Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.



Volume 17, Number 7, April 2012 ISSN 1531-7714

# Exploring Item Order in Anxiety-Related Constructs: Practical Impacts of Serial Position

R. Nicholas Carleton, *University of Regina, Regina, Saskatchewan, Canada*
Michel A. Thibodeau, *University of Regina, Regina, Saskatchewan, Canada*
Jason W. Osborne, *Old Dominion University, Norfolk, Virginia, USA*
Gordon J. G. Asmundson, *University of Regina, Regina, Saskatchewan, Canada*

The present study was designed to test for item order effects by measuring four distinct constructs that contribute substantively to anxiety-related psychopathology (i.e., anxiety sensitivity, fear of negative evaluation, injury/illness sensitivity, and intolerance of uncertainty). Participants (n = 999; 71% women) were randomly assigned to complete measures for each construct presented in one of two modalities: (a) items presented cohesively as measures or (b) items presented randomly interspersed with one another. The results suggested that item order had a relatively small impact on item endorsement, response patterns, and reliabilities. The small impact was such that item order appears unlikely to influence clinical decisions related to these measures. These findings not only have implications for these and other similar measures, but further inform a long-standing debate about whether item grouping is a substantial concern in measurement.

There has been long-standing recognition that features of questionnaire presentation can readily influence the reactions of respondents (Marsden & Wright, 2010). Item order is a particularly important questionnaire feature that can impact responses in several ways (Lam, Green, & Bordignon, 2002; Schwarz, 1995). Respondents may (a) become confused at being asked redundant questions; b) alter their answers due to question context (e.g., being asked if they would ever commit suicide after being asked if they are feeling down); c) be nervous when beginning a questionnaire and more likely to answer sensitive questions if asked later in the questionnaire and; d) produce patterned responses (Krosnick & Presser, 2010; Rossi, Wright, & Anderson, 1983). Research has suggested that item order can influence item-total correlations (Knowles, 1988), item-trait correlations (Steinberg, 1994), and consistency (Knowles & Byers, 1996); moreover, there is evidence that items later in questionnaires better predict traits than earlier items (Hamilton & Shuminsky, 1990; Knowles, 1988). Items can prime respondents (i.e., influence responses through semantic context or implicit anchoring) and have practical impacts on responses to subsequent items (Epley & Gilovich, 2010; Hamilton & Shuminsky, 1990; Steinberg, 1994; Strack & Mussweiler, 1997). Indeed, some research has suggested that minor changes in context can affect item responses and inter-item correlations, with the probability of such effects increasing when similar items are grouped (Schuman & Presser, 1996; Tourangeau, Rips, & Rasinski, 2000).

Item order effects have been of interest in the areas of academic and achievement testing (Neely, Springston, & McCann, 1994; Wainer & Kiely, 1987) and personality assessment (Hamilton & Shuminsky, 1990; Knowles, 1988; Steinberg, 1994; Tourangeau & Rasinski, 1988) for several decades; nevertheless, the role of item order has not been a focus in other clinical areas. There are at least two notable exceptions: first, the *Beck Depression Inventory* (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961), which loses validity when its items are presented randomly (Dahlstrom, Brooks, & Peterson, 1990); second, the Trait scale of the State-Trait Anxiety Inventory (Spielberger, Gorsuch, Luschene, Vagg, & Jacobs, 1983), which has demonstrated item order differences in endorsement and response patterns. As such, item order may have a practical impact on responses and response patterns, which typically result in clinical implications (e.g., supporting diagnoses, monitoring treatment progress); however, this issue has received scant empirical attention in anxiety-related literature.

A reasonable starting point to explore this area would be to simultaneously assess the role of item order on several anxiety-related measures that: (a) assess a wide-range of anxiety-related symptoms, (b) have demonstrated reliability and validity, and (c) have been demonstrated as distinct. Anxiety sensitivity, fear of negative evaluation, illness/injury sensitivity, and intolerance of uncertainty appear to meet the aforementioned criteria. Researchers have argued the independent importance of these constructs for a variety of anxiety and related disorders (Boelen & Carleton, in press; Boelen & Reijntjes, 2009; Carleton, Abrams, Kachur, & Asmundson, 2009; Carleton & Asmundson, 2009; Carleton, Collimore, & Asmundson, 2010; Carleton, Sharpe, & Asmundson, 2007; Taylor, 1993, 1999; Taylor, Asmundson, Carleton, & Brundin, 2007); moreover, a literature search of ScienceDirect and Google Scholar articles from 2006-2011 produced hundreds of unique references citing measures of these constructs, evidencing their popularity and indicating that assessing associated item order effects may be particularly beneficial.

The current study was designed to evaluate whether item order (grouped vs. interspersed) might impact the assessment of several anxiety-related constructs. To achieve this goal, participants were asked to complete measures in one of two presentation modalities: (a) items presented grouped as measures (i.e., standard), or (b) items from all of the measures presented randomly interspersed (i.e., random). Endorsement rates and response patterns of participants from both presentation modalities were then compared to determine if the order of items had any influence.

## Method

### *Participants*

Participants included 428 undergraduates (*n* = 103 men, ages 18-34; *M* = 20.6; *SD* = 3.0; *n* = 325 women, ages 18-45; *M* = 20.5; *SD* = 3.9) and 571 community volunteers (*n* = 187 men, ages 18-55; *M* = 27.9; *SD* = 10.4; *n* = 384 women, ages 18-55; *M* = 28.7; *SD* = 10.8) who completed measures of the constructs of interest as part of a larger study. Undergraduates were solicited through the university research pool, whereas community participants were solicited with web-based advertising to participate in research exploring fear. Web-based data collection has been demonstrated to be a valid approach for questionnaire-based research that is comparable to other data collection methods (Gosling, Vazire, Srivastava, & John, 2004). Many undergraduate participants reported being employed or working at home (5% full-time, 51% part-time, and <1% as homemakers). Most of the undergraduate participants identified their ethnicity as Caucasian (87%), First Nations (i.e., Canadian aboriginals; 2%), or Asian (5%). Most reported being single (82%) or married (12%), with the remainder reporting being divorced (1%) or selecting "Rather Not Say" (5%). The majority of community participants (67%) reported having at least some postsecondary education, being employed or working at home (35% full-time, 21% part-time, and 6% as homemakers) or unemployed but seeking work (12%). Most community participants identified their ethnicity as Caucasian (84%), First Nations (3%), or Asian (5%). Approximately half (55%) reported being single, another third (34%) reported being married, with the remainder reporting being divorced (8%) or selecting "Rather Not Say" (3%).

### *Procedure*

Participants were randomly assigned to conditions based on whether the time at the start of the session ended in an odd or even number. The resulting randomization facilitated data collection such that approximately half of all participants completed the items from the measures described below in the standard fashion (i.e., as cohesive measures; *n* = 485, 48.5%), while the other half completed the items presented in a random order (*n* = 514, 51.5%). In the random presentation modality, item order was randomized for each participant using the computerized testing program. The randomization extended to the item order within each instrument. All of the measures use the same Likert response scales, which facilitated item transitions during responses in the random presentation modality.

