Computers in Human Behavior 25 (2009) 393–401

Contents lists available at ScienceDirect

Computers in Human Behavior

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

Web-based measurement: Effect of completing single or multiple items per webpage

Frances P. Thorndike [a,*], Per Carlbring [b], Frederick L. Smyth [c], Joshua C. Magee [a], Linda Gonder-Frederick [a], Lars-Göran Ost [d], Lee M. Ritterband [a]

[a] Department of Psychiatry and Neurobehavioral Sciences, Behavioral Health and Technology, University of Virginia Health System, P.O. Box 801075, Charlottesville, VA 22908, United States
[b] Department of Behavioral Sciences at Linkoping University, Sweden
[c] Department of Psychology, University of Virginia, United States
[d] Department of Psychology, Stockholm University, Sweden

ARTICLE INFO

Article history:
Available online 4 January 2009

Keywords:
Online
Web
Measurement
Assessment

ABSTRACT

The current study was conducted to determine whether participants respond differently to online questionnaires presenting all items on a single webpage versus questionnaires presenting only one item per page, and whether participants prefer one format over the other. Of participants seeking self-help treatment on the Internet (for depression, social phobia, or panic disorder), 710 completed four questionnaires (Beck Depression Inventory, Beck Anxiety Inventory, Quality of Life Index, Montgomery-Åsberg Depression Rating Scale) on the Internet on two occasions. The questionnaires were either presented with one questionnaire on one webpage (e.g., BDI on one webpage) or on multiple webpages (e.g., BDI on 21 webpages with one item each). Results suggest that the four web questionnaires measure the same construct across diagnostic group (depression, social phobia, panic), presentation type (single versus multiple items per page), and order of presentation (which format first). Within each diagnostic group, factor means for all questionnaires were equivalent across presentation method and time. Furthermore, factor means varied as expected across samples (e.g., depressed group scored higher on BDI), providing evidence of construct validity. The majority of participants in each diagnostic group preferred the single item per page format, even though this format required more time.

1. Introduction

The World Wide Web is now frequently being used to collect clinical and research data, particularly from individuals engaged in treatment delivered over the Internet. Online data collection presents several advantages over more traditional measurement methods, allowing investigators to expand the scale and scope of psychology research (Kraut et al., 2004). Web-based measurement allows for greater recruitment opportunities, increased automation and control, fewer data recording errors, improved efficiency, more flexibility (e.g., tailoring of questions based on previous responses), reduced costs, the ability to more easily construct and revise measures, and the ability to provide the user with prompt feedback (Barak & English, 2002; Kraut et al., 2004). Critics of Internet measurement, however, have raised a range of concerns, including low completion rates, sampling biases, uncertain data security, and questionable data quality (Kraut et al., 2004).

To address concerns about the validity of web measurement, researchers have started to examine whether data collected over the Internet can be compared with data collected via paper-and-pencil measures. Equivalence between Internet-based findings and paper-and-pencil findings has been documented with measures of panic/agoraphobia (Carlbring, Brunt, et al., 2007; Carlbring, Gunnarsdóttir, et al., 2007), self-monitoring (Buchanan & Smith, 1999), youth independence living (Bressani & Downs, 2002), emotional functioning and parenting attachment (Fouladi, McCarthy, & Moller, 2002), and general psychopathology (Vallejo, Jordan, Diaz, Comeche, & Ortego, 2007).

1.1. Best way to conduct web measurement

While there is some research to suggest that paper-and-pencil and online measures do not produce significantly different results (Buchanan, 2003), determining the best way to present online measures has yet to be established. Studies informed by cognitive processing theorists may offer some guidance, particularly studies focused on cognitive load theory (CLT; Sweller, 1988). CLT addresses how cognitive resources are used during learning and problem solving, making the assumption that working memory is limited and long-term memory is unlimited. Given the limitations of working memory, CLT posits that information should be

* Corresponding author. Tel.: +1 434 243 2704; fax: +1 434 924 0185.
E-mail address: fthorndike@virginia.edu (F.P. Thorndike).

0747-5632/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.chb.2008.05.006

presented in a way that does not exceed the limits of working memory.

After a series of experiments testing CLT in instructional design, Chandler and Sweller (1991) concluded that materials should be simplified (not integrated) when mental integration of the content is not necessary. This notion of simplifying presentation of materials to reduce cognitive load warrants consideration in the design of web materials. More specifically, integrating a web format that reduces extraneous load on working memory could have a beneficial effect on user experience.

Mayer and Moreno (2002) note that all multimedia messages deliver information but not all are equally successful in promoting understanding. Pollock, Chandler, and Sweller (2002) tested the effects of two presentations styles on learning: 1) presenting all information at once versus 2) presenting information broken into a series and then viewed all at once. For certain groups of learners, performance improved when the information was presented sequentially.

Ardac and Unal (2008) examined how web-design might impact learning, measuring whether the amount of on-screen text impacted student learning. They found that, although there were no group differences between whole-text and short-text versions, students with lower memory scores performed worse on the whole-text version than students with better memory scores. Interestingly, both groups (high and low memory) performed equally well on the short-text version. This highlights the need to consider both user characteristics and the task at hand when designing web materials.

1.2. Design of web surveys

Although cognitive processing studies offer guidance into how to construct online measures, few studies have been undertaken to actually evaluate the design of web-based surveys. It has been suggested that the design of a survey can affect the response rate, the dropout rate, and even the responses themselves (Couper, Traugott, & Lamias, 2001). More specifically, choosing between the single item per page versus multiple items per page designs could affect the quality of survey data, including measurement error, item non-response and partial non-response (Manfreda, Batagelj, & Vehovar, 2002). Dillman and Bowker (2001) argue that presenting one question per page results in a lack of context, forcing subjects to lose sight of the big picture. However, placing several items on a single page may force respondents to view the items as related entities, thereby increasing the correlations among them (Couper et al., 2001).

Data comparing various features of web survey design has only recently been found in the literature. Couper and colleagues (2001) embedded a series of design experiments in a survey of undergraduate attitudes about affirmative action, including a test of the effect of single versus multiple items per webpage. Although the correlations were consistently higher among the multiple items per page design, the differences between the correlations did not reach significance. They also found similar factor structure across the two versions (single item and multiple items per page). Lastly, they found users spent less time completing the measures in the multiple items per page format than the single item per page format. This has since been confirmed by Manfreda and colleagues (2002) who conducted a similar series of web experiments.

