ARTICLE IN PRESS

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Computers in Human Behavior xxx (2004) xxx–xxx

Computers in Human Behavior

www.elsevier.com/locate/comphumbeh

Short Web-based versions of the perceived stress (PSS) and Center for Epidemiological Studies-Depression (CESD) Scales: a comparison to pencil and paper responses among Internet users

Juan Herrero [a,*], Julio Meneses [b]

[a] *Facultad de Psicología, University of Oviedo, Despacho 211, Plaza Feijoo s/n, 33003 Oviedo, Spain*
[b] *Estudis de Psicología i Ciencies de l'Educaciò, Avinguda Tibidabo 39-43, Universitat Oberta de Catalunya, 08035 Barcelona, Spain*

Abstract

This study reports results from a format comparison between Web-based and pencil and paper questionnaires measuring two well-validated measures of psychological distress among 530 frequent Internet users studying at an on-line University (Open University of Catalonia, UOC). A four-item version of the Perceived Stress Scale (PSS) and a seven-item version of the Center for Epidemiological Studies-Depression Scale (CESD) were randomly administered in Web-based and pencil and paper format. Internal consistency analyses showed that these two brief versions were reliable for on-line assessment. Also, no differences in Cronbach αs were found between formats. Exploratory as well as multigroup factor analysis techniques showed that factor structures were invariant across formats. Overall, results revealed that respondents did not bias their responses in the Web-based condition. As these findings suggest, when familiarity with technology is high among respondents and sample procedures are followed to maximize control for potential participants, Web-based and paper and pencil formats are virtually equivalent, as in the two brief measures of psychological distress used in this study.

Keywords: Computer; Internet; Web-based questionnaires; Perceived stress; Depression

* Corresponding author. Tel.: +34-985-103-282; fax: +34-985-104-144.
E-mail address: olaizola@uniovi.es (J. Herrero).

0747-5632/$ - see front matter
doi:10.1016/j.chb.2004.03.007

1. Introduction

In the last few decades, social scientists have been increasingly paying more attention to the potential of Internet as a source of data collection. For the more optimistic, on-line surveys will increasingly be more a part of the future while other methods of data collection such as telephone or mail will probably belong to the past (Dillman, 2000). Among the advantages of on-line surveys, they not only provide fast access, reliable and verifiable delivery and ease of feedback (Miller et al., 2002). They are also less expensive in terms of production, administration and coding (Buchanan, 2002; Cotten, 2001; Miller et al., 2002) and saving in total costs may be considerable when compared with regular mail surveys (Nesbary, 2000; Weible & Wallace, 1998). In addition, when the population is small or group members are difficult to find, on-line surveys are a very promising research tool to access and involve people (Braithwaite, Waldron, & Finn, 1999; Buchanan, 2002; Coomber, 1997; Evans, Garcia, Garcia, & Baron, 2003; Nosek, Banaji, & Greenwald, 2002).

However, despite these appealing features, Internet-based questionnaire responses could be different from those observed from conventional formats (face-to-face, telephone, self-administered questionnaires, etc.) and psychologists planning to use the Internet to collect data should be aware of this circumstance (Joinson, 1999). Moreover, 'it would be unfortunate if psychologists dismissed this potential prematurely jumping to the conclusion that Internet questionnaires were mere novelties and lacked serious research credibility' (Buchanan & Smith, 1999; p. 128).

In this sense, several efforts have been made to understand how the Internet may affect responses to questionnaires measuring psychological variables including self-monitoring, personality, anxiety, self-esteem, self-consciousness or social desirability (Barbeite & Weiss, 2004; Buchanan & Smith, 1999; Davis, 1999; Joinson, 1999; Knapp & Kirk, 2003; Lang, Raver, White, Hogarty, & Kromrey, 2000; Salgado & Moscoso, 2003). In their comparison of the Gangestad and Snyder's (1985) revised self-monitoring questionnaire between Web-based and pencil and paper versions of the instrument, Buchanan and Smith (1999) applied psychometric and factor analysis (both exploratory and multigroup confirmatory) techniques to show that: (a) psychometric properties were favored for the Internet condition; (b) factor structure was invariant across formats, and, (c) it seemed reasonable to assume that, indeed, the Internet version would provide a better measure of self-monitoring than the pencil and paper version. They advanced two reasons to explain their results. First of all, self-disclosure associated to computerized assessment could be increasing honesty and self-revelation in participants' responses. Secondly, the instrument behaved in a similar manner across conditions, but the heterogeneity of the samples was dissimilar. Thus, 'due to the greater heterogeneity of the Internet sample a clearer picture emerges of the test's factor structure, not that Internet in itself is any better than its conventional equivalent' (p. 139). In our opinion, these two possible explanations (self-disclosure and sampling bias) summarize the two most common reasons adduced by researchers to explain why Internet-based and pencil and paper instruments might not be equivalent.

1.1. Self-disclosure

Much of the research of the effect of on-line assessment on participants' responses can benefit from four decades of studies documenting that computer assessment may be affected by social desirability (Booth-Kewley, Edwards, & Rosenfeld, 1992; Connolly, Jessup, & Valacich, 1990; Levine, Ancill, & Roberts, 1989; Loke & Gilbert, 1995) (Bartram & Bayliss, 1984; Epstein & Klinkenberg, 2001; Richman, Kiesler, Weisb, & Drasgow, 1999; for reviews). In their meta-analysis of 61 studies that compared computer versus pencil and paper versions of the same instrument, Richman et al. (1999) found that social desirability distortion was less in the computer than in the pencil and paper condition, especially if respondents were alone and they could respond fairly freely (e.g., backtrack to previous responses). As Weisband and Kiesler (1996) have noted, whenever social desirability is reduced, self-disclosure is increased. But if computer-mediated administration encourages self-disclosure, non-equivalence of scores on measures in which respondents are asked to reveal personal or sensitive information are to be observed. But why such a disclosure?