### Measures

***Anxiety Sensitivity Index-3* (ASI-3; Taylor, Zvolensky, et al., 2007)**. The ASI-3 is an 18-item self-report measure assessing the tendency to fear anxiety-related sensations based on the belief that they may have harmful or even catastrophic consequences (e.g., "It scares me when my heart beats rapidly"; Reiss & McNally, 1985; Taylor, 1999). Anxiety sensitivity has been extensively and independently related to fearful responding and several psychopathologies (Fedoroff, Taylor, Asmundson, & Koch, 2000; Reiss, 1991; Sexton, Norton, Walker, & Norton, 2003; Taylor, 2004, 1999; Taylor, Koch, McNally, & Crockett, 1992). ASI-3 items are rated on a 5-point Likert scale ranging from 0 (*very little*) to 4 (*very much*). Rather than a simple unitary construct, anxiety sensitivity subsumes fears of physical, mental, and social consequences of anxiety-related sensations (Zinbarg, Barlow, & Brown, 1997). The three factors are: 1) fear of somatic sensations (i.e., somatic; e.g., "When my stomach is upset, I worry that I might be seriously ill"), 2) fear of cognitive dyscontrol (i.e., cognitive; e.g., "It scares me when I am unable to keep my mind on a task"), and 3) fear of socially observable anxiety reactions (i.e., social; e.g., "When I begin to sweat in a social situation, I fear people will think negatively of me"). Factor analyses have supported a robust 3-factor structure for the ASI-3 corresponding to the three originally theorized dimensions of anxiety sensitivity (Taylor, Koch, Woody, & McLean, 1996; Zinbarg et al., 1997), with each subscale and corresponding factor consisting of six items. The ASI-3 has demonstrated evidence for good convergent, discriminant, and criterion-related validity (Taylor, Zvolensky, et al., 2007).

***Brief Fear of Negative Evaluation scale, Straightforward Items* (BFNE-S; Carleton, Collimore, McCabe, & Antony, 2011)**. The BFNE-S is an 8-item revised version of the *Brief Fear of Negative Evaluation scale* (BFNE; Leary, 1983) used for measuring fears of negative evaluation (e.g., "I am afraid that others will not approve of me"). Fear of negative evaluation is "apprehension about others' evaluations, distress over their negative evaluations, avoidance of evaluative situations, and the expectation that others would evaluate oneself negatively" (p. 449; Watson & Friend, 1969). Fear of negative evaluation represents a defining characteristic of Social Anxiety Disorder (American Psychiatric Association, 2000; Carleton, Collimore, et al., 2010; Turk, Heimberg, & Hope, 2001; Turk, Lerner, Heimberg, & Rapee, 2001). BFNE-S items are rated on a 5-point Likert scale ranging from 0 (*not at all characteristic of me*) to 4 (*extremely characteristic of me*). Revisions to the BFNE were made in accordance with previously suggested changes to remove a methodological issue stemming from four reverse-worded items by including only straightforwardly worded items (Carleton, Collimore, & Asmundson, 2007; Carleton et al., 2011; Carleton, McCreary, Norton, & Asmundson, 2006; Weeks et al., 2005). The BFNE-S has been shown to have excellent internal consistency ($\alpha > .90$), to correlate highly with the original scale (*r*s > .95), to correlate well with convergent scales (*r*s > .60), and factor analyses have supported a unitary solution items (Carleton, Collimore, et al., 2007; Carleton et al., 2011; Carleton, McCreary, et al., 2006).

***Illness/Injury Sensitivity Index-Revised* (ISI-R; Carleton, Asmundson, & Taylor, 2005; Carleton, Park, & Asmundson, 2006).** The ISI-R is a 9-item revision of the original Illness/Injury Sensitivity Index (Taylor, 1993), designed to measure fears of illness and injury (e.g., "I am frightened of being injured"). Illness/injury sensitivity represents fears of becoming ill or injured (Reiss, 1991) and has been related to the development of psychopathology (Asmundson, 1999; Carleton, Abrams, Asmundson, Antony, & McCabe, 2009; Sexton et al., 2003). Illness/injury sensitivity has been related to the development and maintenance of chronic pain and health anxiety (Vancleef, Peters, Roelofs, & Asmundson, 2006; Watt, O'Connor, Stewart, Moon, & Terry, 2008), as well as phobias related to physical harm, such as fear of needles, spiders, or heights (Olatunji, Williams, Sawchuk, & Lohr, 2006; Taylor, 1993); however, illness/injury sensitivity appears independent from fears of pain and movement (Kori, Miller, & Todd, 1990; Roelofs, Peters, Fassaert, & Vlaeyen, 2005; Vlaeyen, Kole-Snijders, Boeren, & van Eek, 1995). ISI-R items are rated on a 5-point Likert scale ranging from 0 (*agree very little*) to 4 (*agree very much*). The two factors, fear of illness and fear of injury, are considered to be distinct and loading onto a higher order fear of physical harm as measured by the total ISI-R score (Carleton et al., 2005). Internal consistency (α > .85) and convergent validity (*r*s > .65) for both factors have been shown to be sufficient (Carleton, Park, et al., 2006).

***Intolerance of Uncertainty Scale, Short Form* (IUS-12; Carleton, Norton, & Asmundson, 2007).** The IUS-12 is a 12-item short-form of the original 27-item Intolerance of Uncertainty Scale (Freeston, Rhéaume, Letarte, Dugas, & Ladouceur, 1994) that measures reactions to uncertainty, ambiguous situations, and the future (e.g., "*Unforeseen events upset me greatly*"). Intolerance of uncertainty is a dispositional characteristic resulting from negative beliefs about uncertainty and its implications (Dugas & Robichaud, 2007). Items are scored on a 5-point Likert scale ranging from 1 (*not at all characteristic of me*) to 5 (*entirely characteristic of me*). Evidence suggests IU has a continuous latent structure (Carleton et al., in press). The IUS-12 has a strong correlation with the original scale, *r*s = .94 to .96 (Carleton, Norton, et al., 2007; Khawaja & Yu, 2010), and has been shown to have two factors, prospective IU and inhibitory IU (McEvoy & Mahoney, 2011), with identically high internal consistencies, α = .85 (Carleton, Norton, et al., 2007). The IUS-12 has been shown to have excellent internal consistency and convergent validity with the original (Carleton, Norton, et al., 2007; Carleton, Sharpe, et al., 2007). The psychometric properties of the IUS-12 have all been replicated and reified in clinical and nonclinical samples (Carleton, Sharpe, et al., 2007; Khawaja & Yu, 2010; McEvoy & Mahoney, 2011). The IUS-12 is particularly useful for research because it is psychometrically comparable to the longer original (Khawaja & Yu, 2010) and the symptom-focused Intolerance of Uncertainty Index (IUI; Carleton, Gosselin, & Asmundson, 2010; Gosselin et al., 2008).

### *Analyses*

Descriptive statistics were performed on the items, subscales, and total scores for each measure within each of the samples (i.e., undergraduates and community members). A series of demographic comparisons were performed within and across the two presentation modalities to assess whether the samples could be reasonably collapsed for the subsequent analyses.

Endorsement rates, inter-item correlations, and item-total correlations for both presentation modalities were compared to determine if item-order had an influence on response patterns. Multiple-group confirmatory factor analyses procedures as described by Byrne (2001, 2004) were calculated with AMOS and the results were used to assess whether measurements weights (i.e., factor loadings) differed between each of the two presentation modalities (i.e., random and standard). This multiple-group CFA procedure requires first that the factor structures are tested for each measure; as such, the factor structure as described in the literature for each measure was tested for each presentation modality (i.e., standard vs. random; Table 2). Thereafter, the multiple-group CFA procedure for testing invariance was performed for each measure across each modality.