Although a few empirical studies have begun investigating the impact of web survey design on responses, there is still very limited information. Researchers have been calling for "a massive effort" to examine the psychometric properties of Internet-based psychological tests and understand the impact of web design (Barak & English, 2002). Some have described the issue of single item per page format versus multiple items per page as one of the central issues in web survey design, and although both designs have theoretical advantages and disadvantages, no conclusions have been reached (Clayton & Werking, 1988; Farmer, 1998; Kottler, 1997; Manfreda et al., 2002). Furthermore, no studies have surveyed and reported users' preferences.

1.3. Current study

The current study was conducted to determine whether participants respond differently to online questionnaires presenting all items on a single webpage (multiple items per page) versus measures illustrating only one item per webpage (single item per page). Although we did not intend to establish whether one format was better than another, we hoped to ascertain whether there was a response difference between administration types in a clinical population using well-established measures. In addition, the study sought to determine whether such differences, if they existed, varied according to the population measured (e.g., participants presenting for self-help treatment for depression versus social phobia) or the measure in question (e.g., BDI versus BAI). Lastly, information regarding subject preference was examined.

2. Method

2.1. Participants

Seven hundred and ten participants were recruited from a wait-list of people who had registered for one of three self-help treatment programs on the Internet (social phobia, panic, and depression). Of note, these participants self-referred for treatment and were not diagnosed with the target disorder. Participants were originally recruited from newspaper articles in Swedish national and regional papers, as well as the web pages of the Swedish Anxiety Association. At the time of recruitment, participants had been on the wait-list an average of 11.2 months ($SD = 2.8$). An e-mail outlining the study and inviting participation was randomly sent to approximately 30% of the wait-listed individuals. In exchange for answering four web questionnaires twice, participants were promised an empirically tested electronic self-help manual that was diagnosis specific (Andersson et al., 2005; Carlbring et al., 2006; Carlbring, Brunt, et al., 2007; Carlbring, Gunnarsdóttir, et al., 2007).

After the initial e-mail and two reminders, 268 out of 300 (89.3%) approached persons on the wait-list for social phobia treatment responded. For the social phobia group, the average age was 35.1 years ($SD = 11.2$), and the majority were female (56.0%). Among participants waiting for treatment for panic disorder, 201 out of the 220 (91.4%) approached persons responded, the average age was 35.2 years ($SD = 10.2$), and the majority were female (63.7%). In the depression wait-list group, 241 out of the 280 (86.1%) approached persons responded, the average age was 34.9 years ($SD = 10.9$), and the majority were female (61.0%). Only three people declined participation, whereas the rest did not respond.

2.2. Measures

Beck Depression Inventory. The Beck Depression Inventory (BDI) is a 21-item, forced-choice, self-report scale designed to assess current depressive symptoms (Beck & Steer, 1996). Each of the 21 items includes four statements, and respondents are asked to circle the statement that most reflects how they have been feeling during the past week. Scores on the BDI range from 0 to 63, with higher scores representing more depressive symptomatology. Total scores are labeled using the following scale: 0 – 9 = no depression or mild depression, 10 – 18 = mild to moderate depression, 19 – 29 = mod-

erate to severe depression, and > 30 = severe depression. In a sample of outpatients, the average BDI score was 18 for patients diagnosed with mild major depressive episode (MDE), 27 for patients diagnosed with a moderate MDE, and 34 for patients diagnosed with a severe MDE (Steer, Brown, Beck, & Sanderson, 2001). Extensive studies demonstrate both high internal consistency and high convergent validity with interviewer ratings of current depression symptomatology (Beck, Steer, & Garbin, 1988).

Beck Anxiety Inventory. The Beck Anxiety Inventory (BAI) is a 21-item self-report inventory designed to measure the severity of anxiety in psychiatric populations (Beck, Epstein, Brown, & Steer, 1988). Respondents rate how much they have been bothered by a list of symptoms during the past week on a four-point scale ranging from zero (not at all) to three (terrified). Like the BDI, scores on the BAI range from 0 to 63, with higher scores representing more anxiety symptomatology. In a normative sample, a BAI score of three fell at the 50th percentile, and a score of ten fell at the 80th percentile (Gillis, Haaga, & Ford, 1995). In contrast, the mean BAI score for a sample with panic disorder with agoraphobia was 27.27 (SD = 13.11) (Beck et al., 1988). The BAI has demonstrated high internal consistency in panic patients (Cronbach's α = .92) and good test-retest reliability (.83) (de Beurs, Wilson, Chambless, Goldstein, & Feske, 1997).

Quality of Life Inventory. The Quality of Life Inventory (QOLI) asks subjects to rate 16 domains in both importance (0-2) and satisfaction (-3 to +3) (Frisch, Cornell, Villanueva, & Retzlaff, 1992). These two ratings are then multiplied, yielding a score between -6 and +6 for each area, with higher scores representing a better quality of life. Total quality of life scores are then calculated by taking the average of scores across the domains judged as important or very important for life satisfaction. On the QOLI, Swedish outpatients with anxiety scored 0.84 (SD = 3.05), whereas normal controls scored 2.76 (SD = 2.29) (Ost, Breitholtz, & Thulin, 1997). Studies of the QOLI have documented both good internal consistency (between α = .77 and .89) and excellent one month test-retest reliability (between r = 0.80 and 0.91) (Frisch et al., 1992).

Montgomery Åsberg Depression Rating Scale. The Montgomery Åsberg Depression Rating Scale (MADRS) was designed to measure the severity of depressive symptoms, with the intention of being particularly sensitive to treatment effects (Svanborg & Åsberg, 1994). Although initially developed by Montgomery and Åsberg (1979), Svanborg and Åsberg (1994) later developed a self-assessment version of MADRS (the MADRS-S), which was utilized in this study. The nine item self-report checklist measures a range of symptoms, including mood changes, anxiety, changes in sleeping patterns, appetite, ability to concentrate, initiative-taking, emotional engagement, pessimism and attitude to life. Scores on the MADRS-S range from 0 to 54, with higher scores representing more depressive symptomatology. Among a group of depressed primary care patients, the average MADRS-S score before treatment was 28.0 (McIntyre et al., 2005). The instrument has been demonstrated to be reliable (Montgomery & Åsberg, 1979) and shows high correlations (from r = 0.80 to r = 0.94) between expert ratings and self-reports (Svanborg & Åsberg, 1994).

