Internet-Based Attitude Assessment: Does Gender Affect Measurement Equivalence?

Author: **Brian G. Whitaker**

Abstract

As researchers and practitioners increasingly turn to the Internet as a data collection medium, investigating the differential effects of administration mode on data quality becomes increasingly important. To date, no research has investigated whether data gathered from differing subgroups demonstrate measurement equivalence across Internet and paper-and-pencil administration modes despite the large literature suggesting that males and females differ in terms of computer anxiety. The present study, employing a repeated measures design, investigated systematic cross-mode differences in responding as a function of gender. Results demonstrate that both sexes use the same psychological metric when providing satisfaction ratings regardless of Internet or paper-and-pencil administration (equivalent factor structure and factor loadings). Furthermore, there were no statistically significant differences in scale/item means over the modes of administration according to gender. These results suggest that gender does not lead to a negative impact on cross-mode measurement equivalence for self-report ratings.

Brian G. Whitaker (2007) "Internet-Based Attitude Assessment: Does Gender Affect Measurement Equivalence?" *Computers in Human Behavior* #23, pp. 1183-1194. Version of Record available At www.springer.com [DOI:10.1016/j.chb.2004.11.016]

1. Introduction

More and more researchers and practitioners are turning to Internet-based administration as a method for collecting survey/questionnaire data (Church, 2001; Thompson, Surface, Martin, & Sanders, 2003). This phenomenon is not surprising given the reported advantages of collecting data over the Internet (for re- views, see Barak & English, 2002; Thompson et al., 2003; Yun & Trumbo, 2000).

Fortunately, to date, research has consistently demonstrated that survey respondents view the dimensionality of the constructs similarly and that items measure the constructs in the same way across administration mode. Indeed, measurement equivalence across Internet and paper-and-pencil administrations has been replicated for measures such as self-monitoring (Buchanan & Smith, 1999a, 1999b), personality (Pas- veer & Ellard, 1998), self-efficacy and technology course satisfaction (Burnkrant & Taylor, 2001), organizational justice (Stanton, 1998), 360 feedback (Fenlason, 2000), selection (Ployhart, Weekley, Holtz, & Kemp, 2003), performance appraisal (Facteau
& Craig, 2001; Maurer, Raju, & Collins, 1998), and employee attitudes (Collins, Raju, & Edwards, 2000; Etchegaray, Sexton, Helmreich, & Thomas, 2003; Magnan, Lundby, & Fenlason, 2000; Spera & Moye, 2001; Ryan, Chan, Ployhart, & Slade, 1999). Thus, it seems reasonable to assume that, as a whole, respondents use the same psychological metric when responding to Internet or paper-and-pencil surveys.

To date, relatively few studies have investigated whether differing subgroups of respondents view the dimensionality of the constructs equivalently when providing ratings via identical administration method. Facteau and Craig (2001) demonstrated measurement equivalence of subordinate and peer ratings of managerial team-building skills. Similarly, Robie, Zickar, and Schmidt (2001) demonstrated the measurement equivalence of personality test scores across applicant and incumbent groups. However, measurement equivalence has not been established for all studies examining subgroups. In contrast to the aforementioned findings, Vandenberg and Riordan (1994) demonstrated that different cultural groups provided inequivalent responses to organizational commitment measures. In addition, Maitland, Dixon, Hultsch, and Hertzog (2001) found that male and female responses to the Bradburn Affect Scale (a measure of well-being) were not equivalent. However, no research has investigated whether data gathered from differing subgroups demonstrate measurement equivalence across Internet and paper-and-pencil administration modes. The dearth of research into cross-mode subgroup measurement equivalence stands in stark contrast to the plethora of findings indicating that males and females differ with respect to computer anxiety.

Many studies have demonstrated that females report higher levels of computer anxiety than males (Brosnan, 1998; Liu, Reed, & Phillips, 1992; Rosen & Weil, 1994; Teasdale & Lupart, 2001). Furthermore, these effects have been demonstrated in a variety of environments (Hacklin, 1989) and across age groups (Williams, Ogletree, Woodburn, & Raffeld, 1993).

Computer anxiety has generally been conceptualized as a fear related to the use of computers, or feelings of intimidation and hostility towards this form of technology (Meier, 1985; Glass & Knight, 1988). Research has shown that computer anxiety is

negatively related to performance on computer-based selection tests (Frericks, Ehrhart, & O'Connell, 2003) as well as performance on simple tasks that require the use of a computer (Brosnan, 1998; Mahar, Henderson, & Deane, 1997; Rosen & Weil, 1995). Furthermore, Mahar et al. (1997) demonstrated that those with higher levels of computer anxiety tended to complete simple computer tasks more slowly.

Some individuals have explained the gender-based differences in attitude regarding computer technology as an outcome of the socialization process (Cooper & Weaver, 2003). The basic framework of this explanation is that a sex-role is first formed within the family, wherein norms are internalized. Subsequently, attitudes are learned and a self-image is acquired. Later, these behaviors are shaped and reinforced in the scholastic environment and in the workplace where society's underlying culture is transmitted. As such, gender differences in attitudes towards computers may ultimately be a reflection of differing social-cultural expectations and experiences. That is, in general, gender differentiation is a product of the social construction that determines what models of correct behavior are given to members of each gender.