The raw data from each sample were used as input, along with maximum likelihood estimation. The following fit indices and 90 percent confidence intervals – where applicable – were considered representative of excellent fit and values approaching these cut off scores as indicating an increasingly good fit (Hu & Bentler, 1999; Tabachnick & Fidell, 2007): (1) chi-square (values should not be significant); 2) chi-square/df ratio (values should be less than 2.0); 3) Comparative Fit Index (CFI; values must be greater than .90, and ideal fits approach or are greater than .95); 4) the Standardized Root Mean Square Residual (SRMR; values must be less than .10 and ideal fits approach or are less than .05); 5) Root Mean Square Error of Approximation (RMSEA; values must be less than .08 and ideal fits approach or are less than .05, with 90% confidence interval values below .10); and (6)

Expected Cross-Validation Index (ECVI; when comparing these scores across different models, lower values indicate a closer fit (Browne & Cudeck, 1989, 1993). Goodness of fit evaluations should emphasize the latter four fit indices because of potential chi-square inflation (Hu & Bentler, 1999). Statistically significant differences (i.e., significant values with Cramer's *V* effect sizes greater than .10; Cohen, 1988) in measurement weights would suggest differences in the pattern of responses. This procedure serves as a relatively stringent test of invariance across the presentation modes.

Differential item functioning was estimated to assess whether there were differences in item response patterns across the two presentation modalities. Specifically, differential item functioning occurs when two groups with the same latent traits (i.e., the anxiety constructs of interest) respond to items differently due to test characteristics (Embretson & Reise, 2000; Zumbo, 2007). The Mantel-Haenszel chi-square test was performed to test for differential item functioning and was calculated for each subscale, with the standard presentation modality being the reference group and the random presentation modality being the focal group. Broadly speaking, the Mantel-Haenszel test involves a three-way contingency table that simultaneously considers whether an individual endorses a response, the group membership of the individual, and the total score on the latent variable (Hidalgo & Lopez-Pina, 2004; Holland & Thayer, 1988). As per current protocols, values of statistical significance for the Mantel-Haenszel test were reported in conjunction with effect sizes to prevent the flagging of unimportant differences (Monahan, McHorney, Stump, & Perkins, 2007). Effect sizes were grouped based on standardized mean differences and standardized *p*-differences for all data (i.e., five item Likert scale data): a) negligible differential item functioning was coded as "AA" when the Mantel-Haenszel test was non-significant or if absolute value of the effect size was less than or equal to 0.20; b) marginal differential item functioning was coded as "BB" when the Mantel-Haenszel test was statistically significant and the absolute value of the effect size was greater than 0.20 and less than or equal to 0.40; c) definite differential item functioning was coded as "CC" when the Mantel-Haenszel test was statistically significant and the absolute value of the effect size was greater than 0.40 (Dorans & Kulick, 1986; Monahan et al., 2007; Zwick & Thayer, 1996). Codes with a "-" sign (e.g., BB-) indicate that individuals in the focal group (i.e., random presentation modality) were less likely to endorse higher items on the Likert scale.

## Results

### *Descriptive Statistics*

The descriptive statistics for each item in each dependent variable are available from the authors upon request. None of the indices of univariate skewness and kurtosis were sufficiently out of range to preclude the planned analyses (i.e., had positive standardized skewness values that exceeded 2 or positive standardized kurtosis values that exceeded 7; see Curran, West, & Finch, 1996; Tabachnick & Fidell, 2007). The reliabilities and inter-item correlations for each measure from each of the samples are presented in Table 1. Multivariate normality was assessed using Mardia's coefficient of multivariate kurtosis (Byrne, 2001, 2004) for all models and the results suggested nonnormal data; however, parameter estimates and most CFA model fit indices are robust to nonnormality given maximum-likelihood estimation and a sample size of 100 or more participants (Lei & Lomax, 2005). Nonetheless, the Bollen-Stine bootstrap chi-square was used and bootstrapped parameter estimates were compared with estimates from a maximum-likelihood procedure (Byrne, 2001, 2004; Nevitt & Hancock, 2001). In all cases, the statistical significance value for the Bollen-Stine bootstrap chi-square produced results comparable with those from the maximum-likelihood procedure for the CFA.

### *Comparative Statistics*

The undergraduate and community samples included comparable proportions of women in each presentation modality. Differences in ratio of men to women and in mean age between the samples are reported in Table 1. There were no differences between the standard and random viewing modalities within the undergraduate and community samples with regards to the ratio of men to women or in mean age. In contrast, there were slightly more

Table 1. Sample Compositions

| | UG sample | | Dif between ST and RM | Com sample | | Dif between ST and RM | ST and RM combined | | Dif between UG and Com |
|---|---|---|---|---|---|---|---|---|---|
| | ST ($n$=206) | RM ($n$=222) | | ST ($n$=279) | RM ($n$=292) | | UG ($n$=428) | Com ($n$=571) | |
| Percentage of Women | 78% | 74% | $V$ <.01 | 69% | 66% | $V$ <.01 | 76% | 67% | $V$=.09* |
| Mean Age | 20.72 | 20.70 | $r^2$<.01 | 28.94 | 27.96 | $r^2$<.01 | 24.40 | 28.44 | $r^2$=.18* |

Notes: *$p$<.01; V - Cramer's V; UG - Undergraduates; Com - Community; ST – Standard presentation modality; RM – Random presentation modality

women in the undergraduate sample than the community sample; however, the effect size was small and such disparity is common in undergraduate samples. As expected, participants in the community sample were significantly older than participants in the undergraduate sample.

Total and subscale scores were compared between the undergraduate and community samples within each of the presentation modalities (Table 2). In both presentation modalities, participants from the community sample endorsed higher levels of each construct relative to participants from the undergraduate sample; however, following a Bonferroni correction, few were statistically significant and in all cases the effect sizes were very small (Cohen, 1988). As such, the samples were combined for subsequent analyses.

Total and subscale scores were compared between the presentation modalities (i.e., standard vs. random; Table 2). Following a Bonferroni correction, few differences were statistically significant and in all cases the effect sizes were very small (Cohen, 1988). Specifically, there were mean differences of more than one point on the BFNE-S as well as on the ASI-3 social subscale and, therefore, on the ASI-3 total score. In all cases, participant item endorsements were slightly higher in the standard presentation modality relative to the random presentation modality. The largest difference in scores was on the ASI-3 total score, wherein participants in the standard presentation modality scored 2.15 points higher (out of a possible maximum score of 72). Smaller differences were found for the ASI-3 social subscale (i.e., 1.45 points higher out of a possible maximum score of 24; ~6%) and for the BFNE-S (i.e., 1.50 points higher out of a possible maximum score of 32; ~5%); however, only two of the 11 differences were statistically significant and the effect sizes were very small. As such, the differences can be argued to be minor and relatively unimportant.

Cronbach's alpha was calculated for each subscale and total score (Table 3). Using the procedure recommended by Feldt, Woodruff, and Salih (1987), there were no statistically significant differences between the Cronbach's alpha values across the two modalities (all $F$s<1.65) following a Bonferroni correction (Feldt, Woodruff, & Salih, 1987). As such, the modality differences did not appear to impact internal reliability.