2.3. Procedure

All participants had registered for an Internet-based treatment program for depression, social phobia or panic disorder. A randomly selected portion of the individuals (approximately 30%) were asked if they wanted to take part in an Internet experiment that includes answering four questionnaires on the Internet on two separate occasions. After providing informed consent, participants were given a unique code and Internet link, which prompted them to complete the four questionnaires (BDI, BAI, MADRS and QOLI). The order of the questionnaires was randomized to balance for order effects. Participants were free to backtrack and review all item responses (and amend if desired) before submitting the questionnaires. After finishing the first battery, participants were instructed to return and answer the same questionnaires again no sooner than one hour and no later than four hours. The average time elapsed between the two sets of questionnaires was 3.2 hours.

Table 1
Group Assignment

Group	Presentation		
	Time 1	Time 2	Label
Depression			
1	S	M	SM
2	M	M	MM
3	M	S	MS
4	S	S	SS
Social Phobia			
5	S	M	SM
6	M	M	MM
7	M	S	MS
8	S	S	SS
Panic Disorder			
9	M	S	MS
10	S	M	SM

Note. S refers to single item per page presentation and M refers to multiple items per page presentation.

The questionnaires were either presented with one entire questionnaire on one web page (e.g., the BDI on one web page with 21 items – multiple items per page) or on multiple web pages (i.e., the BDI on 21 web pages with one item each – single item per page). Hence, the questionnaires were either answered on a total of four web pages (multiple items per page) or on 83 web pages (single item per page).

For purposes of counter balancing, participants in the social phobia and depression groups were divided into four equally large groups. Half the participants completed the multiple items per page questionnaires first and the single item per page last, or vice versa. The other half of the participants answered either the multiple items per page questionnaires twice, or the single item per page questionnaires twice. Participants in the panic disorder sample were only divided into two groups – first answering multiple items per page questionnaires, then single item per page, or vice versa. Hence, there were a total of ten different groups across the three diagnostic categories. See Table 1 for an illustration of the division of these groups.

There was no experimenter present during the administration, and all communication with participants was done remotely (by email). Participants took an average of 19.28 minutes (SD = 9.56) to complete the four questionnaires on the first occasion and 13.30 minutes (SD = 6.47) to complete them on the second occasion.

3. Results

3.1. Modeling Procedure

The main purpose of the current study was to examine whether participants responded differently according to the presentation format (single or multiple items per page) of several Internet questionnaires. As a secondary question, we were also interested in whether the responses varied according to the sample measured (i.e., participants seeking help for depression, social phobia or panic disorder).

After ensuring that there was a significant correlation between scores on the single item per page and multiple items per page formats (see Table 2), structural equation modeling (SEM) was used to compare responses across presentation formats and clinical samples. For each measure, a four-step approach was taken. First, the literature for each scale was reviewed to identify factor models proposed by other researchers based on paper-and-pencil administrations. Then, using a confirmatory factor analytic approach, a baseline model for each scale was determined by comparing the fits of the various proposed models in the relevant diagnostic group (e.g. for the BDI, only the depressed sample was included when comparing models). The best-fitting factor structure of each scale was selected to use as the baseline; if the fit of different proposed models did not differ for a scale, the one with the simplest factor structure was chosen.

Second, multiple group comparisons (McArdle & Hamagami, 1996) were conducted to evaluate factorial invariance, i.e., whether the questionnaires were measuring the same construct regardless of presentation format (single versus multiple items per page) or order of presentation (whether participants received single or multiple items per page first or second). This was accomplished by testing misfit in the models and constraining factor loadings to be equivalent across the presentation format and order of presentation groups. Significant misfit would have indicated that scale items were not functioning in the same way across groups and render any cross-group comparison of factor means and correlations meaningless. This step tested the primary question of whether the scale constructs being measured remained comparable across 1) presentation format and 2) order of presentation (timing) within the relevant diagnostic group (e.g., depression scale tested with sample of participants seeking treatment for depression).

If constructs were deemed comparable across format and timing, a third step compared factor means across format and order of presentation within relevant diagnostic groups. Factor means were the basis for comparisons (rather than summed overall scores) because factor means reduce measurement error compared to overall summed scores; that is, factor means are more 'clean' representations of the population means. Comparing factor means also capitalized on the existing factor analytic literature for each scale, which in all cases suggested more than one construct underlying the scale (rather than unnecessarily assuming that each scale was one-dimensional). The tests of factor means addressed whether the means of the construct in the relevant group (e.g. depressed participants on the BDI) varied according to the presentation format or order of presentation.

Table 2
Correlations between single item and multiple items per page administrations

	N	% Female	% prefer single	BDI *r*	BAI *r*	MADRS-S *r*	QOLI *r*
Depression							
SM	61	57	92	0.99	0.97	0.97	0.97
MS	57	56	93	0.98	0.96	0.98	0.97
MM	65	60	94	0.96	0.96	0.97	0.98
SS	58	70	100	0.96	0.99	0.97	0.97
Panic Disorder							
SM	103	59	79	0.97	0.97	0.95	0.98
MS	98	68	72	0.98	0.96	0.96	0.95
Social Phobia							
SM	69	58	67	0.98	0.94	0.95	0.97
MS	69	57	75	0.97	0.97	0.94	0.97
MM	64	50	73	0.98	0.97	0.98	0.98
SS	66	59	77	0.97	0.98	0.98	0.98

Note. p value for each correlation is < .001. BDI = Beck Depression Inventory; BAI = Beck Anxiety Inventory; MADRS-S = Montgomery Åsberg Depression Rating Scale – Self Report; QOLI = Quality of Life Inventory.

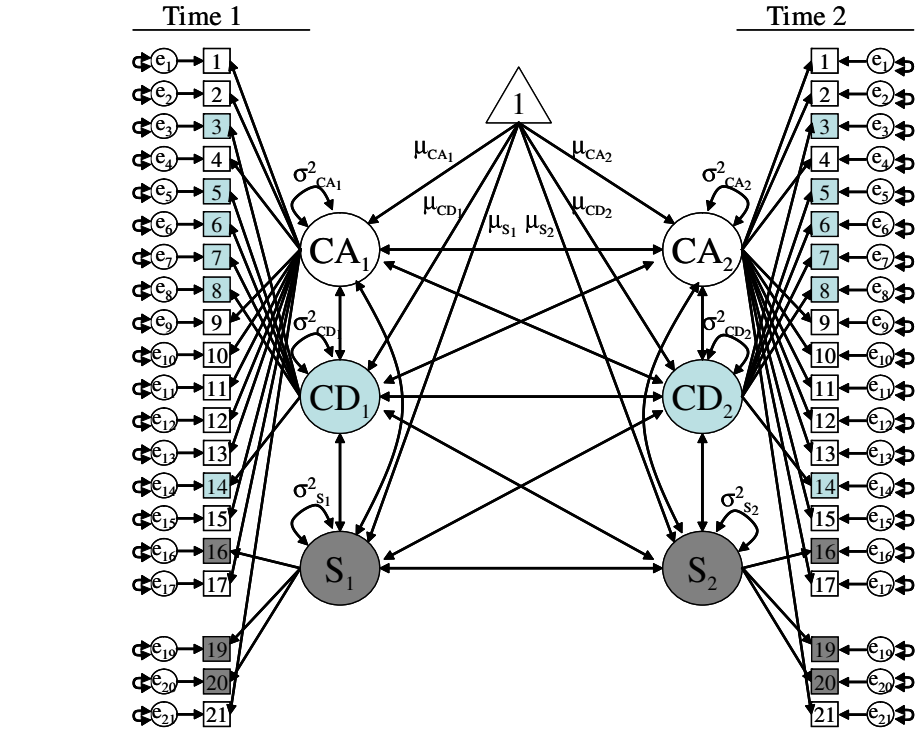
In the fourth and final step, a secondary analysis was conducted, extending the multiple group comparisons of factor means across all three diagnostic groups (depression, social phobia, panic) as another variable. All models were fit to the data using *Mplus* structural equation modeling software (Muthen & Muthen, 2001) and full information maximum likelihood methods.