Several reasons have been proposed to explain why computer mediated assessment elicits more disclosure among participants than other traditional formats (face-to-face interviews and pencil and paper formats). Firstly, the absence of an experimenter engaged in face-to-face interaction with the participant reduces bias for a correct assessment of highly sensitive psychological issues away from the surveillance of an experimenter (Evans et al., 2003; Nosek et al., 2002). Although this could explain why computer interviews reduce social desirability as compared to face-to-face interviews, it could not explain differences between computer and self-administered paper and pencil questionnaires, because in both conditions the observation of the experimenter is removed.

Secondly, it has been suggested that participants interacting with the computer are probably immersed in the task and several misattributions may cause respondents to reveal personal or sensitive information. A sense of invulnerability to criticism, an illusion of privacy, the impression that responses 'disappear' into the computer are among them (Weisband & Kiesler, 1996). However, these naïve misconceptions of computers might disappear whenever familiarity with computer technology increases.

Thirdly, and most importantly, familiarity with technology has also been used to explain different rates of self-disclosure in computer assessment. When a technology is first introduced and most of the people lack experience with computers they probably are not aware of the risks of self-disclosure of personal information to a computer. According to this explanation, due to the rapid growth of residential computing, the general public will become more computer literate and the effects of lack of experience on computerized instruments' responses will probably decline over the years. As Weisband and Kiesler (1996) have reported in their meta-analysis study, the impact of computer administration self-disclosure from 1964 to 1994 has declined over the years and this decline is not explained fully by changes in the use of various study characteristics in evaluation studies. Despite this decline, however,

these authors found that the effect of computer administration on self-disclosure was still significant for measures eliciting sensitive information (e.g., mental health measures).

1.2. Sampling bias

One question that researchers are concerned with knowing is whether differences between Web-based and other traditional forms of tests are caused by the method of recruiting respondents or the method of collecting responses. In this sense, on-line surveys present a potential bias in terms of who is responding and of a lack of representation of those who do not have access to the Internet. Also, since the Web has no central registry of users and e-mail addresses are so varied that they are virtually impossible to construct randomly, current Internet sampling techniques only permit the generation of diverse, not representative, samples (Best, Krueger, Hubbard, & Smith, 2001). In this vein, Wu and Weaver (1997) have defined on-line surveys as volunteer samples because respondents are self-selected, showing lower response rates than other traditional survey methods (telephone, mails, etc.). These voluntary participants might be more interested, informed, and concerned about the survey topic than other individuals, generating results that are not representative of the larger population. Furthermore, although on-line surveys allow sampling from the broadest pool possible of potential participants, the researcher usually loses control of the type of sample who ultimately chooses to participate (Nosek et al., 2002). As Dillman and Salent (1994) have stated, non-respondent error is an issue here if those who did not respond to a questionnaire are different from those who did respond to the questionnaire in a way that is important to the study.

These potential sampling biases should be taken seriously if the researcher intends to claim generalizability from the sample to the population. There are, however, some ways to lessen the effect of sampling bias in Internet surveys. In their meta-analysis of 68 Internet-based surveys, Cook, Heath, and Thompson (2000) found that the number of contacts with participants, the type of contact (whether personalized or not) and pre-contacts were factors most associated with a higher response rate in the Web studies analyzed (see also, Solomon, 2001). Also, response rates may be somewhat increased in Web surveys if they load quickly (Dillman, Tortora, & Conrad, 2001) and complex questions or personal information are not asked (i.e. e-mail address) (Jeavons, 2000).

Also, to gain control over participants in the study it could be useful to narrow the range of potential respondents by selecting Internet target populations. Fortunately, in certain populations Web-based surveys may have only minor coverage problems, as is the case with University professors, government employees or members of some corporations since they have Internet address and access (Dillman, 2000). These populations are candidates for targeted advertising: a method that can help to increase the researcher's control over who hears about the study and how the study is described, mostly contacting directly particular groups of interest (Nosek et al., 2002). In addition, providing each individual with a unique access code and a link to the website allows the researcher to gain control over the participants in the study.

This recruiting technique, known as invited participation, allows the researcher to verify that each participant is engaged in the study on only one occasion and when combined with targeted advertising, control is maximized. In this sense, researchers willing to compare Web-based tests to other traditional form of tests could work with this especially networked populations (University staff, government employees, etc.) and assign participants randomly to each condition (see Couper, Blair, & Triplett, 1999; Kwak & Radler, 2002). With this procedure, participants will probably be equivalent in socio-demographic variables and differences in responses were due to the test's condition and not to the respondent's profile in each condition.

1.3. The present study

Drawing from these ideas, the present study aims to evaluate Web-based and pencil and paper responses of two well-validated measures of psychological distress among Internet users. Most of the research conducted about the characteristics of Web-based surveys have been carried out among participants who used Internet to complete the surveys, but no indications were provided as to what extent they were Internet users (Joinson, 1999; Epstein, Klinkenberg, Wiley, & McKinley, 2001; see Barbeite & Weiss, 2004 for an exception) or simply frequent Internet access was inferred due to the fact that respondents were college students where campus connection was available (Knapp & Kirk, 2003). In these cases, participants were recruited by traditional means and randomly assigned to a Web-based or traditional form of the test.

However, when comparing Web-based to other traditional formats researchers need to assure that participants are familiar with Internet technologies in order to control for potential self-disclosure bias for participant's responses in the Internet condition. In this present study, we selected an on-line University as a target population and students in the behavioral on-line courses were invited to participate and were granted access to the questionnaires (whether on-line or pencil and paper format). This procedure has the advantage of controlling for two potentially confounding influences when comparing Web-based and pencil and paper responses. Firstly, it homogenizes the groups in terms of socio-demographic variables (age, income, education, etc.). Secondly, it guarantees that all participants are Internet users and that they are all familiar with this technology to a certain extent. Since on-line University studies imply that students use Internet technologies in their learning process (electronic forums, instant messengers, e-mail, electronic research, etc.), it is reasonable to assume that familiarity with technology was present in all participants.