### *Invariance Analyses Results*

The CFA fit indices for each measure generally supported the prescribed factor structures (Table 4); however, in the standard presentation modality, the RMSEA values for the BFNE-S, ISI-R, and IUS-12 were all slightly beyond the recommended range. In all cases the fit indices indicated a statistically significantly (all $p$s<.05) better fit for the random presentation modality relative to the standard presentation modality. This counterintuitive finding (i.e., better factor structures would have been expected from grouping scale items together) serves to underscore that there appear to be no important differences between the two modalities. There were also statistically significant differences between

Table 2. Sample and Presentation Comparisons

| | ST | | Within-ST dif | RM | | Within-RM dif | UG and Com combined | | Dif between ST and RM |
|---|---|---|---|---|---|---|---|---|---|
| | UG (*n*=206) | Com (*n*=279) | $r^2$ | UG (*n*=222) | Com (*n*=292) | $r^2$ | ST (*n*=485) | RM (*n*=514) | $r^2$ |
| ASI-3 – Somatic | 4.46 (4.32) | 5.50 (5.15) | .01 | 3.74 (4.49) | 4.96 (5.36) | .02* | 5.06 (4.83) | 4.43 (5.03) | <.01 |
| ASI-3 – Cognitive | 3.18 (4.26) | 4.63 (5.37) | .02* | 2.84 (3.54) | 4.77 (5.70) | .04* | 4.02 (4.98) | 3.94 (4.97) | <.01 |
| ASI-3 – Social | 8.61 (4.75) | 9.14 (6.13) | <.01 | 6.86 (4.59) | 7.92 (5.86) | .01 | 8.91 (5.59) | 7.46 (5.37) | .02* |
| ASI-3 – Total | 16.26 (10.70) | 19.27 (14.09) | .02* | 13.45 (10.42) | 17.65 (14.39) | .03* | 17.99 (12.84) | 15.84 (12.99) | <.01* |
| BFNE-S | 23.78 (13.52) | 23.56 (14.94) | <.01 | 20.50 (11.86) | 21.56 (14.19) | <.01 | 23.65 (14.34) | 21.13 (13.23) | <.01 |
| ISI-R – Injury | 4.58 (4.55) | 4.40 (4.41) | <.01 | 3.95 (3.93) | 3.96 (4.28) | <.01 | 4.47 (4.47) | 3.96 (4.13) | <.01 |
| ISI-R – Illness | 6.30 (5.44) | 6.36 (5.66) | <.01 | 6.31 (4.69) | 6.10 (5.12) | <.01 | 6.33 (5.56) | 6.20 (4.93) | <.01 |
| ISI-R Total | 10.87 (9.38) | 10.76 (9.45) | <.01 | 10.27 (7.88) | 10.07 (8.75) | <.01 | 10.81 (9.41) | 10.16 (8.38) | <.01 |
| IUS-12 – Prospective | 17.52 (5.99) | 18.89 (6.99) | .01 | 17.51 (5.40) | 18.21 (5.98) | <.01 | 18.31 (6.61) | 17.90 (5.74) | <.01 |
| IUS-12 – Inhibitory | 10.17 (4.58) | 10.97 (5.34) | .01 | 9.84 (4.04) | 11.01 (5.13) | .02 | 10.63 (5.04) | 10.56 (4.72) | <.01 |
| IUS-12 – Total | 27.69 (9.87) | 29.86 (11.60) | .01 | 27.35 (8.71) | 29.22 (10.32) | .01 | 28.94 (10.94) | 28.41 (9.69) | <.01 |

Notes: *Bonferroni-corrected $p<.05$; SDs in parentheses; ST – Standard presentation modality; RM – Random presentation modality; UG – Undergraduates; Com – Community; ASI-3 – Anxiety Sensitivity Index-3; BFNE-S – Brief Fear of Negative Evaluation Index-Straightforward Items; ISI-R – Illness/Injury Sensitivity Index-Revised; IUS-12 – Intolerance of Uncertainty Scale, Short Form

presentation modalities based on measurement weights, suggesting the response patterns were different. Despite these statistically significant differences, the comparative Cramer's *V* effect sizes were all small. Given the substantive power available, the effect sizes are critical to the interpretation of the results, which do not indicate the existence of important differences associated with item order.

### *Differential Item Functioning*

The results of the Mantel-Haenszel chi-square test and associated effect sizes are reported in Table 5. A total of 18 out of 47 Mantel-Haenszel chi-square tests reached statistical significance; however, only three items demonstrated differential item functioning beyond the prescribed negligible range. Items 1 and 9 of the ASI-3, both from the social subscale, and item 3 of the IUS-12, from the inhibitory subscale, were the items that displayed differential item functioning in a range that warrants attention (Dorans & Kulick, 1986; Monahan et al., 2007; Zwick & Thayer, 1996). For item 1 of the ASI-3 and item 3 of the IUS-12, the individuals in the random presentation modality were less likely to endorse higher scores on the Likert scales, while the effect was opposite for item 9 of the ASI-3. The effect sizes were marginal and no items scored in a range suggesting definitive differential item functioning.

Table 3. Reliabilities

| | Undergraduate Sample | | | | Community Sample | | | | Combined Samples | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ST | | RM | | ST | | RM | | ST | | RM | |
| | α | AIC | α | AIC | α | AIC | α | AIC | α | AIC | α | AIC |
| ASI-3 – Somatic | .87 | .43 | .86 | .50 | .85 | .49 | .87 | .52 | .84 | .47 | .86 | .51 |
| ASI-3 – Cognitive | .88 | .55 | .83 | .45 | .90 | .60 | .91 | .61 | .89 | .58 | .89 | .58 |
| ASI-3 – Social | .78 | .37 | .78 | .37 | .85 | .48 | .83 | .44 | .83 | .44 | .81 | .42 |
| ASI-3 – Total | .89 | .31 | .90 | .33 | .92 | .40 | .92 | .40 | .91 | .37 | .92 | .38 |
| BFNE-S | .97 | .78 | .95 | .69 | .96 | .77 | .95 | .72 | .96 | .76 | .95 | .69 |
| ISI-R – Injury | .92 | .74 | .90 | .68 | .92 | .75 | .92 | .73 | .92 | .71 | .87 | .57 |
| ISI-R – Illness | .93 | .63 | .85 | .53 | .92 | .71 | .88 | .60 | .92 | .75 | .91 | .71 |
| ISI-R Total | .94 | .65 | .91 | .52 | .94 | .65 | .93 | .59 | .94 | .65 | .92 | .56 |
| IUS-12 – Prospective | .88 | .51 | .82 | .39 | .89 | .55 | .82 | .39 | .89 | .53 | .82 | .39 |
| IUS-12 – Inhibitory | .88 | .60 | .85 | .52 | .91 | .66 | .87 | .56 | .90 | .64 | .86 | .55 |
| IUS-12 – Total | .92 | .50 | .89 | .41 | .94 | .54 | .90 | .42 | .93 | .53 | .90 | .42 |

Notes: ST – Standard presentation modality; RM – Random presentation modality; AIC – Average Inter-item Correlation; ASI-3 – Anxiety Sensitivity Index-3; BFNE-S – Brief Fear of Negative Evaluation Index-Straightforward Items; ISI-R – Illness/Injury Sensitivity Index-Revised; IUS-12 – Intolerance of Uncertainty Scale, Short Form.

Table 4. Confirmatory Factor Analyses Fit Indices

| | | $\chi^2$ | *df* | $\chi^2/df$ | CFI | SRMR | RMSEA | CI 90% | ECVI | CI 90% | Measurement Weights† |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ASI-3 | ST | 404.31 | 132 | 3.06 | .934 | .048 | .065 | .058; .073 | .997 | .879; 1.129 | |
| | RM | 336.97 | 132 | 2.55 | .954 | .050 | .055 | .048; .062 | .809 | .711; .922 | |
| | CP | | | | | | | | | | $\chi^2(15)=42.07^{**}$, $V = .05$ |
| BFNE-S | ST | 143.56 | 20 | 7.18 | .970 | .021 | .113 | .096; .131 | .363 | .292; .449 | |
| | RM | 72.00 | 20 | 3.60 | .985 | .018 | .071 | .054; .089 | .203 | .159; .261 | |
| | CP | | | | | | | | | | $\chi^2(7)=19.03^{**}$, $V = .05$ |
| ISI-R | ST | 226.36 | 26 | 8.71 | .948 | .041 | .126 | .111; .142 | .546 | .455; .653 | |
| | RM | 55.92 | 26 | 2.15 | .990 | .019 | .047 | .030; .064 | .183 | .148; .233 | |
| | CP | | | | | | | | | | $\chi^2(7)=10.66$, $V = .04$ |
| IUS-12 | ST | 301.41 | 53 | 5.69 | .930 | .053 | .098 | .088; .109 | .726 | .621; .847 | |
| | RM | 169.24 | 53 | 3.19 | .954 | .043 | .065 | .054; .077 | .427 | .358; .512 | |
| | CP | | | | | | | | | | $\chi^2(10)=46.86^{**}$, $V = .07$ |

Notes: **p*<.05; ST – Standard presentation modality; RM – Random presentation modality; CP – Comparing the ST and RM; † Invariance is indicated by a lack of statistical significance; CFI – Comparative Fit Index (must be greater than .90); SRMR – Standardized Root Mean Square Residual (must be less than .10); RMSEA – Root Mean Square Error of Approximation (must be less than .08); ECVI – Expected Cross-Validation Index (lower values indicate closer fits when comparing models); ASI-3 – Anxiety Sensitivity Index-3; BFNE-S – Brief Fear of Negative Evaluation Index-Straightforward Items; ISI-R – Illness/Injury Sensitivity Index-Revised; IUS-12 – Intolerance of Uncertainty Scale, Short Form.