BDI. For purposes of illustration, each step of the analytic plan will be outlined in detail for the BDI and then the results of this same procedure will be summarized for the remaining measures. In the first step, an oblique (correlated) three factor model (Steer, Beck, Riskind, & Brown, 1987) fit the data better than a two-factor model (Enns, Cox, Parker, & Guertin, 1998) or a baseline one-factor model. The model was modified by dropping item 18 from all analyses, because the lack of variance caused convergence problems for the SEM program —between 70-80% of all depressed group members indicated they had not experienced a change in appetite. Fig. 1 shows an illustration of the resulting model that was the starting point for subsequent comparisons using the depressed group.

Next, the questionnaire was tested to determine if it measured the same construct regardless of presentation format or order of presentation. The pattern of loadings of individual items on the three factors of the BDI was constrained to be equivalent across the four combinations, referred to as MM, SS, MS, or SM (see Table 1 for explanation of labels) (Ahmavaara, 1954; Horn & McArdle, 1992; Meredith, 1964; Thurstone, 1947). If the misfit introduced by these constraints was relatively small, it demonstrated that the assumption of factorial (construct measurement) invariance was justified, i.e., factor loading patterns were essentially the same. To judge model fit, Browne's FITMOD statistical program (Browne, 1992) was used to compare the fits of nested structural equation models. The 95% confidence interval around the *root-mean-square error of approximation* (RMSEA or ε_a) *of the change* (Δ) in model fit was assessed to see whether it included the .05 value commonly used to indicate a "close" fit (MacCallum, Browne, & Sugawara, 1996). If the 95% CI around $\varepsilon_a\Delta$ included (or was lower than) .05, the two model fits were judged to be statistically close to one another, indicating that the scale was functioning to measure the same construct regardless of presentation format or order. The results for the BDI indicated that the same construct held across the four depression groups (MM, SS, MS, SM), 95% CI $\varepsilon_a\Delta$ = .00-.05. Thus, we concluded that the constructs were equivalent across the presentation format and order of presentation in the BDI for the depression diagnostic group.

Given that the constructs were deemed equivalent, factor means were then evaluated to determine whether there was an effect for presentation format or order of presentation. This was accomplished by constraining estimates of the BDI factor means to be equal across the four groups (MM, SS, MS, SM) and evaluating the change in fit relative to the baseline model in which means were freely estimated for each group. Constraining the means to be equivalent did not add significant misfit according to presentation format or order of presentation, 95% CI $\varepsilon_a\Delta$ = .00-.08. Thus, we concluded that the presentation format and order of presentation had no effect on the BDI factor means of depressive symptoms.

This analysis was extended to the other diagnostic groups (participants seeking treatment for social phobia or panic disorder) by fitting the same progression of models but also testing for factorial invariance across diagnostic group. These tests used ten distinct combinations of presentation format and order of presentation for the diagnostic groups (see Table 1 for breakdown of these diagnostic groups). Constraining the pattern of loadings for items across presentation format, order of presentation, and diagnostic group did not add significant misfit, since the 95% CI around the RMSEA of the change was below or encompassed .05 in each com-



Note. CA = cognitive-affective & performance difficulties, CD = cognitive distortions, S = somatic complaints. Square = observed variable, circle = latent variable, triangle = a constant allowing estimates of factor means, e = unique residual (which are allowed to correlate across occasion in all models).

Fig. 1. Diagram of the 3-factor model for the BDI data for both measurement times, based on Steer et al. 1987.

parison. Therefore, we concluded that the same construct was being measured across the ten diagnostic groups.

Next, the factor means were constrained to be equivalent *within* diagnostic groups, yielding results that fell within the $\varepsilon_a\Delta = .05$ range. This confirmed that the BDI factor means were equivalent across presentation format and order of presentation within each diagnostic group.

Finally, when factor means were evaluated *between* diagnostic groups, differences were found. The BDI factor means for participants seeking treatment for depression were significantly higher than those seeking treatment for panic disorder (95% CI $\varepsilon_a\Delta = .15$-$.23$), and the means for the panic disorder groups were significantly higher than those in social phobia group (95% CI $\varepsilon_a\Delta = .09$-$.17$). Table 3 lists the total means and standard deviations for each group.

In sum, the results suggested that the BDI web-questionnaire measured the same construct across diagnostic groups (depression, social phobia, panic), presentation format (single item per page versus multiple items per page), and order of presentation (which format appeared first). As expected, participants in the depression diagnostic group had higher factor means than those in the panic or social phobia groups, providing evidence for the construct validity of the Internet measure.