In addition, comparison among Internet-based and pencil and paper formats should first be conducted on well-established instruments with satisfactory psychometric properties and, if possible, on short tests since they are less likely to induce boredom or fatigue in the volunteer participants (Buchanan & Smith, 1999; Kaye & Johnson, 1999). In our study we tried to accomplish both requirements. On one hand, we selected two well-validated measures of psychological distress measuring perceived stress (Perceived Stress Scale, Cohen, Kamarck, & Mermelstein, 1983) and depressive mood (Center for Epidemiological Studies, Depression, Radloff, 1977).

Although these scales have acceptable psychometric properties when traditionally measured (Gracia & Herrero, 2004) the reliability of an instrument when converted to on-line format must be assessed prior to its use (Buchanan, 2002; Buchanan & Smith, 1999; Joinson, 1999; Salgado & Moscoso, 2003). As Tabachnik and Fidell (1989) have pointed out, tests have to show a similar factor structure, the same number of factors should account for similar proportions of variance and the same items should load on each factor, to be equivalent. Buchanan and Smith (1999) have also added that similar distribution characteristics of the underlying constructs have to be observed across groups. In this present study while we explored psychometric properties and mean differences for items and scales scores across conditions, we also contrasted the hypothesis that the measurement model for stress and depression was identical for both conditions and that the structural relationship between stress and depression constructs remained invariant across formats.

On the other hand, on-line surveys should be as short as possible for quick completion so it is therefore necessary to conduct research examining the possible use of short and reliable measures of psychological constructs. Because of this, the present study focuses on the psychometric properties of the two on-line shortened versions of four items for the perceived stress scale and seven items for the depression scale.

Finally, we used Web-based instead of e-mail-based questionnaires because they are easier to complete and the process is more straightforward – partially eliminating biased responses due to socio-demographic variables –, and have been recommended when compared to e-mail based techniques (Batinic, 1997). In this sense, studies documenting lower responses for e-mail questionnaires as compared to mail surveys have been partially explained by the technical difficulties inherent to download, open, complete, and submit the attached survey. Thus, respondents who were more likely to adapt to new technologies used to be over-represented in e-mail based surveys (Couper et al., 1999; Kwak & Radler, 2002) and this probably also affected mean responses.

2. Method

2.1. Procedure

College students at the on-line Open University of Catalonia (UOC, see www.uoc.edu) were assigned at random to either an Internet or to a pencil and paper survey (format condition) to complete two measures of psychological distress (depression and stress). An e-mail explaining the procedure for participating in the study as well as a compromise of confidentiality and anonymity was sent out to students in the on-line behavioural science courses. We enclosed a cover letter with the instruction sheet and a stamped return envelope (for the pencil and paper condition) and placed envelopes ordered at random in several boxes at the entrance of the rooms where students had off-line presentations at the beginning of the semester. The envelopes looked alike but were actually of two kinds (whether they belonged to

the Internet or to the pencil and paper condition). For each student, an eight-digit code was created according to his/her original e-mail address at the UOC. A username and password is provided by the UOC at the time of students' course registration, and several security measures are regularly implemented to keep the system secure. Thus, it was possible to assign a unique and non-meaningful eight-digit code to each of the participants in the study.

For the Internet condition, the enclosed instruction sheet listed a Universal Resource Location (URL) to access the on-line questionnaire. Also, in the upper right side of this page, the eight-digit code was provided. To complete the survey, participants had to first log-on with their original username and password to get access to the UOC's server. Once the participants tried to access the URL containing the on-line survey, access was granted for eight-digit codes that matched with original UOC's e-mail address in our database. This procedure did not decrease response rates while it increased data quality (Heerwegh & Loosveldt, 2002). Once the participants submitted their responses by clicking the corresponding button, their entry in our database was blocked, not permitting further trials to complete the questionnaire. This procedure allowed controlling for additional possible questionnaire abuse.

For the paper and pencil condition, participants found enclosed with a stamped return envelope a cover letter which contained a different URL where they could download a printer-friendly version of the survey, instructions to return the questionnaires in the enclosed envelope, as well as the eight-digit code to enter the URL. This eight-digit code was used to assure that only one survey per respondent was finally coded into the database.

2.2. Participants

Participants were recruited from three undergraduate courses in the behavioral science disciplines. To obtain an equivalent number of participants in the Internet and pencil and paper conditions, we randomly placed a ratio of 6/4 envelopes favouring the pencil and paper condition. The rationale for this was based on our expectation that the response rate would be higher for the Internet condition because of the basic technical requirements to be registered at the UOC. Since potential respondents were all Internet users, owned a computer and had Internet connection at home – 67% of participants had broadband connections at that time (cable modem or DSL) –, we expected a higher rate of responses in the Internet condition. Increasing the probability of choosing a pencil and paper questionnaire would probably cancel out the effect of lower response rates from the pencil and paper condition in the final sample.

Of the 1618 participants contacted through e-mail, 970 took envelopes with instructions to complete the survey (about 60%). These 970 packets were distributed as follows: 407 (42%) for the Internet condition and 563 (58%) for the pencil and paper condition. A total of 610 surveys were finally submitted (62% of those taken) from which 80 were excluded from the analyses because they failed to provide valid data for all of the responses, yielding a final number of 530 surveys. Of those 80 ques-

tionnaires with non-valid data, 75% corresponded to surveys in the pencil and paper condition. Finally, the 530 valid surveys were distributed almost equally for each condition (Internet: $N = 262$, 50.6%; pencil and paper: $N = 268$, 49.4%). Thus, a response rate of 62% and 48% was found for the Internet and pencil and paper condition, respectively. Participants were 18 years older ($M = 29.06$, SD = 5.14), with a larger proportion of females (66%). Age F (1,528) = 0.48, n.s., and income F (1,528) = 2.71, n.s., were equally distributed across conditions. However, the proportion of females was significantly greater in the Internet condition ($\chi^2 = 9.70$, $p < 0.01$). In other words, females in the Internet condition were over-sampled because of their higher response rates due maybe to higher rates of participation in on-line activities at the campus. This tendency was also confirmed by ratios of gender/participation in on-line discussions observed in the on-line classrooms. We selected three on-line classrooms and calculated the average messages posted by males and females. The messages of 83 females and 30 males to the on-line board were followed for two weeks. Females posted 255 messages whereas only 30 messages came from males. A simple ratio of number of messages posted by females/number of females showed that on the average, each female posted 3.07 messages to the board. For males, this ratio descended to 1.