Table 5. Differential item functioning (DIF) for each item of each scale

| | Chi-square | Effect size | | | Chi-square | Effect size | |
|---|---|---|---|---|---|---|---|
| ASI-3 Social | | | | ASI-3 Somatic | | | |
| ASI-3, 1 | 19.06* | -.27 | BB- | ASI-3, 3 | 12.90* | -.15 | AA |
| ASI-3, 6 | 1.02 | -.04 | AA | ASI-3, 4 | 3.24 | .08 | AA |
| ASI-3, 9 | 19.61* | .23 | BB | ASI-3, 7 | 5.87* | -.10 | AA |
| ASI-3, 11 | .08 | .02 | AA | ASI-3, 8 | .12 | -.01 | AA |
| ASI-3, 13 | .00 | .01 | AA | ASI-3, 12 | 10.93* | .12 | AA |
| ASI-3, 17 | .60 | .05 | AA | ASI-3, 15 | 2.68 | .06 | AA |
| ASI-3 Cognitive | | | | BFNE-S | | | |
| ASI-3, 2 | 21.81* | -.16 | AA | BFNE-S, 1 | 2.45 | -.08 | AA |
| ASI-3, 5 | 4.20* | .07 | AA | BFNE-S, 2 | 4.93* | -.11 | AA |
| ASI-3, 10 | .21 | -.01 | AA | BFNE-S, 3 | 1.61 | -.05 | AA |
| ASI-3, 14 | 8.98* | .13 | AA | BFNE-S, 4 | .00 | -.01 | AA |
| ASI-3, 16 | .19 | .01 | AA | BFNE-S, 5 | 3.01 | .08 | AA |
| ASI-3, 18 | .43 | -.04 | AA | BFNE-S, 6 | 10.04* | .12 | AA |
| | | | | BFNE-S, 7 | .01 | .01 | AA |
| | | | | BFNE-S, 8 | .27 | .03 | AA |
| ISI-R Injury | | | | ISI-R Injury Illness | | | |
| ISI-R, 1 | 20.35* | -.15 | AA | ISI-R, 3 | 5.57* | -.10 | AA |
| ISI-R, 2 | .00 | .00 | AA | ISI-R, 4 | .01 | .00 | AA |
| ISI-R, 5 | .11 | -.01 | AA | ISI-R, 6 | .04 | .01 | AA |
| ISI-R, 9 | 16.71* | .15 | AA | ISI-R, 7 | 3.76 | .10 | AA |
| | | | | ISI-R, 8 | .14 | -.01 | AA |
| IUS-12 Prospective | | | | IUS-12 Inhibitory | | | |
| IUS-12, 1 | 6.67* | -.11 | AA | IUS-12, 3 | 14.53* | -.21 | BB- |
| IUS-12, 2 | .59 | -.04 | AA | IUS-12, 6 | 1.99 | .04 | AA |
| IUS-12, 4 | 4.95* | -.12 | AA | IUS-12, 7 | .51 | .03 | AA |
| IUS-12, 5 | .21 | .02 | AA | IUS-12, 10 | 1.70 | .08 | AA |
| IUS-12, 8 | 6.34* | .15 | AA | IUS-12, 12 | .86 | .05 | AA |
| IUS-12, 9 | .36 | -.04 | AA | | | | |
| IUS-12, 11 | 5.95* | .14 | AA | | | | |

Notes: *$p$<.05; ASI-3 – Anxiety Sensitivity Index-3; BFNE-S – Brief Fear of Negative Evaluation Index- Straightforward Items; ISI-R – Illness/Injury Sensitivity Index-Revised; Intolerance of Uncertainty Scale, Short Form; AA - negligible DIF (i.e., Mantel-Haenszel test was non-significant or absolute value of the effect size was less than or equal to 0.20); BB - marginal DIF (i.e., Mantel-Haenszel test was statistically significant and the absolute value of the effect size was greater than 0.20 and less than or equal to 0.40); CC - definite DIF (i.e., Mantel-Haenszel test was statistically significant and the absolute value of the effect size was greater than 0.40).

## Discussion

The current study was designed to evaluate how item order may impact the assessment of several anxiety-related constructs. To that end, participants completed several anxiety-related measures presented either as grouped measures (i.e., standard) or with all items randomly interspersed (i.e., random). The design allowed for comparisons of endorsement rates and response patterns to determine the degree of impact from item order. The study serves, in part, to extend previous research exploring the impact of item presentation on response patterns, such as with the *Beck*

*Depression Inventory* (Beck et al., 1961) and the Trait scale of the State-Trait Anxiety Inventory (Spielberger et al., 1983). Evaluating item order effects on more specific anxiety-related constructs (e.g., as represented by the ASI-3, BFNE-S, ISI-R, and IUS-12), rather than general anxiety (e.g., as represented by the State-Trait Anxiety Inventory), should facilitate confidence for disorder-specific item-order presentations (i.e., there appear to be no psychometric reasons to worry about grouping items by disorder).

Mean responses, skew, and kurtosis for each individual item and for each of the subscales and total scores were very similar across both presentation modalities. There was a tendency for slightly higher scores on items and subscales presented in the standard modality (i.e., grouped as measures). Despite the tendency, the effect sizes indicated these differences would be unimportant from a practical clinical perspective.

Cronbach's alphas were comparable across the two modalities. In most cases the reliability for items presented in the random modality was very slightly lower, but none was statistically significantly lower. As such, it appears that construct cohesion as measured by Cronbach's alpha is not substantially impacted by the order of individual items.

The CFA results were unexpected in that the fit indices were consistently superior for the random presentation modality relative to the standard presentation modality. In the random presentation modality, all of the prescribed factor solutions met all fit index requirements, suggesting the fit was excellent in all cases. Similarly, most of the prescribed factor solutions met all fit index requirements with data from the standard modality. The exceptions were for the RMSEA indices, for which only the ASI-3 solution produced values under the recommended cut-off for the standard presentation modality. Using the stringent tests of invariance (Byrne, 2001, 2004) across the presentation modalities identified differences based on measurement weights; however, in all instances the comparative effect sizes were all small. As such, it appears that while item order does produce minor loading differences in response patterns, the differences themselves may not be sufficient to warrant practical considerations. The fact that the directions of the relationships were counterintuitive (i.e., better factor structures would have been expected from grouping scale items together) serves to highlight the conclusion that grouping items as measures does not inflate their similarities and influence responses due to context.

Differential item functioning was minimal between presentation modalities. Indeed, only three out of 47 items produced non-negligible differential item functioning. Consistent with differences on subscale scores, two items from the ASI-3 social subscale demonstrated differential item functioning; moreover, item order appeared to have influenced response patterns on one item on the IUS-12 inhibitory subscale. Overall, however, it appears that individuals with similar levels of the constructs of interest responded to the individual items similarly. For example, two individuals who have relatively extreme fears of negative evaluation (e.g., two standard deviations above the mean) would generally respond similarly to items on the BFNE-S (i.e., would chose the same choice on the Likert scale) regardless of whether the items were presented coherently as a measure or randomly interspersed with other items.

