., Brillon, P., Fürst, G., & Van der Linden, M. A. (2012). A French adaptation of the posttraumatic diagnostic scale. *Canadian Journal of Behavioural Science, 44*, 16–28. http://dx.doi.org/10.1037/a0025591.

Hewson, C., & Charlton, J. P. (2005). Measuring health benefits on the internet: A comparison of paper and Internet administrations of the multidimensional health locus of control scale. *Behavior Research Methods, Instruments, & Computers, 37*, 691–702. http://dx.doi.org/10.3758/BF03192742.

Joinson, A. N. (1999). Social desirability, anonymity and internet-based questionnaires. *Behavior Research Methods, Instruments and Computer, 31*, 433–438. http://dx.doi.org/10.3758/BF03200723.

Karlsdóttir, E. (2013). *Skilningur á atriðum íslenskrar útgáfu kvarða Marlowe og Crowne á félagslega æskilegri svörun* [Understanding of items in the Icelandic version of the Marlowe Crowne social desirability scale]. Unpublished bachelor's thesis. University of Iceland. <http://hdl.handle.net/1946/15356>.

Kaufmann, E., & Reips, U.-D. (2008). *Internet-basierte Messung sozialer Erwünschtheit [internet-based measurement of social desirability]*. Saarbrücken: VDM Verlag Dr. Müller.

Lautenschlager, G. J., & Flaherty, V. L. (1990). Computer administration of questions: More desirable or more social desirability? *Journal of Applied Psychology, 75*, 310–314. http://dx.doi.org/10.1037/0021-9010.75.3.310.

Leite, W. L., & Beretvas, S. N. (2005). Validation of scores on the Marlowe–Crowne social desirability scale and the balanced inventory of desirable responding. *Educational and Psychological Measurement, 65*, 140–154. http://dx.doi.org/10.1177/0013164404267285.

Loo, R., & Loewen, P. (2004). Confirmatory factor analysis of scores from full and short versions of the Marlowe–Crowne social desirability scale. *Journal of Applied Social Psychology, 34*, 2343–2352. http://dx.doi.org/10.1111/j.1559-1816.2004.tb01980.x.

Loo, R., & Thorpe, K. (2000). Confirmatory factor analyses of the full and short versions of the Marlowe–Crowne social desirability scale. *The Journal of Social Psychology, 140*, 628–635. http://dx.doi.org/10.1080/00224540009600503.

Lozar Manfreda, K., & Vehovar, V. (2008). Internet surveys. In E. D. de Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International handbook of survey methodology* (pp. 264–284). London: Lawrence Erlbaum Associates.

Martin, C. L., & Nagao, D. H. (1989). Some effects of computerized interviewing on job applicant responses. *Journal of Applied Psychology, 74*, 72–80. http://dx.doi.org/10.1037/0021-9010.74.1.72.

Noyes, J. M., & Garland, K. L. (2008). Computer- vs paper-based tasks: Are they equivalent? *Ergonomics, 51*, 1352–1375. http://dx.doi.org/10.1080/00140130802170387.

Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw Hill.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

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). San Diego, CA: Academic Press.

Ragnarsson, P. (2011). *Próffræðilegir eiginleikar íslenskrar útgáfu IPIP-HEXACO persónuleikaprófsins* [Psychometric properties of an Icelandic version of the IPIP-HEXACO personality test]. Unpublished bachelor's thesis. University of Iceland. <http://hdl.handle.net/1946/8798>.

Ramanaiah, N. V., Schill, T., & Leung, L. S. (1977). A test of the hypothesis about the two-dimensional nature of the Marlowe–Crowne social desirability scale. *Journal of Research in Personality, 11*, 251–259. http://dx.doi.org/10.1016/0092-6566(77)90022-8.

Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlowe–Crowne social desirability scale. *Journal of Clinical Psychology, 38*, 119–125. http://dx.doi.org/10.1002/1097-4679(198201)38:1<119::AID-JCLP2270380118>3.0.CO;2-I.

Rudmin, F. (1999). Norwegian short-form of the Marlowe–Crowne social desirability scale. *Scandinavian Journal of Psychology, 40*, 229–233. http://dx.doi.org/10.1111/1467-9450.00121.

Sârbescu, P., Costea, I., & Rusu, S. (2012). Psychometric properties of the Marlowe–Crowne social desirability scale in a Romanian sample. *Procedia –Social and Behavioral Sciences, 33*, 707–711. http://dx.doi.org/10.1016/j.sbspro.2012.01.213.

Strahan, R., & Gerbasi, K. C. (1972). Short, homogeneous versions of the Marlowe–Crowne social desirability scale. *Journal of Clinical Psychology, 28*, 191–193. http://dx.doi.org/10.1002/1097-4679(197204)28:2<191::AID-JCLP2270280220>3.0.CO;2-G.

Ventimiglia, M., & MacDonald, D. A. (2012). An examination of the factorial dimensionality of the Marlowe Crowne social desirability scale. *Personality and Individual Differences, 52*, 487–491. http://dx.doi.org/10.1016/j.paid.2011.11.016.

Vésteinsdóttir, V., Reips, U.-D., Joinson, A., & Thorsdottir, F. (2012, November). *Psychometric properties of an internet administered version of the Marlowe–Crowne social desirability scale*. Presented at the 6th ISM Workshop on Internet Survey Methodology, Ljubljana (Slovenia). Supported by COST Action IS1004. <www.webdatanet.eu>. <http://www.websm.org/db/12/15501/Web_Survey_Bibliography/Psychometric_properties_of_an_internet_administered_version_of_the_MarloweCrowne_Social_Desirability_Scale/?menu=1&lst=&q=vesteinsdottir_0_1111110_-1&qdb=12&qsort=0>.

Vésteinsdóttir, V., Reips, U.-D., Joinson, A., & Thorsdottir, F. (2013, July). *Evaluating the psychometric properties of the Marlowe–Crowne social desirability scale in internet format using item response theory (IRT) and intensive individual interviews*. In Presented at the 5th Conference of the European Survey Research Association (ESRA), Ljubljana (Slovenia). http://www.europeansurveyresearch.org/conference/programme?sess=149#637.

Yu, C. Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*. Doctoral dissertation, University of California, Los Angeles 2002.