BAI. Initial analyses of the BAI were conducted only with participants from the panic disorder group. The higher-order factor structure suggested by Osman et al. (2002)—four first-order factors

Table 3
Total means (and SD) for each group

Group	BDI		BAI		MARS-S		QOLI	
	Total Mean		Total Mean		Total Mean		Total Mean	
	Time 1	Time 2	Time 1	Time 2	Time 1	Time 2	Time 1	Time 2
Depression								
SM	26.8 (11.7)	25.8 (11.8)	20.4 (13.1)	20.6 (12.5)	25.0 (9.9)	25.3 (9.9)	−0.3 (1.9)	−0.4 (2.0)
MS	25.2 (10.9)	25.2 (11.4)	21.3 (10.1)	18.9 (10.7)	25.2 (9.2)	24.5 (9.6)	−0.9 (1.6)	−1.0 (1.7)
MM	24.0 (10.8)	23.6 (11.2)	19.2 (8.9)	18.4 (9.6)	24.5 (9.2)	24.4 (7.2)	−0.03 (1.9)	−0.4 (1.9)
SS	26.2 (8.2)	25.1 (8.4)	19.3 (10.5)	18.4 (10.2)	24.4 (24.4)	24.0 (7.2)	−1.0 (1.6)	−1.0 (1.6)
Panic Disorder								
SM	17.9 (9.1)	17.0 (9.0)	20.7 (11.0)	21.1 (10.3)	17.9 (8.1)	17.7 (8.6)	0.7 (1.8)	0.7 (1.9)
MS	16.8 (9.5)	16.5 (9.1)	23.5 (9.9)	21.4 (10.3)	19.3 (8.4)	17.9 (8.2)	0.8 (1.4)	0.7 (1.4)
Social Phobia								
SM	15.1 (8.9)	14.3 (8.2)	14.5 (7.1)	15.2 (6.7)	14.7 (7.6)	15.0 (7.7)	0.6 (1.5)	0.5 (1.6)
MS	14.6 (7.4)	13.9 (7.4)	17.8 (8.2)	15.3 (8.5)	17.2 (6.9)	15.7 (6.6)	0.3 (1.6)	0.2 (1.6)
MM	16.0 (9.1)	15.9 (9.9)	20.5 (7.9)	19.9 (7.7)	17.8 (9.0)	16.8 (9.1)	0.4 (1.9)	0.4 (1.9)
SS	15.2 (8.6)	15.1 (8.4)	16.4 (8.4)	15.5 (9.0)	15.8 (8.2)	14.8 (7.6)	0.6 (1.8)	0.5 (1.8)

Note. S refers to single item per page presentation and M refers to multiple items per page presentation. BDI = Beck Depression Inventory; BAI = Beck Anxiety Inventory; MADRS-S = Montgomery Åsberg Depression Rating Scale – Self Report; QOLI = Quality of Life Inventory.

(neurophysiological, subjective, autonomic, and panic) and one second-order factor—provided a better fit to the data than the two- and four-factor alternatives (Beck et al., 1988; Hewitt & Norton, 1993; Steer, Ranieri, Beck, & Clark, 1993).

Once again, comparisons of the planned sequence of increasingly constrained models demonstrated that the constructs being measured, as well as the factor means for the BAI factors, did not differ among the two panic groups (MS and SM). Because there were only two panic groups, presentation format and order of presentation were confounded for this initial analysis. However, given that the constructs and factor means were equivalent across these combinations, it is unlikely that either significantly affected the measurement.

When including all three diagnostic groups (depression, social phobia, panic disorder) for the BAI, construct measurement was equivalent regardless of presentation format or order of presentation. Factor means were again equivalent *within* diagnostic groups, but differed *between* diagnostic groups, supporting the construct validity of the measure. In evaluation of the panic factor on this scale, the group seeking treatment for panic disorder scored significantly higher than the group seeking depression treatment (95% CI $\varepsilon_a\Delta$ = .08-.22), but the depression group was not significantly different than the social phobia group (95% CI $\varepsilon_a\Delta$ = .04-.18); factor means for the panic factor were .50, .20, and .04, respectively, for the panic, depression, and social phobia diagnostic groups.

In sum, the BAI web-questionnaire functioned similarly, i.e., measured the same constructs across differences in diagnostic group (depression, social phobia, panic), presentation format (single item per page versus multiple items per page), and order (single versus multiple items per page first). The group seeking treatment for panic disorder scored higher on the panic factor mean than the other two diagnostic groups.

QOLI. A search of the literature did not yield clear factor analytic specifications for the QOLI and attempts to fit even a one-factor model across more than one treatment group at a time ended with program non-convergence. These problems appeared to result from measurement properties of the QOLI, which include missing items by design (i.e., any item receiving an importance rating of zero is treated as missing). Instead of comparing factor models, multiple-group structural regression models for the total QOLI scores were estimated, with age and gender as covariates. Similar to the previous approaches, increasing levels of constraints were systematically applied to test the effects of presentation format and treatment group. With this approach it was not possible to test the assumption that the same construct was being measured across diagnostic groups, but, if this assumption were made, it would be possible to examine whether QOLI composite means differed according to the key variables.

In the unconstrained (saturated) baseline model, 20 composite scores were estimated, one for each of the two measurements for the ten combinations (four depression, four social phobia, two panic). To test whether composite means differed as a function of presentation format or order, the fit of this baseline model was compared with one in which total score estimates were constrained to be equal within diagnostic group. As evidenced by the *RMSEA*-of-the-change 95%CI interval (.00-.04), the difference in fit between these models was trivial. Thus, QOLI total score means were judged to have not differed as a function of either presentation format or order of presentation. Means did differ once again, however, when comparing diagnostic groups. Specifically, model comparisons confirmed that the mean of the depression group was significantly lower—poorer quality of life—than those of the panic and social phobia groups, while the latter two did not differ from one another.

In sum, the QOLI web-questionnaire showed the same total scores across differences in presentation format (single item per page versus multiple item per page) and order (single versus multiple items per page first or second). The total scores showed that the group seeking depression treatment had the lowest quality of life scores, and the panic and social phobia groups reported similar quality of life scores.

MADRS-S. For the MADRS-S, the three-factor model specified by Parker, Flint, Bosworth, Pieper, and Steffens (2003) proved superior to the three-factor models offered by Suzuki et al. (2005) and Benazzi (2001), as well as to a one-factor specification.

The results were the same as with the BDI; the same construct was being measured regardless of presentation format, order of presentation, or group, as all results fell within the $\varepsilon_a\Delta$ = .05 range. Also, the factor means did not differ across presentation format, order of presentation, or *within* the diagnostic groups.

Comparison of the factor means between diagnostic groups revealed that participants in the group seeking social phobia treatment had the lowest means for each of the three factors, although they did not differ significantly from the means for the panic group (95% CI of the change included .05). Participants in the group seeking depression showed significantly higher factor means than the group seeking panic treatment (95% CI of the change = .13-.22). By inference, the factor means for the participants seeking depression treatment were higher than both panic and social phobic groups.

In sum, the MADRS-S web-questionnaire functioned similarly across differences in diagnostic treatment groups (depression, social phobia, panic), presentation type (single item per page versus multiple items per page), and order of presentation (single versus multiple items per page first or second). Comparison of factor means showed that the group seeking treatment for depression had higher MADRS-S means than either of the other groups, while the mean for participants seeking treatment for panic disorder did not differ significantly from those seeking treatment for social phobia.