2.3. Instruments

Two widely accepted measures of psychological distress were selected: a measure of stress (Perceived Stress Scale; Cohen et al., 1983); and a measure of depressive mood (Center for Epidemiology Studies-Depression Scale CES-D; Radloff, 1977). These two measures have been studied extensively and brief versions were available in the literature. We used these brief versions to obtain a short and reliable survey of stress and depressive mood.

2.3.1. Perceived stress

The original version of PSS (Perceived Stress Scale) is a 14-item scale (Cohen et al., 1983) that measures the degree to which respondents appraised situations as stressful in the last month. The authors have provided reliable psychometric properties for shorter versions of the original scale (Cohen & Williamson, 1988). In this study we used the PSS-4, the four-item version with responses scored on a one to five-point scale from (1) *never* to (5) *very often*. These four items were: In the last month, (1) *how often have you felt that you were unable to control the important things in your life*? (2) *how often have you felt confident about your ability to handle your personal problems*? (3) *how often have you felt that things were going your way*? and (4) *how often have you felt difficulties were piling up so high that you could not overcome them*?

2.3.2. Depression

The original CESD is a 20-item scale representing four subscales of depressive sypmtomatology: negative affect, positive affect, somatic symptoms, and interpersonal problems (Radloff, 1977). Research has traditionally supported this four-factor

structure pointing out, however, that the use of combined score loses little information (Iwata & Roberts, 1996; Knight, Williams, McGee, & Olaman, 1997; MacKinnon, McCallum, Abdrews, & Anderson, 1998; McCallum, MacKinnon, Simons, & Simons, 1995), which explains the widely accepted use of the total score of CESD as a valid measure of depressed mood. In this study, we used a seven-item version of the original CESD suggested by Santor and Coyne (1997). Item responses were rated on a four-point scale from (1) *rarely or none of the time (less than once a week)* to (4) *most or all of the time (57 days a week)*. This shortened version included three items tapping dysphoric mood ('*I felt that I could not shake off the blues even with the help from my family or friends*', '*I felt depressed*', '*I felt sad*'), and one item for each of the following domains: motivation ('*I felt everything I did was an effort*'), concentration ('*I had trouble keeping my mind on what I was doing*'), pleasure ('*I enjoyed life*') and poor sleep ('*My sleep was restless*').

3. Results

3.1. Multivariate ANOVA's

We conducted multivariate analysis of variance (MANOVA) to search for differences in item responses across format conditions (Internet, and pencil and paper). Two separate MANOVA's were performed for the PSS (four items) and CESD (seven items) scales. Because of the significantly greater proportion of females in the Internet condition, we analyzed main effects as well interaction effects (gender × format) to control for possible confounding effects of gender on format. Table 1 presents means, standard deviations, kurtosis and skewness for the 12 items and total scores.

For the four items of perceived stress, non-interaction effects were found F (4,523) = 1.01, n.s. The main effects of format F (4,523) = 9.58, $p < 0.001$, and gender F (4,533) = 4.13, $p < 0.01$ were significant. Univariate statistics showed that the mean of the item 2 (*confident about ability to handle personal problems*) was significantly greater for the Internet condition F (1,539) = 34.18, $p < 0.001$. Also, women scored significantly higher in items 1 and 4 (*unable to control the important things in life* and *difficulties were piling up so high that could not overcome them*) (F's (1,526) > 5.21, p's < 0.05) while males scored higher in item 2 (*confident about ability to handle personal problems*) (F (1,526) = 14.84, $p < 0.001$).

For the seven items of depression, neither gender × format interaction F (7,520) = 1.78, n.s., nor format effects were found (F (7,530) = 1.13, n.s.). Gender effects were statistically significant (F (7,520) = 3.19, $p < 0.01$), with females scoring higher than males in items 1, 3, 5, 7 (*could not shake off the blues*, *felt depressed*, *sleep was restless*, *felt sad*) (F's (1,526) > 2.99, p's < 0.05).

Also, two gender × format MANOVA's were conducted for the total scores of PSS-four and CESD-seven scales. These analyses revealed that gender was the only variable accounting for differences in stress and depression (F (2,525) = 6.89, $p < 0.01$). The univariate tests indicated that females scored higher than males both

Table 1
Means, standard deviations, kurtosis and skewness for items and total scores of PSS-4 and CESD-7 scales for Internet and pencil and paper conditions