When questions related to social anxiety were asked in a series, participants endorsed very slightly higher levels than when the same items were asked sporadically. Despite the statistical significance of the difference, the practical difference remained unimportant. The relative increases may be the result of social anxiety specific biases, such as priming. These findings were supported by differential item functioning for two items in the social subscale of the ASI-3; however, these results did not generalize to items of the BFNE-S. ASI-3 item one – an item in the ASI-3 social subscale – would have been the very first item participants saw in the standard modality. ASI-3 item one demonstrated the greatest differential item functioning and previous research (Steinberg, 1994) has suggested that the first item in questionnaires may be particularly prone to differential item functioning. In other words, the differences in the ASI-3 item responding and the subsequent differences on the ASI-3 social subscale may have been exacerbated by the fact that one of its items was presented as the very first item in the battery of questionnaires. Researchers may benefit from further studies examining how the very first item of anxiety-related questionnaire batteries may be responded to differently.

### Caveats and Limitations

The current study has several limitations and findings that provide directions for future studies. First, randomization was the only alternative presentation modality presented. Alternative presentation modalities (e.g., backwards or from most severe symptoms to least severe; Dahlstrom et al., 1990; Kornblith, Greenwald, Michelson, & Kazdin, 1984) may produce different results, though such alternatives are rare in the literature and in practice. Second, because these were a community and a university sample, it is unlikely that these findings can definitively inform response patterns from persons with clinical levels of anxiety-related psychopathology. People suffering from clinical levels of fear may be more profoundly influenced by item order effects; however, this remains to be tested using clinical samples. Third, although the majority of fit indices support the a priori factor structures, the fact that the random modality produced better fit indices runs counter-intuitive and warrants replication and further investigation.

Despite the aforementioned limitations, the current study is the first to examine the role of item order in assessing anxiety-related constructs and adds to earlier evidence regarding item order on participant responses (e.g., Dahl, Wilson, & Nilsson, 2004; DeMoranville, Bienstock, & Judson, 2008). Specifically, the items were presented in a standard format (i.e., visually grouped together as measures) and in a random format (i.e., the items were randomly interspersed for each participant). The random presentation format controlled for potential item order effects and the possibility that the constructs were independent as a result of learned responding and serial effects (Krosnick & Presser, 2010). Item response means, summed response means, reliabilities, and response patterns for individual items were not substantially different across presentation modalities. The largest difference in scores only represented an approximate increase of 2 points (out of a possible maximum score of 72) for the standard presentation modality relative to the random presentation modality. As such, when using prescribed cut-off scores for either the ASI-3 social subscale or the BFNE-S, clinicians should be aware that the score may be slightly inflated relative to the "true" latent score. Based on these results, it may be prudent to consider a very slight margin of error for all social anxiety measures (e.g., the Social Interaction Phobia Scale; Carleton, Collimore, et al., 2009; Weeks, Carleton, Asmundson, McCabe, & Antony, 2010); The Social Avoidance and Distress Scale; Turner, McCanna, & Beidel, 1987); however, from a practical standpoint, clinical conclusions based on subscale or total scores are unlikely to be significantly skewed by item order presentation because all clinical decisions should allow for at least as much error as would be supported by these results. Accordingly, clinicians likely need not worry about the order of item presentation for the ASI-3, BFNE-S, ISI-R, or IUS-12. By way of extension, it would seem that the order of questionnaire presentation for these constructs is also unlikely to have a significant impact on responses; however, this presumption remains to be tested.

There is a paucity of research exploring these types of questions; as such, we encourage researchers to similarly examine other scales and constructs for order effects, particularly where scale scores are used in important decision making (e.g., hiring, selection into academic programs, diagnoses). Although we were gratified to see these scales not subject to substantial item order effects, it is not necessarily the case that other scales or other constructs are immune from these effects.

## References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text revision ed.). Washington, DC: Author.

Asmundson, G. J. G. (1999). Anxiety sensitivity and chronic pain: Empirical findings, clinical implications, and future directions. In S. Taylor (Ed.), *Anxiety sensitivity: Theory, research and treatment of the fear of anxiety* (pp. 269-285). New Jersey: Erlbaum.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4*, 561-571.

Boelen, P. A., & Carleton, R. N. (2012). Intolerance of uncertainty, hypochondriacal concerns, OCD, and worry. *Journal of Nervous and Mental Disease*, *300*, 208-213.

Boelen, P. A., & Reijntjes, A. (2009). Intolerance of uncertainty and social anxiety. *Journal of Anxiety Disorders, 23*, 130-135. doi: 10.1016/j.janxdis.2008.04.007

Browne, M. W., & Cudeck, R. (1989). Single sample cross-validation indices for covariance structures. *Multivariate Behavioral Research, 24*, 445-455. doi: 10.1207/s15327906mbr2404_4

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models*. Newbury Park, CA: Sage Publications Inc.

Byrne, B. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming.* Mahwah, NJ: Erlbaum.

Byrne, B. (2004). Testing for multigroup invariance using amos graphics: A road less traveled. *Structural Equation Modeling, 11*, 272-300. doi: 10.1207/s15328007sem1102_8

Carleton, R. N., Abrams, M. P., Asmundson, G. J. G., Antony, M. M., & McCabe, R. E. (2009). Pain-related anxiety and anxiety sensitivity across anxiety and depressive disorders. *Journal of Anxiety Disorders, 23*, 791-798.

Carleton, R. N., Abrams, M. P., Kachur, S. S., & Asmundson, G. J. G. (2009). Waddell's symptoms as correlates of vulnerabilities associated with fear-anxiety-avoidance models of pain: Pain-related anxiety, catastrophic thinking, perceived disability, and treatment outcome. *Journal of Occupational Rehabilitation, 19*, 364-374.

Carleton, R. N., & Asmundson, G. J. G. (2009). The multidimensionality of fear of pain: Construct independence for the Fear of Pain Questionnaire-short form and the Pain Anxiety Symptoms Scale-20. *The Journal of Pain, 10*, 29-37.

Carleton, R. N., Asmundson, G. J. G., & Taylor, S. (2005). Fear of physical harm: Factor structure and psychometric properties of the injury/illness sensitivity index. *Journal of Psychopathology and Behavioral Assessment, 27*, 235-241.

Carleton, R. N., Collimore, K. C., & Asmundson, G. J. G. (2007). Social anxiety and fear of negative evaluation: Construct validity of the BFNE-II. *Journal of Anxiety Disorders, 21*, 131-141. doi: S0887-6185(06)00047-8 [pii] 10.1016/j.janxdis.2006.03.010

Carleton, R. N., Collimore, K. C., & Asmundson, G. J. G. (2010). "It's not just the judgements – it's that I don't know": Intolerance of uncertainty as a predictor of social anxiety. . *Journal of Anxiety Disorders, 24*, 189-195. doi: 10.1016/j.janxdis.2009.10.007

Carleton, R. N., Collimore, K. C., Asmundson, G. J. G., McCabe, R. E., Rowa, K., & Antony, M. M. (2009). Refining and validating the Social Interaction Anxiety Scale and the Social Phobia Scale. *Depression and Anxiety, 26*, E71-81.