Table 4 provides a summary of the results detailed above for each of the four measures, illustrating whether each measure shows: 1) construct equivalence across presentation format and order of presentation, 2) factor mean equivalence within diagnostic groups, and 3) factor mean equivalence between diagnostic groups.

3.2. Preference for Single Item or Multiple Items per Page

As shown in Table 2, a majority of participants in each of the ten diagnostic groups preferred the single item per page presentation format. This was true even for the two groups, one among the depression group and one among the social phobia group, which received questionnaires only in the multiple items per page format. This preference was especially pronounced among the participants in the depressed group, with over 90% in each group (MM, SS, MS, SM) preferring one item to a page.

Table 4
Results summary for each measure

	BDI	BAI	MADRS-S
Construct equivalence across presentation format and order of presentation	Yes	Yes	Yes
Factor mean equivalence within diagnostic groups	Yes	Yes	Yes
Factor mean equivalence between diagnostic groups	No	No	No

Note. BDI = Beck Depression Inventory; BAI = Beck Anxiety Inventory; MADRS-S = Montgomery Åsberg Depression Rating Scale – Self Report. Construct equivalence was tested within relevant diagnostic group (BDI tested with depression group, BAI tested with panic disorder group, and MADRS-S tested with depression group). The Quality of Life Inventory is not included in this table because it was analyzed using multiple-group structural regression models rather than structural equation modeling due to problems with non-covergence.

We tested group differences in preference by employing the same multiple group structural model comparison approach we used in the previous analyses, allowing us to conduct logistic regressions for preference (coded 0 for preferring multiple items per page, and 1 for single item per page). Our baseline (saturated) model included age and gender as control covariates of the preference and allowed the preference proportion to be freely estimated for each group (the SS group of depressed participants was excluded, since they had no variance in the outcome, all preferring the single item method). In a second model we constrained the preference proportion to be equal across presentation format within each diagnostic group, and found that the loss of fit compared with the baseline was trivial (95% CI $\varepsilon_a\Delta$ = .00-.06). This indicates that there were no differences in preference as a function of presentation format within the three diagnostic groups. In a third model we tested whether preference differed between the social phobia and panic treatment groups by constraining the proportion to be equal for both. As evidenced by the CI around RMSEA of the change relative to the previous model, .00-.06, these groups did not differ significantly in their preference for the single item format (about 70% with gender and age held constant). Finally, we constrained the preference proportion to be equal across all three diagnostic groups as a test of whether the participants seeking depression treatment differed from the other two groups. The difference in fit of this model, in comparison to the model constraining the social phobia and panic diagnostic groups to be equal, yielded a 95% CI $\varepsilon_a\Delta$ = .12-.27. Thus, the depression group's estimated 92% preference for the single item format (not including the SS group which was 100%), was significantly higher than that of the other two groups.

Interestingly, participants preferred the single item per page format despite it requiring more time than the multiple item per page format. On the first occasion of measurement, participants took an average of 20.14 minutes (SD = 10.21) to complete the single item per page format, whereas participants took an average of 18.41 (SD = 8.78) minutes to complete the multiple-item-per-page format, F (1, 708) = 5.80, p = .02. On the second occasion of measurement, participants took an average of 13.99 minutes (SD = 6.86) to complete the single item per page format, whereas participants took an average of 12.63 (SD = 6.01) minutes to complete the multiple item per page format, F (1, 708) = 7.96, p = .01.

4. Discussion

In sum, we found validity for the Internet administration of four measures whether using the single item per page format or the multiple items per page format. There were extremely high correlations between scores on single item and multiple items per page presentations, and each of the four web questionnaires (BDI, MADRS-S, BAI, QOLI) functioned similarly across different methods of presentation (single item per page versus multiple items per page), times (Time 1 versus Time 2), and clinical samples (those seeking treatment for depression, social phobia, or panic disorder). *Within* each of the three diagnostic groups, factor means (or composite mean in the case of the QOLI) for all scales were equivalent across presentation format and time. As could reasonably be expected, however, there were differences in factor means *across* diagnostic group (i.e., those seeking treatment for depression versus treatment for social phobia).

Participants seeking treatment for depression endorsed more depressive symptomatology on two self-reports of depression (BDI, MADRS-S) than participants seeking treatment for either panic disorder or social phobia. Participants seeking panic treatment scored higher on a panic factor (BAI) than either the depression or social phobia groups. Finally, the group seeking depressed treatment scored lower on a measure of quality of life (QOLI) than did the panic or social phobia group, while the latter two did not differ from one another. In sum, mean differences suggest construct validity.

The majority of participants preferred the single item per page presentation format. This was true despite the increase in time required by the single item per page format. The preference for the single item per page presentation was especially pronounced among the participants seeking treatment for depression, with over 90% in each group (MM, SS, MS, SM) preferring one item to a page. This may be explained, at least in part, by cognitive load theory (CLT) which suggests that the increased cognitive load generated by presenting unnecessary or irrelevant items impedes learning. To apply this to our example, grouping multiple items on a page may force the user to integrate these items, which takes additional time and effort. Grouping these items, however, is not intrinsically necessary, as each item can stand on its own. This fits with Pollock et al. (2002) findings that, at least for some groups of learners, information is better learned by breaking up information (and creating discrete units), which allows users to process information serially.

We offer no evidence to indicate that one presentation format should be used over the other. Instead, we argue that the formats yield comparable results and suggest that researchers and survey designers consider participant preference, along with other factors, when designing surveys. With that recommendation, however, three significant warnings should be considered.

First, the preference for single items presented on each web page may not apply across all samples. Perhaps this presentation format is preferable for individuals who endorse significant symptoms of depression, panic, social phobia, or other psychopathology. It seems plausible that users with psychiatric symptoms might benefit from the more straightforward task of answering one item per web page, willingly trading time for this approach. Other samples, however, may prefer the multiple items per web page approach. For example, it seems plausible that certain samples (e.g., normative controls) who prioritize time may favor the quicker method. This would fit with research from Ardac and Unal (2008) who found that both the nature of the task and characteristics of the learner needed to be considered when determining web design.

Second, given the findings from this study, we caution readers not to confuse preference with actual behavior. One could hypothesize that matching the format to the user's preference would result in changed outcomes. In this study, for example, we could have speculated that giving users the format they prefer would lead to more accurate data or greater completion rates, but we have no data to indicate how preference impacts behavior. It is possible that users prefer the single item format but are equally as likely to complete the questionnaire whether it is presented in the single item format or multiple items per page format. Just because subjects prefer the single item format does not necessarily mean that using this method of presentation will affect the quality of survey data, including response rates and measurement error.