	Internet				Pencil and paper			
	M	SD	Kurtosis	Skewness	M	SD	Kurtosis	Skewness
1. How often have you felt that you were unable to control the important things in your life?	2.44	0.88	0.11	0.41	2.58	0.87	−0.13	0.23
2. How often have you felt confident about your ability to handle your personal problems?	4.12	0.78	1.38	−0.90	3.72	0.77	0.72	−0.62
3. How often have you felt that things were going your way?	3.65	0.78	0.60	−0.57	3.68	.74	−0.16	−0.23
4. How often have you felt difficulties were piling up so high that you could not overcome them?	2.58	0.83	0.39	0.43	2.70	0.88	0.06	0.42
PSS-4 total score	9.25	2.37	0.58	0.74	9.86	2.34	0.41	0.48
1. I felt that I could not shake off the blues even with the help from my family or friends	1.46	0.73	1.32	1.47	1.50	0.74	1.45	1.38
2. I had trouble keeping my mind on what I was doing	1.85	0.73	−0.51	0.41	1.87	0.81	−0.38	0.56
3. I felt depressed	1.48	0.73	1.17	1.39	1.75	0.82	0.06	0.82
4. I felt everything I did was an effort	1.63	0.74	0.23	0.94	1.68	0.75	−0.05	0.82
5. My sleep was restless	1.74	0.85	−0.14	0.87	1.61	0.75	1.06	1.15
6. I enjoyed life	3.06	0.84	−0.01	−0.71	23.88	0.76	−0.19	−0.31
7. I felt sad	1.61	0.80	0.49	1.12	1.63	0.75	−0.18	0.82
CESD-7 total score	11.57	3.79	0.28	1.14	11.85	3.78	0.38	0.86

in stress (F (1,526) = 11.25, $p < 0.001$) and depression (F (1,526) = 12.09, $p < 0.001$). Neither gender × format nor format effects were found ($p > 0.05$).

3.2. Internal consistency analyses

Internal consistency analyses were carried out for item responses in each condition. Table 2 presents the results of these analyses. Cronbachs' αs for items corresponding to the PSS and CESD scales in each condition are also presented.

As seen in Table 2, αs were from moderate to high ($0.68 > \alpha s < 0.83$) indicating an acceptable internal consistency for the two scales in each format. The PSS-four scale showed a slightly greater α for the Internet condition whereas the CESD-seven scale virtually equalled αs for both conditions. Item-total intercorrelations indicated that the Pearson correlation coefficients between the score of each item and the sum of the scores on the remaining items were from moderate to large for each scale across formats ($0.31 > r$'s < 0.74).

3.3. Principal components and confirmatory factor analyses

We conducted Principal Component analyses for each scale and each format. For the four items of the perceived stress scale, one principal component was found for both the Internet (explained variance = 54%) and the pencil and paper condition (explained variance = 51%). All factor loadings for both conditions were greater than |0.50|. For the seven items of the depression scale, one principal component was also obtained in both conditions. Explained variance was approximately equal for both conditions (51%) and factor loadings were greater than |0.45|.

To further explore the factorial structure across format conditions, we used EQS (Bentler, 1995) structural equation program to estimate a set of models examining the measurement model of the 11 items of the PSS and CESD scales. Maximum Likelihood estimator and corrected χ^2 were used for the calculation of robust CFI fit index, standard errors, and statistical significance of the parameters. For correcting departure from multinormality, the Satorra–Bentler corrected χ^2 was used.

Firstly, a measurement model (see Fig. 1) with PSS items loading in a Stress latent variable and CESD items loading in a Depression latent variable was fit for the complete sample ($N = 530$). This measurement model allowed latent factors to covary while imposing the restriction of no cross-loadings as well as no correlated errors neither among measures of the same construct nor among measures of different constructs. This model fit the data well: (robust) χ^2 (43, $N = 530$) = 103.56, $p < 0.001$, CFI = 0.95, Robust CFI = 0.96, GFI = 0.96, AGFI = 0.93, SRMR = 0.042; RMSEA = 0.048 (90% confidence interval 0.035, 0.060). As Hu and Bentler (1999) have suggested, a SRMR < 0.08 indicated that the structure was adequately predicted, while CFI (>0.95) and RMSEA (<0.05) confirmed the adequacy of the measurement part of the model.

Secondly, to test hypotheses regarding the factorial invariance of the questionnaires across format conditions we conducted multigroup analyses. Two models were tested. In the first between-group model (unrestricted model) all of the pa-

Table 2
Internal consistency analyses for PSS-4 and CESD-7 scales for Internet and pencil and paper condition

	Internet				Pencil and paper			
	Scale Mean if item deleted	Scale Variance if item deleted	Item-total correlation	Alpha if Item deleted	Scale Mean if item deleted	Scale Variance if item deleted	Item-total correlation	Alpha if Item deleted
1. How often have you felt that you were unable to control the important things in your life?	6.80	3.07	0.57	0.62	7.29	3.08	0.54	0.56
2. How often have you felt confident about your ability to handle your personal problems?	7.36	3.72	0.43	0.70	7.59	3.44	0.52	0.58
3. How often have you felt that things were going your way?	6.89	3.79	0.50	0.67	7.54	4.02	0.31	0.70
4. How often have you felt difficulties were piling up so high that you could not overcome them?	6.67	3.30	0.54	0.64	7.17	3.16	0.49	0.60
		Cronbachs' $\alpha = 0.72$				Cronbachs' $\alpha = 0.68$		
1. I felt that I could not shake off the blues even with the help from my family or friends	11.23	10.47	0.70	0.78	11.71	10.31	0.72	0.78
2. I had trouble keeping my mind on what I was doing	10.84	11.33	0.50	0.81	11.35	11.24	0.44	0.83
3. I felt depressed	11.21	10.32	0.74	0.77	11.47	9.88	0.75	0.77
4. I felt everything I did was an effort	11.07	11.39	0.50	0.81	11.55	10.85	0.58	0.80
5. My sleep was restless	10.95	11.84	0.31	0.84	11.62	11.81	0.37	0.84
6. I enjoyed life	9.75	10.63	0.54	0.80	10.11	11.25	0.48	0.82
7. I felt sad	11.30	10.04	0.72	0.77	11.57	10.29	0.71	0.78
		Cronbachs' $\alpha = 0.82$				Cronbachs' $\alpha = 0.83$		