Ultimately, this rationale suggests that because society views computer use as masculine, females are simply not expected to become comfortable with computer use as readily as men. Indeed, it has been argued that the gender gap in male-dominated fields such as science and technology may be at least partly due to computer anxiety, as higher levels of computer anxiety would lead females to simply self-select out of these careers to begin with (Cooper & Weaver, 2003). For example, Teasdale and Lupart (2001) surveyed over 1300 students in grades seven through ten regarding their liking for computers, perceptions of their ability to use computers, amount of time spent per day using computers, and whether they might choose a career Information Technology. Their results demonstrate that, on average, males reported higher perceived ability, more liking for computers, and more time spent per day on computers than females. Males were also found to be more likely to select a career in Information Technology than females.

Given the previously outlined findings, awareness of differential attitudes towards the use of computers may be an important factor to consider when developing and designing Internet-based survey initiatives or programs. As research has shown that even minimal involvement in computer-related activities can become an anxiety-provoking situation for those suffering from computer anxiety (Paxton & Turner, 1984), those prone to computer anxiety may read instructions improperly, ignore disclaimers more frequently, or respond in a half-hearted manner to a greater extent than those who do not suffer from computer anxiety. More recently, Durndell and Haag (2001), investigating the effects of gender on computer anxiety, demonstrated that males tend to report greater computer self-efficacy, lower computer anxiety, more positive attitudes towards the Internet and longer use of the Internet than females. The authors argue that the literature on attitudes and anxiety towards computers is likely to extrapolate to the Internet. Thus, if females generally exhibit more computer anxiety than males, one may expect a disruption the quality of data obtained from Internet administration and, which in turn, may negatively impact measurement equivalence across pencil-and-paper and Internet administration modes.

Given the recent surge in online surveying, it is not hard to imagine a scenario in which a researcher/practitioner gathers attitudinal data at Time 1 using traditional paper-and-pencil surveying techniques, then later implements Internet surveying at Time 2. If female employees' computer anxiety negatively impacts responding to the Time 2 survey, cross-mode measurement equivalence may break down as females may no longer respond similarly or view the dimensionality of the constructs in the same way across administration mode. Thus, it is possible that observed attitudinal differences in data sets between Times 1 and 2 may lead one to conclude that a shift in attitudes had occurred for females where, in reality, any observed differences in responding may have simply been a reflection of and reaction to computer anxiety.

The present study seeks to investigate whether differences in administration mode (Internet vs. paper-and-pencil) lead to differential survey/questionnaire responding for males and females. As no one has investigated cross-mode equivalence using a repeated-measures framework yet, this study seeks to further contribute to the literature by employing a repeated-measures design to investigate the impact of differing survey methodologies on a widely-used job satisfaction measure, the Job Descriptive Index (JDI: Smith, Kendall, & Hulin, 1969). The fundamental advantage of the repeated-measures design over a between subjects design is the reduction in error variance associated with individual differences (Robinson, 1981). That is, the use of a repeated measures design removes error variance from the denominator of the *F*-ratio test thereby increasing statistical power to detect mean differences. If gen- der leads to small systematic differences in cross-mode responding, a repeated measures design may be required to detect cross-mode differences as a between-subjects design may lack the power to discern any existing differences.

The problems outlined in the preceding section can serve as a basis for a set of hypotheses concerning quality of male- and female-generated data collected via traditional paper-and-pencil administration methods compared with that obtained via the Internet.

Hypothesis One: Satisfaction ratings will demonstrate measurement equivalence across administration modes for males, but not females.

Hypothesis Two: There are no statistically significant differences in scale means over the modes of administration for males, but significant scale mean differences will be found for females.

Hypothesis Three: There are no statistically significant differences in item means over the modes of administration for males, but significant item mean differences will be found for females.

2. Method

2.1. Participants

Survey responses were obtained from 117 employed MBA and doctoral students from a private, mid-sized Midwestern university. The mean age of the participants

was 29.37 (SD = 7.25) and the sample was split between genders; male 53%, females 47%.

2.2. Measures

Participants anonymously completed all six items from each of three subscales of the JDI (*Promotions*, *Supervision*, and *Work Itself*) on both a paper-and-pencil and Internet version of the survey. The JDI was chosen because it is considered one of the most carefully constructed measures of job satisfaction in use today (Rosznowski, 1989). These scales and items were chosen on the basis of Gregson's (1987) findings that using the six items that loaded highest on each dimension, even when converted into a five-point scale, loaded the same as they did for Smith et al. (1969) when the test was developed. As such, researchers can abbreviate the JDI and use a Likert ratings scale (*strongly agree* to *strongly disagree*) without interfering with dimensionality. Furthermore, these scales demonstrated high internal consistency (.90, .86, and
.84, respectively), and do not inter-correlate very highly; (*Promotions – Supervision*, .41), (*Promotions Work Itself*, .45), and (*Supervision – Work Itself*, .37).