Third, there may be a number of moderating variables that affect preference. It seems plausible that users may prefer different formats under different circumstances. For example, when completing surveys in a single sitting (e.g., pre-treatment or post-treatment questionnaires), perhaps users would be less concerned about time and prefer the single item format. When completing daily measures, however, a task the user is familiar with and has to do regularly, time may become more important, pushing the user to favor the multiple items per page, less timely approach. In short, we agree with Couper and colleagues who suggested that web survey designers consider the task at hand, rather than

arguing for one approach over another for all applications (Couper et al., 2001).

4.1. Limitations

There are several limitations to the current study that need to be noted. First, the data were obtained from a group of participants waiting to seek online treatment for one of three psychiatric disorders (depression, social phobia, panic). These participants self-referred for treatment and were not diagnosed with the target disorder. Thus, participants in each diagnostic group may not meet criteria for that disorder or may meet criteria for multiple disorders. For example, it is possible that a participant waiting to receive self-help treatment for social phobia would have been more appropriately diagnosed with panic disorder. Review of questionnaire means, however, suggests that these participants, particularly the group waiting for depression treatment, self-reported more symptoms than most wait-list controls. In short, as with many studies, the findings may not generalize to other populations, particularly healthy controls.

Second, the nature of the study (web measurement) prevented tighter experimenter control over the context in which the data was collected. It is possible that non-serious responders affected the quality of the data. Although this is possible, the findings of construct validity (i.e., that there were meaningful factor mean distinctions between diagnostic groups) argues against the notion that non-serious responders significantly impacted the data. It seems more plausible that the high motivation of these participants might yield uncharacteristically stable results.

Third, the process for identifying a baseline structural measurement model for each scale was limited by the existing literature concerning that scale. It is possible that alternative models not found in the literature would have provided better fits to these data. Fourth, methodological constraints (e.g., short timing between questionnaires) also prevent drawing more robust conclusions.

4.2. Clinical Implications and Future Directions

Despite the above limitations, the current study extends our knowledge about the validity of capturing data over the Internet and provides information about how to design online measures. Results suggest that the two methods (single item versus multiple items per webpage) are comparable and that researchers and survey designers should consider preference, along with other factors, when designing surveys. Additional web design experiments, however, are needed with different populations (e.g., normal controls), different tasks (e.g., daily symptom diaries versus one-time assessments), and different measures (e.g., non-psychiatric questionnaires).

An important next step is to investigate whether preference for a particular presentation method impacts behavior; for example, examining whether users who complete questionnaires in the format they prefer are more likely to finish surveys. If it turns out that preference impacts behavior (e.g., if users get their preferred method, they are more likely to complete the survey), we should consider tailoring our measures to user preference as the power of the Internet makes this possible.

In conclusion, web surveys offer researchers a greater range of options for data collection than more traditional methods. The rich multimedia capabilities of the web (e.g., audio, animation, illustrations, interaction, personalization, video) allow researchers to manipulate multiple variables to help improve comprehension of the measure, motivate the user, and increase completion rates. But, before taking advantage of this technology it is important to better understand how these features impact the quality of responses. Researchers should also be mindful of the potential problems associated with web measurement, considering how factors like lack of access, lack of familiarity with technology, bandwidth, language, and security could all interfere with the administration of a questionnaire. Although technology and the Internet are sophisticated tools with multiple capabilities, this field of research is still in its infancy. In sum, it is necessary to better understand how the increased opportunities associated with web measurement affect data, using empirical investigations to further the science behind web measurement.

Acknowledgements

The authors thank Emily Cumming and Drew Saylor for their support with literature reviews and manuscript preparation.

References

Ahmavaara, Y. (1954). The mathematical theory of factorial invariance under selection. *Psychometrika, 19*, 27–38.

Andersson, G., Bergström, J., Holländare, F., Carlbring, P., Kaldo, V., & Ekselius, L. (2005). Internet-based self-help for depression: Randomised controlled trial. *British Journal of Psychiatry, 187*(5), 456–461.

Ardac, D., & Unal, S. (2008). Does the amount of on-screen text influence student learning from a multimedia-based instructional unit? *Instructional Science, 36*, 75–88.

Barak, A., & English, N. (2002). Prospects and limitations of psychological testing on the internet. *Journal of Technology in Human Services, 19*(2/3), 65–89.

Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology, 56*(893), 897.

Beck, A. T., & Steer, R. A. (1996). *Beck Depression Inventory manual, Svensk version (Swedish version)*. Fagernes, Norway: Psykologiforlaget AB.

Beck, A. T., Steer, R. A., & Garbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review, 8*, 77–100.

Benazzi, F. (2001). Factor analysis of the Montgomery Asberg Depression Rating Scale in 251 Bipolar II and 306 unipolar depressed outpatients. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 25*(7), 1369–1376.

Bressani, R., & Downs, A. (2002). Youth independent living assessment: Testing the equivalence of web and paper/pencil versions of the Ansell-Casey Life Skills Assessment. *Computers in Human Behavior, 18*(4), 453.

Browne, M. W. (1992). *FITMOD: Point and interval estimates of measures of fit of a model [computer software]*.

Buchanan, T. (2003). Internet-based questionnaire assessment: Appropriate use in clinical contexts. *Cognitive Behaviour Therapy, 32*(3), 100–109.

Buchanan, T., & Smith, J. (1999). Using the internet for psychological research: Personality testing on the World Wide Web. *British Journal of Psychology, 90*(1), 125.

Carlbring, P., Bohman, S., Brunt, S., Buhrman, M., Westling, B. E., Ekselius, L., et al. (2006). Remote treatment of panic disorder: A randomized trial of internet-based cognitive behavior therapy supplemented with telephone calls. *American Journal of Psychiatry, 163*(12), 2119–2125.

Carlbring, P., Brunt, S., Bohman, S., Austin, D., Richards, J., Öst, L., et al. (2007). Internet vs paper and pencil administration of questionnaires commonly used in panic/agoraphobia research.. *Computers in Human Behavior, 23*(3), 1421–1434.

Carlbring, P., Gunnarsdóttir, M., Hedensjö, L., Andersson, G., Ekselius, L., & Furmark, T. (2007). Treatment of social phobia: Randomised trial of internet-delivered cognitive-behavioural therapy with telephone support. *British Journal of Psychiatry, 190*(2), 123–128.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293–332.