Carleton, R. N., Collimore, K. C., McCabe, R. E., & Antony, M. M. (2011). Addressing revisions to the brief fear of negative evaluation scale: Measuring fear of negative evaluation across anxiety and mood disorders. *Journal of Anxiety Disorders, 25*, 822-828. doi: 10.1016/j.janxdis.2011.04.002

Carleton, R. N., Gosselin, P., & Asmundson, G. J. G. (2010). The Intolerance of Uncertainty Index: Replication and extension with an english sample. *Psychological Assessment, 22*, 396-406. doi: 10.1037/a0019230

Carleton, R. N., McCreary, D. R., Norton, P. J., & Asmundson, G. J. G. (2006). Brief Fear of Negative Evaluation Scale-revised. *Depression and Anxiety, 23*, 297-303. doi: 10.1002/da.20142

Carleton, R. N., Norton, M. A., & Asmundson, G. J. G. (2007). Fearing the unknown: A short version of the Intolerance of Uncertainty Scale. *Journal of Anxiety Disorders, 21*, 105-117. doi: S0887-6185(06)00051-X [pii]10.1016/j.janxdis.2006.03.014

Carleton, R. N., Park, I., & Asmundson, G. J. G. (2006). The Illness/Injury Sensitivity Index: An examination of construct validity. *Depression and Anxiety, 23*, 340-346. doi: 10.1002/da.20138

Carleton, R. N., Sharpe, D., & Asmundson, G. J. G. (2007). Anxiety sensitivity and intolerance of uncertainty: Requisites of the fundamental fears? *Behaviour Research and Therapy, 45*, 2307-2316. doi: S0005-7967(07)00095-2 [pii]10.1016/j.brat.2007.04.006

Carleton, R. N., Weeks, J. W., Howell, A. N., Asmundson, G. J. G., Antony, M. M., & McCabe, R. E. (2012). Assessing the latent structure of the intolerance of uncertainty construct: An initial taxometric analysis. *Journal of Anxiety Disorders, 26,* 150-157.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* (2nd ed.). Mawah, NJ: Erlbaum.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1*, 16-29.

Dahl, J., Wilson, K. G., & Nilsson, A. (2004). Acceptance and commitment therapy and the treatment of persons at risk for long-term disability resulting from stress and painsymptoms: A preliminary randomized trial. *Behavior Therapy, 35*, 785-801.

Dahlstrom, W. G., Brooks, J. D., & Peterson, C. D. (1990). The Beck Depression Inventory: Item order and the impact of response sets. *Journal of Personality Assessment, 55*, 224-233.

DeMoranville, C. W., Bienstock, C. C., & Judson, K. (2008). Using question order for predictive service quality measures. *The Journal of Services Marketing, 22*, 255-262.

Dorans, N. J., & Kulick, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the scholastic aptitude-test. *Journal of Educational Measurement, 23*, 355-368.

Dugas, M. J., & Robichaud, M. (2007). *Cognitive-behavioral treatment for generalized anxiety disorder: From science to practice*. New York, NY: Routledge.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, N.J.: L. Erlbaum Associates.

Epley, N., & Gilovich, T. (2010). Anchoring unbound. *Journal of Consumer Psychology, 20*, 20-24. doi: DOI 10.1016/j.jcps.2009.12.005

Fedoroff, I. C., Taylor, S., Asmundson, G. J. G., & Koch, W. J. (2000). Cognitive factors in traumatic stress reactions: Predicting PTSD symptoms from anxiety sensitivity and beliefs about harmful events. *Behavioural and Cognitive Psychotherapy, 28*, 5-15.

Feldt, L. S., Woodruff, D. J., & Salih, F. A. (1987). Statistical-inference for coefficient-alpha. *Applied Psychological Measurement, 11*, 93-103.

Freeston, M., Rhéaume, J., Letarte, H., Dugas, M. J., & Ladouceur, R. (1994). Why do people worry? *Personality and Individual Differences, 17*, 791-802. doi: 10.1016/0191-8869(94)90048-5

Gosselin, P., Ladouceur, R., Evers, A., Laverdiere, A., Routhier, S., & Tremblay-Picard, M. (2008). Evaluation of intolerance of uncertainty: Development and validation of a new self-report measure. *Journal of Anxiety Disorders, 22*, 1427-1439. doi: 10.1016/j.janxdis.2008.02.005

Hamilton, J. C., & Shuminsky, T. R. (1990). Self-awareness mediates the relationship between serial position and item reliability. *Journal of Personality and Social Psychology, 59*, 1301-1307.

Hidalgo, M. D., & Lopez-Pina, J. A. (2004). Differential item functioning detection and effect size: A comparison between logistic regression and mantel-haenszel procedures. *Educational and Psychological Measurement, 64*, 903-915. doi: Doi 10.1177/0013164403261769

Holland, P. W., & Thayer, D. T. (1988). Differential item functioning and the mantel-haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129-145). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hu, L., & Bentler, P. M. (1999). Fit indices in covariance structure modeling: Sensitivity to underparameterized model mis-specification. *Psychological Methods, 3*, 424-453.

Khawaja, N. G., & Yu, L., N. H. (2010). A comparison of the 27-item and 12-item Intolerance of Uncertainty Scales. *Clinical Psychologist, 14*, 97-106. doi: Doi 10.1080/13284207.2010.502542

Knowles, E. S. (1988). Item context effects on personality-scales - measuring changes the measure. *Journal of Personality and Social Psychology, 55*, 312-320.

Knowles, E. S., & Byers, B. (1996). Reliability shifts in measurement reactivity: Driven by content engagement or self-engagement? *Journal of Personality and Social Psychology, 70*, 1080-1090.

Kori, S. H., Miller, R. P., & Todd, D. D. (1990). Kinesiophobia: A new view of chronic pain behavior. *Pain Management, 3*, 35-43.

Kornblith, S. J., Greenwald, D. P., Michelson, L., & Kazdin, A. E. (1984). Measuring the effects of demand characteristics on the Beck Depression Inventory responses of college students. *Journal of Psychopathology and Behavioral Assessment, 6*, 45-49.

Krosnick, J. A., & Presser, S. (2010). Question and questionnaire design. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of survey research* (pp. 263-314). Bingley, UK: Emerald Group Publishing Limited.

Lam, T. C. M., Green, K. E., & Bordignon, C. (2002). Effects of item grouping and position of the "don't know" option on questionnaire response. *Field Methods, 14*, 418-432. doi: 10.1177/152582202237730

Lei, M., & Lomax, R. G. (2005). The effect of varying degrees of nonormality in structural equation modeling. *Structural Equation Modeling, 12*, 1-27. doi: 10.1207/s15328007sem1201_1

Marsden, P. V., & Wright, J. D. (2010). *Handbook of survey research*. Bingley, UK: Emerald Group Publishing Limited.

McEvoy, P. M., & Mahoney, A. E. J. (2011). Achieving certainty about the structure of intolerance of uncertainty in a treatment-seeking sample with anxiety and depression. *Journal of Anxiety Disorders, 25*, 112-122. doi: 10.1016/j.janxdis.2010.08.010

Monahan, P. O., McHorney, C. A., Stump, T. E., & Perkins, A. J. (2007). Odds ratio, delta, ETS classification, and standardization measures of dif magnitude for binary logistic regression. *Journal of Educational and Behavioral Statistics, 32*, 92-109. doi: Doi 10.3102/1076998606298035

Neely, D. L., Springston, F. J., & McCann, S. J. (1994). Does item order affect performance on multiplechoice exams? *Teaching of Psychology, 21*, 44-45.