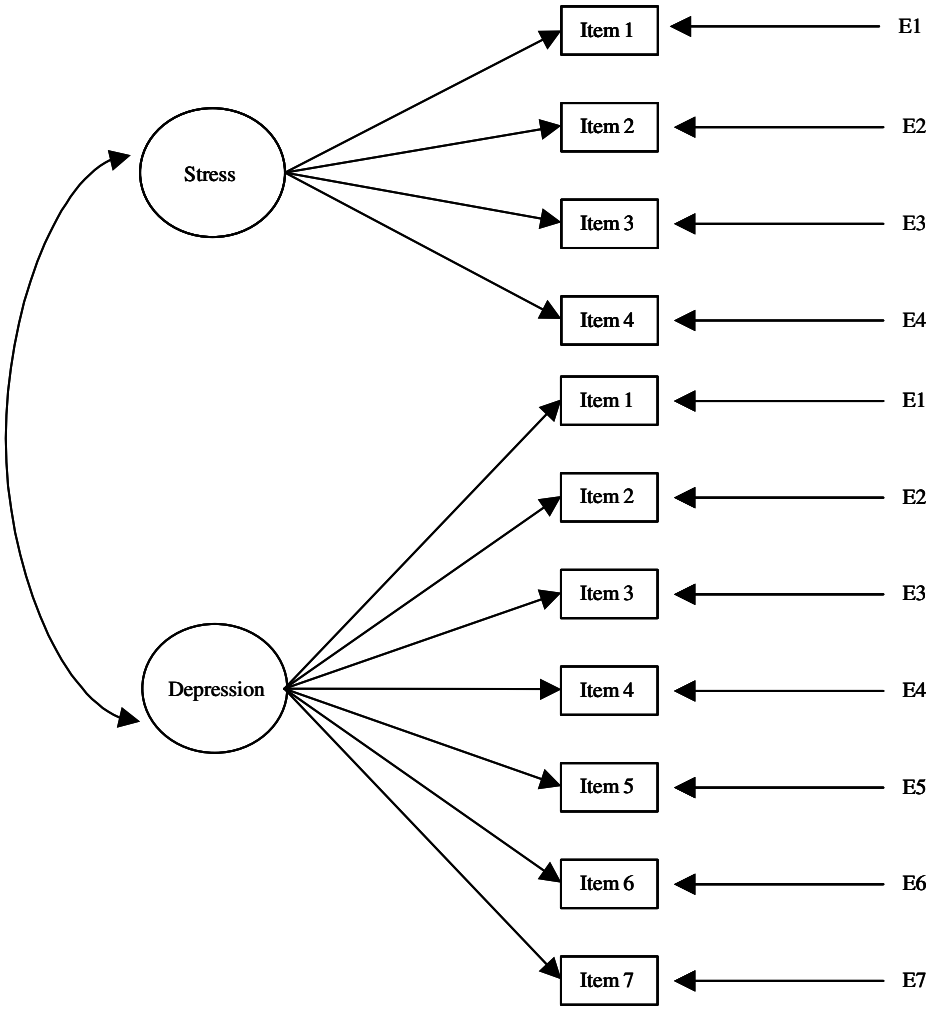

Fig. 1. Measurement model for PSS-4 and CESD-7 items.

rameter estimates (factor loadings and correlation among latent variables) were freely estimated across format conditions. In the second between-group model (restricted model) we constrained each of the factor loadings as well as the correlation between latent variables to be invariant across format conditions. If the chi-square for the restricted model were significantly larger than the chi-square of the unrestricted model, the assumption of invariance would not be tenable. Results of the multigroup analyses indicated a non-significant difference between the unrestricted and restricted model: $\Delta\chi^2(10, N = 530) = 14.54$, n.s. These results supported the factorial invariance of the measurement model across groups. Table 3 presents the unstandardized and standardized parameter estimates of the measurement model.

Table 3
Maximum-likelihood factor loadings for the measurement model

	Factor loadings[a]		
	Unstandardized	Standardized	Z[b]
Stress			
Item 1[c]	1.00	0.70 (0.68)	–
Item 2	−0.76	−0.59 (−0.64)	−12.39
Item 3	−0.59	−0.54 (−0.46)	−9.28
Item 4	0.93	0.69 (0.63)	13.93
Depression			
Item 1[c]	1.00	0.83 (0.83)	–
Item 2	0.62	0.53 (0.46)	12.20
Item 3	1.07	0.86 (0.85)	22.83
Item 4	0.67	0.56 (0.56)	12.85
Item 5	0.46	0.34 (0.37)	7.58
Item 6	−0.75	−0.56 (−0.58)	−13.22
Item 7	1.05	0.82 (0.85)	22.13
Correlation among latent variables	0.28	0.73 (0.78)	8.90

[a] The unstandardized factor loadings were constrained to be equal across the two conditions. Factor loadings are for Internet condition (in parenthesis factor loadings for pencil and paper condition).
[b] All significant at $p < 0.001$.
[c] Parameter is fixed to 1.00 during estimation.

As seen in Table 3, all factor loadings were from moderate to large and highly significant ($p < 0.001$). Also, the correlation coefficient between latent variables (stress and depression) showed a highly significant positive relationship ($0.73 > r$'s < 0.78, $p < 0.001$) across formats. Finally, as expected because of the observed invariance of the measurement model in both formats, the standardized parameter estimates were almost equivalent for both conditions (Internet and pencil and paper).

4. Discussion

This research has explored the internal consistency and factor structure of short on-line versions for two widely accepted measures of psychological distress among Internet users. The main objective of the study was to analyze whether two short versions of CESD and PSS were reliable for use in Web-based surveys among Internet users. Questionnaires were completed in two randomly assigned conditions (Internet and pencil and paper) to see if psychometric properties and factor structure were invariant across conditions. Also, item mean responses and global scale scores were compared across conditions. Finally, factor structure was examined using exploratory as well as confirmatory multigroup techniques.