Clayton, R. L., & Werking, G. S. (1988). Business surveys of the future: The World Wide Web as a data collection methodology. In M. P. Couper et al. (Eds.), *Computer Assisted Survey Information Collection* (pp. 543–562). New York: John Wiley & Sons.

Couper, M., Traugott, M., & Lamias, M. (2001). Web survey design and administration. *Public Opinion Quarterly, 65*, 230–253.

de Beurs, E., Wilson, K. A., Chambless, D. L., Goldstein, A. J., & Feske, U. (1997). Convergent and divergent validity of the Beck Anxiety Inventory for patients with panic disorder and agoraphobia. *Depression and Anxiety, 6*(4), 140–146.

Dillman, D. A., & Bowker, D. K. (2001). The web questionnaire challenge to survey methodologists. In M. Bosnjak & U. Reips (Eds.), *Dimensions of Internet Science* (pp. 159–178). Lengerich: Pabst Science Publishers.

Enns, M. W., Cox, B. J., Parker, J. D. A., & Guertin, J. E. (1998). Confirmatory factor analysis of the Beck Anxiety and Depression Inventories in patients with major depression. *Journal of Affective Disorders, 47*(1), 195–200.

Farmer, T. (1998). *Using the internet for primary research data collection*, from <http://www.researchinfo.com/library/infotek/index.html> Retrieved 28.04.99.

Fouladi, R., McCarthy, C., & Moller, N. (2002). Paper-and-pencil or online? Evaluating mode effects on measures of emotional functioning and attachment. *Assessment, 9*(2), 204.

Frisch, M. B., Cornell, J., Villanueva, M., & Retzlaff, P. J. (1992). Clinical validation of the Quality of Life Inventory. A measure of life satisfaction for use in treatment planning and outcome assessment. *Psychological Assessment, 7*, 450–455.

Gillis, M. M., Haaga, D. A. F., & Ford, G. T. (1995). Normative values for the Beck Anxiety Inventory, fear questionnaire, Penn state worry questionnaire, and social phobia and anxiety inventory. *Psychological Assessment, 7*, 450–455.

Hewitt, P. L., & Norton, G. R. (1993). The Beck Anxiety Inventory: A psychometric analysis. *Psychological Assessment, 5*(4), 408–412.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research, 18*(3–4), 117–144.

Kottler, R. E. (1997). *Search no further. Yahoo!'s audience analysis project – A case study*., from <http://www.spss.com/spssmr/web_bureau/knowledge/yahooqq.htm> Retrieved 21.11.01.

Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J., & Couper, M. (2004). Psychological research online: Report of board of scientific affairs' advisory group on the conduct of research on the internet. *American Psychologist, 59*(2), 105.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130–149.

Manfreda, K. L., Batagelj, Z., & Vehovar, V. (2002). Design of web survey questionnaires: Three basic experiments. *Journal of Computer-Mediated Communication, 7*(3).

Mayer, R., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and Instruction, 12*, 107–119.

McArdle, J. J., & Hamagami, F. (1996). Multilevel models from a multiple group structural equation perspective. In G. Marcoulides & R. Schumacker (Eds.), *Advanced Structural Equation Modeling Techniques* (pp. 89–124). Hillsdale, NJ: Erlbaum.

McIntyre, R. S., Konarski, J. Z., Mancini, D. A., Fulton, K. A., Parikh, S. V., Grigoriadis, S., et al. (2005). Measuring the severity of depression and remission in primary care: Validation of the HAMD-7 scale. *CMAJ: Canadian Medical Association Journal = journal de l'Association medicale canadienne, 173*(11), 1327–1334.

Meredith, W. (1964). Notes on factorial invariance. *Psychometrika, 29*, 177–185.

Montgomery, S. A., & Åsberg, M. (1979). A new depression scale designed to be sensitive to change. *British Journal of Psychiatry, 134*, 382–389.

Muthen, L. K., & Muthen, B. O. (2001). *Mplus user's guide* (2nd ed.). Los Angeles, CA: Muthen & Muthen.

Osman, A., Hoffman, J., Barrios, F. X., Kopper, B. A., Breitenstein, J. L., & Hahn, S. K. (2002). Factor structure, reliability, and validity of the Beck Anxiety Inventory in adolescent psychiatric in patients. *Journal of Clinical Psychology, 58*(4), 443–456.

Ost, L. G., Breitholtz, E., Thulin, U. (1997). *Quality of life in anxiety disorder patients.* Unpublished manuscript.

Parker, R. D., Flint, E. P., Bosworth, H. B., Pieper, C. F., & Steffens, D. C. (2003). A three-factor analytic model of the MADRS in geriatric depression. *International Journal of Geriatric Psychiatry, 18*(1), 73–77.

Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and Instruction, 12*, 61–86.

Steer, R. A., Beck, A. T., Riskind, J. H., & Brown, G. (1987). Relationships between the Beck Depression Inventory and the Hamilton psychiatric rating scale for depression in depressed outpatients. *Journal of Psychopathology and Behavioral Assessment, 9*(3), 327–339.

Steer, R. A., Brown, G. K., Beck, A. T., & Sanderson, W. C. (2001). Mean Beck Depression Inventory-II scores by severity of major depressive episode. *Psychological Reports, 88*(3 Pt 2), 1075–1076.

Steer, R. A., Ranieri, W. F., Beck, A. T., & Clark, D. A. (1993). Further evidence for the validity of the Beck Anxiety Inventory with psychiatric outpatients. *Journal of Anxiety Disorders, 7*(3), 195–205.

Suzuki, A., Aoshima, T., Fukasawa, T., Yoshida, K., Higuchi, H., Shimizu, T., et al. (2005). A three-factor model of the MADRS in major depressive disorder. *Depression and Anxiety, 21*(2), 95–97.

Svanborg, P., & Åsberg, M. (1994). A new self-rating scale for depression and anxiety states based on the Comprehensive Psychopathological Rating Scale. *Acta Psychiatrica Scandinavica, 89*(1), 21–28.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257–285.

Thurstone, L. L. (1947). *Multiple-factor analysis: A development and expansion of the vectors of mind*. Chicago: The University of Chicago Press.

Vallejo, M., Jordan, C., Diaz, M., Comeche, M., & Ortego, J. (2007). Psychological assessment through the Internet: A reliability and validity study of online (vs. paper-and-pencil versions of the General Health Questionnaire-28 (GHQ-28) and the Symptoms Check-list-90-Revised (SCL-90-R). *Journal of Medical Internet Research 9(1),* e2, from <http://www.jmir.org/2007/1/e2/> Retrieved 28.04.07.