Nevitt, J., & Hancock, G. R. (2001). Performance of bootstrapping approaches to model test statistics and parameter standard error estimation in structural equation modeling. *Structural Equation Modeling, 8*, 353-377. doi: 10.1207/S15328007SEM0803_2

Olatunji, B. O., Williams, N. L., Sawchuk, C. N., & Lohr, J. M. (2006). Disgust, anxiety and fainting symptoms associated with blood-injection-injury fears: A structural model. *Journal of Anxiety Disorders, 20*, 23-41. doi: 10.1016/j.janxdis.2004.11.009

Reiss, S. (1991). Expectancy model of fear, anxiety, and panic. *Clinical Psychology Review, 11*, 141-153. doi: 10.1016/0272-7358(91)90092-9

Reiss, S., & McNally, R. J. (1985). The expectancy model of fear. In S. Reiss & R. R. Bootzin (Eds.), *Theoretical issues in behaviour therapy* (pp. 107-121). New York, NY: Academic Press.

Roelofs, J., Peters, M. L., Fassaert, T., & Vlaeyen, J. W. S. (2005). The role of fear of movement and injury in selective attentional processing in patients with chronic low back pain: A dot-probe evaluation. *The Journal of Pain, 6*, 294-300.

Rossi, P. H., Wright, J. D., & Anderson, A. B. (1983). *Handbook of survey research*. New York, NY: Academic Press.

Schuman, H., & Presser, S. (1996). *Questions and answers in attitude survey*. London: Sage Publications Inc.

Schwarz, N. (1995). What respondents learn from questionnaires: The survey interview and the logic of conversation. *International Statistical Review, 63*, 153-168.

Sexton, K. A., Norton, P. J., Walker, J. R., & Norton, G. R. (2003). Hierarchical model of generalized and specific vulnerabilities in anxiety. *Cognitive Behaviour Therapy, 32*, 82-94. doi: 10.1080/16506070302321

Spielberger, C. D., Gorsuch, R. L., Luschene, R. E., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Steinberg, L. (1994). Context and serial-order effects in personality measurement - limits on the generality of measuring changes the measure. *Journal of Personality and Social Psychology, 66*, 341-349.

Strack, F., & Mussweiler, T. (1997). Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of Personality and Social Psychology, 73*, 437-446.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Allyn and Bacon.

Taylor, S. (1993). The structure of fundamental fears. *Journal of Behavior Therapy and Experimental Psychiatry, 24*, 289-299. doi: 10.1016/0005-7916(93)90053-Y

Taylor, S. (2004). Anxiety sensitivity and its implications for understanding and treating ptsd. In S. Taylor (Ed.), *Advances in the treatment of posttraumatic stress disorder: Cognitive-behavioral perspectives* (pp. 57-66). New York, NY: Springer Publishing Company.

Taylor, S. (Ed.). (1999). *Anxiety sensitivity: Theory, research, and treatment of the fear of anxiety*. Mahwah, NJ: Erlbaum.

Taylor, S., Asmundson, G. J. G., Carleton, R. N., & Brundin, P. (2007). Acute posttraumatic stress symptoms and depression after exposure to the 2005 Saskatchewan centennial air show disaster: Prevalence and predictors. *American Journal of Disaster Medicine, 2*, 217-230.

Taylor, S., Koch, W. J., McNally, R. J., & Crockett, D. J. (1992). Conceptualizations of anxiety sensitivity. *Psychological Assessment, 4*, 245-250.

Taylor, S., Koch, W. J., Woody, S., & McLean, P. (1996). Anxiety sensitivity and depression: How are they related? *Journal of Abnormal Psychology, 105*, 474-479.

Taylor, S., Zvolensky, M. J., Cox, B. J., Deacon, B., Heimberg, R. G., Ledley, D. R., . . . Cardenas, S. J. (2007). Robust dimensions of anxiety sensitivity: Development and initial validation of the Anxiety Sensitivity Index-3. *Psychological Assessment, 19*, 176-188. doi: 2007-07953-002 [pii]10.1037/1040-3590.19.2.176

Tourangeau, R., & Rasinski, K. A. (1988). Cognitive-processes underlying context effects in attitude measurement. *Psychological Bulletin, 103*, 299-314.

Tourangeau, R., Rips, L. J., & Rasinski, K. A. (2000). *The psychology of survey response*. New York, NY: Cambridge University Press.

Turk, C. L., Heimberg, R. G., & Hope, D. A. (2001). Social anxiety disorder. In D. H. Barlow (Ed.), *Clinical handbook of psychological disorders: A step-by-step treatment manual (3rd ed.)* (pp. 114-153). New York, NY: Guilford Press.

Turk, C. L., Lerner, J., Heimberg, R. G., & Rapee, R. M. (2001). An integrated cognitive-behavioral model of social anxiety. In S. G. Hofmann & P. M. DiBartolo (Eds.), *From social anxiety to social phobia: Multiple perspectives* (pp. 281-303). Needham Heights, MA: Allyn & Bacon.

Turner, S. M., McCanna, M., & Beidel, D. C. (1987). Validity of the social avoidance and distress and fear of negative evaluation scales. *Behaviour Research and Therapy, 25*, 113-115.

Vancleef, L. M., Peters, M. L., Roelofs, J., & Asmundson, G. J. G. (2006). Do fundamental fears differentially contribute to pain-related fear and pain catastrophizing? An evaluation of the sensitivity index. *European Journal of Pain, 10*, 527-536.

Vlaeyen, J. W. S., Kole-Snijders, A. M., Boeren, R. G., & van Eek, H. (1995). Fear of movement/(re)injury in chronic low back pain and its relation to behavioral performance. *Pain, 62*, 363-372. doi: 030439599400279N [pii]

Wainer, H., & Kiely, G. L. (1987). Item clusters and computerized adaptive testing: A case for testlets. *Journal of Educational Measurement, 24*, 185-201.

Watson, D., & Friend, R. (1969). Measurement of social-evaluative anxiety. *Journal of Consulting and Clinical Psychology, 33*, 448-457.

Watt, M. C., O'Connor, R., Stewart, S., Moon, E., & Terry, L. (2008). Specificity of childhood learning experiences related to anxiety sensitivity and illness/injury sensitivity: Implications for health anxiety and pain. *Journal of Cognitive Psychotherapy, 22*, 128-142.

Weeks, J. W., Carleton, R. N., Asmundson, G. J. G., McCabe, R. E., & Antony, M. M. (2010). "Social anxiety disorder carved at its joints": Evidence for the taxonicity of social anxiety disorder. *Journal of Anxiety Disorders, 24*, 734-742. doi: DOI 10.1016/j.janxdis.2010.05.006

Weeks, J. W., Heimberg, R. G., Fresco, D. M., Hart, T. A., Turk, C. L., Schneier, F. R., & Liebowitz, M. R. (2005). Empirical validation and psychometric evaluation of the Brief Fear of Negative Evaluation scale in patients with social anxiety disorder. *Psychological Assessment, 17*, 179-190.

Zinbarg, R. E., Barlow, D. H., & Brown, T. A. (1997). Hierarchical structure and general factor saturation of the Anxiety Sensitivity Index. *Psychological Assessment, 9*, 277-284.

Zumbo, B. D. (2007). Three generations of differential item functioning (DIF) analyses: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly, 4*, 223-233.

Zwick, R., & Thayer, D. T. (1996). Evaluating the magnitude of differential item functioning in polytomous items. *Journal of Educational and Behavioral Statistics, 21*, 187-201.

**Citation:**

Carleton, R. Nicholas, Thibodeau, Michel A., Osborne, Jason W. & Asmundson, Gordon J. G. (2012). Exploring Item Order in Anxiety-Related Constructs: Practical Impacts of Serial Position. *Practical Assessment, Research & Evaluation*, 17(7). Available online: http://pareonline.net/getvn.asp?v=17&n=7

**Corresponding Author:**

R. Nicholas Carleton
Department of Psychology
University of Regina
Regina, Saskatchewan, S4S 0A2.
Tel: (306) 337-2473; Fax: (306) 337-3275;
E-mail: nick.carleton [at] uregina.ca