Results showed that item mean responses were almost invariant across the two conditions. Only 1 item out of 11 was significantly different (*confident about ability to*

handle personal problems) for the Internet and pencil and paper condition, indicative of that condition was not biasing responses in the sample. Also, distributional characteristics (skewness and kurtosis) of the items and the scales were similar across groups. When internal consistency was explored for each brief scale across the two conditions, acceptable Cronbachs' αs were obtained, albeit PSS-4 showed lower αs for both conditions than the CESD-7. Nonetheless, psychometric properties were identical across format suggesting that regarding these measures of psychological distress, Web-Internet based questionnaires behaved the same as their pencil and paper counterparts among respondents.

When analyses extended to the factor structure of the questionnaires across conditions, the same pattern of results emerged. Not only both short scales seemed to have a single-dimension structure for both conditions as showed by the exploratory factor analyses. Moreover, multigroup analyses gave evidence that factor loadings across conditions were of the same magnitude and that the relationship between constructs (stress and depression) remained invariant whether respondents used the Internet or the pencil and paper questionnaires. Overall, results indicated a strong equivalence between Web-based and pencil and paper formats among Internet users.

These results are suggesting that, in fact, respondents did not systematically bias their responses to Web-based questionnaires measuring psychological distress when compared to more traditional pencil and paper versions of the same self-reports. The equivalence between Web-based and pencil and paper questionnaires found here indicates that among Internet users familiarized with the utilization of new technologies, format condition has no effect on participants' responses. Previous studies (Weisband & Kiesler, 1996) have already suggested that as technology evolves and more user-friendly applications are developed, the effect of technology on responses to Web-based surveys will diminish. According to Knapp and Kirk (2003), when participants are equally familiar with the use of technology, there is no reason to believe 'that a test-circumstance foreign enough to cause participants to respond differently' exists (p. 132). In our study, not only the levels of familiarity with Web-based technology were similar among participants but also presumably high among them. This circumstance adds credibility to the thesis that while increasing use of Internet technology is to be observed among the population, more reliable the Internet would be as a source of collecting psychological-related data.

Also, the average participant in our study owned a computer at home with a broadband connection (cable modem or DSL) and went on-line seven days a week. It is therefore tenable that most of the participants in the Internet condition responded to the questionnaires from home, although we did not specifically control for this situation. As Buchanan and Smith (1999) have remarked, people may experience a different set of environmental stimuli or distractions whether they complete the test on a computer at home or elsewhere (work, a noisy public room, etc.). The increasing availability of access to the Internet from home would probably parallel those conditions associated with more conventional self-administered assessments.

In this study, Internet and pencil and paper conditions were much alike in terms of familiarity with the technology and the distribution of socio-demographic variables, two potential influences that have been used to explain differences in Web-based and pencil and paper assessments (Buchanan & Smith, 1999). What the present study clearly shows is that when these circumstances are controlled and participants are distributed equally on socio-demographic variables and familiarity with the technology is high among participants, the effect of format on response is not present. At this point, it seems that technology does not matter and that using the Web-based questionnaire's potential to analyze psychological distress could be a valid resource for researchers.

Acknowledgements

This research was supported by grant 2003FI0001 from the Department of Universities, Research and Information Society of the Catalonian Government, Spain.

References

Barbeite, F., & Weiss, E. (2004). Computer self-efficacy and anxiety scales for an Internet sample: testing measurement equivalence of existing measures and development of new scales. *Computers in Human Behavior, 20*, 1–15.

Bartram, D., & Bayliss, R. (1984). Automated testing: Past, present and future. *Journal of Occupational Psychology, 57*, 221–237.

Batinic, B. (1997). How to make an Internet based survey?. In W. Bandilla & F. Faulbaum (Eds.), *SoftSta '97 Advances in Statistical Software 6*. Stuttgart: Lucius & Lucius.

Bentler, P. M. (1995). *EQS structural equations program manual*. Encino, CA: Multivariate Software.

Best, S., Krueger, B., Hubbard, C., & Smith, A. (2001). An assessment of the generalizability of Internet surveys. *Social Science Computer Review, 19*, 131–145.

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.

Braithwaite, D. O., Waldron, V. R., & Finn, J. (1999). Communications of social support in computer-mediated groups for people with disabilities. *Health Communication, 11*, 123–151.

Buchanan, T. (2002). Online assessment: desirable or dangerous? *Professional Psychology: Research and Practice, 33*, 148–154.

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

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior, 24*, 385–396.

Cohen, S., & Williamson, G. (1988). Perceived stress in a probability sample of the United States. In S. Spacapam & S. Oskamp (Eds.), *The social psychology of health: Claremont Symposium on applied social psychology*. Newbury Park, CA: Sage.

Connolly, T., Jessup, L. M., & Valacich, J. S. (1990). Effects of anonymity and evaluative tone on idea generation in computer-mediated groups. *Management Science, 36*, 689–703.

Cook, C., Heath, F., & Thompson, R. L. (2000). A meta-analysis of response rates in Web- or Internet-based surveys. *Educational and Psychological Measurement, 60*, 821–836.

Coomber, R. (1997). Using the Internet for survey research. *Sociological Research, 2*. Available: http://www.socresonline.org.uk/socresonline/2/2/2.html.

Cotten, S. R. (2001). Implications of Internet technology for medical sociology in the new millennium. *Sociological Spectrum, 21*, 319–340.

Couper, M., Blair, J., & Triplett, T. (1999). A comparison of mail and e-mail for a survey of employees in federal statistical agencies. *Journal of Official Statistics, 15*, 39–56.

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

Dillman, D. A. (2000). *Mail and Internet surveys: The tailored design method* (2nd ed.). New York, NY: John Wiley & Sons.

Dillman, D., & Salent, P. (1994). *How to conduct your own survey*. New York, NY: John Wiley and Sons.

Dillman, D. A., Tortora, R. D., Conrad, J. & Bowker D. (2001). Influence of plan vs. fancy design on response rates of Web surveys. Working paper available from http://survey.sesrc.wsu.edu/dillman/papers.htm [Accessed 11/03].

Epstein, J., & Klinkenberg, W. D. (2001). From Eliza to Internet: A brief history of computerized assessment. *Computers in Human Behavior, 17*, 295–314.

Epstein, J., Klinkenberg, W. D., Wiley, D., & McKinley, L. (2001). Insuring sample equivalence across Internet and paper-and-pencil assessments. *Computers in Human Behavior, 17*, 339–346.

Evans, D. C., Garcia, D. J., Garcia, D. M., & Baron, R. S. (2003). In the privacy of their homes: Using the Internet to assess racial bias. *Society for Personality and Social Psychology, 29*, 273–284.

Gangestad, S., & Snyder, M. (1985). 'To carve nature at its joints': On the existence of discrete classes in personality. *Psychological Review, 92*, 317–340.

Gracia, E., & Herrero, J. (2004). Determinants of social integration in the community: An exploratory analysis of personal, interpersonal, and situational variables. *Journal of Community and Applied Social Psychology, 14*, 1–15.

Heerwegh, D., & Loosveldt, G. (2002). Web surveys: The effect of controlling survey access using PIN numbers. *Social Science Computer Review, 20*, 10–21.

Hu, L., & Bentler, P. M. (1999). Cut-off criterion for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.

Iwata, N., & Roberts, R. (1996). Age differences among Japanese on the Center for Epidemiologic Studies Depression Scale: an ethnocultural perspective on somatization. *Social Science and Medicine, 43*, 967–974.

Jeavons, A. (2000). Ethology and the Web: Observing Respondent Behaviour in Web Surveys, and Paradata: Uses in Web Surveying. In Brookes (Ed.), *Marketing Research in a .com Environment. ESOMAR Monograph*, Vol. 10.

Joinson, A. N. (1999). Social desirability, anonymity, and Internet-based questionnaires. *Behaviour Research Methods, Instruments, and Computers, 31*, 433–438.

Kaye, B. K., & Johnson, T. J. (1999). Research methodology: Taming the cyber frontier. Techniques for improving online surveys. *Social Science Computer Review, 17*, 323–337.

Knapp, H., & Kirk, S. A. (2003). Using pencil and paper, Internet and touch-tone phones for self-administered surveys: does methodology matters? *Computers in Human Behaviour, 19*, 117–134.

Knight, R., Williams, S., McGee, R., & Olaman, S. (1997). Psychometric properties of the Center for Epidemiologic Studies Depression Scale (CES-D) in a sample of women in middle life. *Behavioral Research and Therapy, 35*, 373–380.

Kwak, N., & Radler, B. (2002). A comparison between mail and web surveys: Response pattern, respondent profile, and data quality. *Journal of Official Statistics, 18*, 257–273.

Lang, T., Raver, R., White, J., Hogarty, K. & Kromrey, J. (2000). Survey data collection strategies: Response differences between Web-based and paper-based methods. In Paper presented at the *Annual Meeting of the Florida Educational Research Association*.

Levine, S., Ancill, R. J., & Roberts, A. P. (1989). Assessment of suicide risk by computer-delivered self-rating questionnaire: Preliminary findings. *Acta-Psychiatrica-Scandinavica, 80*, 216–220.

Loke, S. D., & Gilbert, B. O. (1995). Method of psychological assessment, self disclosure, and experiential differences: A study of computer, questionnaire and interview assessment formats. *Journal of Social Behavior and Personality, 10*, 255–263.

MacKinnon, A., McCallum, J., Abdrews, G., & Anderson, I. (1998). The Center for Epidemiologic Studies Depression Scale in older community samples in Indonesia, North Korea, Myanmar, Sri Lanka, and Thailand. *Journals of Gerontology: Series B: Psychological Sciences and Social Sciences, 53*, 343–352.

McCallum, J., MacKinnon, A., Simons, L., & Simons, J. (1995). Measurement properties of the Center for Epidemiologic Studies Depression Scale: An Australian community study of aged persons. *Journals of Gerontology: Series B: Psychological Sciences and Social Sciences, 50*, 182–189.

Miller, E. T., Neal, D. J., Roberts, L. J., Baer, J. S., Cressler, S. O., Metrik, J., & Marlatt, G. A. (2002). Test-reest reliability of alcohol measures: Is there a difference between Internet-based assessment and traditional methods. *Psychology of Addictive Behaviours, 16*, 56–63.

Nesbary, D. K. (2000). *Survey Research and the World Wide Web*. Boston: Allyn and Bacon.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). E-research: Ethics, security, design, and control in psychological research on the Internet. *Journal of Social Issues, 58*, 161–176.

Radloff, L. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385–401.

Richman, W. L., Kiesler, S., Weisb, S., & Drasgow, F. (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *Journal of Applied Psychology, 84*, 754–775.

Salgado, J. F., & Moscoso, S. (2003). Internet-based personality testing: Equivalence of measures and assesses' perceptions and reactions. *International Journal of Selection and Assessment, 11*, 194–205.

Santor, D., & Coyne, J. (1997). Shortening the CES-D to improve its ability to detect cases of depression. *Psychological Assessment, 9*, 233–243.

Solomon, D. (2001). Conducting Web-based surveys. *Practical assessment, Research and Evaluation, 7*, 19.

Tabachnik, B., & Fidell, L. (1989). *Using multivariate statistics* (2nd ed.). New York: Harper & Row.

Weible, R. J., & Wallace, J. (1998). Cyber research: The impact of the Internet on data collection. *Marketing Research, 10*, 19–31.

Weisband, S. & Kiesler, S. (1996). Self-disclosure on computer forms: meta-analysis and implications. In *Proceedings of the SIGCHI on Human factors in computing systems: common ground*. Vancouver: Canada.

Wu, W., & Weaver, D. (1997). Online democracy or online demagoguery – public opinion 'Polls' on the Internet. *Harvard International Journal of Politics/Press, 2*, 71–86.