In order to minimize memory effects across administrations, the 18 items making up the three subscales were first embedded within 52 filler items for the paper-and-pencil administration, then embedded within 52 different filler items for the Internet administration. The 18 JDI items maintained their order and position between administrations, with the filler items serving to ''mask'' their repeated presence. For this task, all filler items for both versions were taken from the Devine Inventory (Conoley & Impara, 1995) and set to the same five-point scale as the JDI items. As analysis was only carried out on the 18 JDI items, data from filler items were omitted from analysis.

In order to model both traditional paper-and-pencil and Internet surveying conditions while counterbalancing for order effects, the following procedure was used: During class time, 62 of the 117 participants (53%) completed the paper-and-pencil survey first. To each survey was attached an identification number which they were instructed to take with them as they would be asked to supply this number during Internet surveying. Once paper-and-pencil surveying had been completed, participants were informed that they would receive an email (the address was supplied to the experimenter by the participants) one week from the then-current date that contained a link to the Internet survey. They were further informed that the Internet survey would remain accessible 24 h a day, for a seven-day period, and the survey may be taken at any point during this one-week period. The email containing the embedded link and access password to the survey was sent via email one week later. The paper- and-pencil and Internet versions of the survey were identical in format, length, and appearance. The only differences between the surveys were the masking items.

The other 55 participants (47%), received essentially the same treatment in the reverse order. Participants were informed that they would receive an email with an embedded link to the survey, which they had one week to complete. Participants were instructed that Internet surveying would be done first and paper- and-pencil administration would be done two weeks from the then-current date.

The email sent to participants followed the same procedure as above. The emailed link took them to the instruction page, and the instruction page was linked to the survey page. Two weeks later, the experimenter returned to the class and paper-and-pencil administration was carried out following the same procedure outlined above.

3. Results

Hypothesis One postulates that job satisfaction ratings will demonstrate measurement equivalence across administration modes for males, but not females. In order to examine measurement equivalence, the author followed the sequence of analyses indicated by Vandenberg and Lance (2000). The first analysis investigates any differences in covariance matrices across both samples. Equivalence of covariance matrices indicate that identical models will fit each dataset well and that the parameter estimates will be comparable across samples. As such, no further testing is required. If the covariance matrices are not equivalent, subsequent tests can determine the source of the difference across samples.

Upon comparing the covariance matrices across the two administration modes for the female-generated data, the v^2 significance test indicated that the covariance matrices are non-equivalent across the administrations, $v^2(244) = 349.23$, $p < .01$, see Table 1. However, given this statistic's extreme sensitivity to model complexity, sample size, and model misspecification, many researchers have recommended assessing model fit based on various absolute and incremental fit indices (Bentler & Bonett, 1980; Cheung & Rensvold, 2001; Ding, Velicer, & Harlow, 1995; Hu & Bentler, 1999). The obtained fit indices indicated acceptable fit for the model testing covariance matrices across administration modes (CFI = .90 and RMSEA = .08). Furthermore, more traditional guidelines also corroborated model fit. For example, Carmines and McIver (1981) suggested that a v^2/df ratio of no more than three serves as an adequate indicator of good fit. Based on this indicator, the obtained cross-mode covariance matrices are equivalent (v^2/df ratio = 1.43). Similarly, comparing male-generated data across administration modes resulted in a significant v^2, $v^2(270) = 406.38$, $p < .01$, yet acceptable fit indices and v^2/df ratio (CFI = .91, RMSEA = .09, v^2/df ratio = 1.51).

Table 1
Fit statistics across internet and paper and pencil samples

Phase	v^2	CFI	RMSEA	v^2/df ratio
Covariance				
Female	349.23**	0.90	0.08	1.43
Male	406.38**	0.91	0.09	1.51

CFI, comparative fit index, RMSEA, root mean square error of approximation.
** $p < .01$.

Table 2
Mean scale responses SD, and alpha reliabilities from internet and paper-and-pencil samples

	Scale	Paper-and-pencil		Internet		*p*
		M	SD	*M*	SD	
Males	Supervision	3.55	.81	3.58	.79	.71
	Promotions	3.17	.88	3.21	.86	.56
	Work itself	3.76	.88	3.72	.87	.44
Female	Supervision	3.67	.95	3.60	.87	.25
	Promotions	3.22	.82	3.17	.72	.47
	Work itself	4.10	.78	4.11	.73	.44

Given that the covariance matrices were equivalent across administration modes for both male and female respondents, as indicated by the fit indices, it was concluded that the sexes use the same psychological metric when providing job satisfaction ratings regardless of Internet or paper-and-pencil administration. Thus, Hypothesis One was partially supported.

Hypotheses Two and Three stated that there are no statistically significant differences in scale or item means over the modes of administration for males, but significant scale or item mean differences would be found for females. Bonferroni adjusted *t*-tests (applied to control for possible experiment-wide error in conducting multiple *t*-tests) showed that there were no statistically significant differences between the

Table 3
Mean item responses from internet and paper-and-pencil samples

Item	Males				*p*	Females				*p*
	P&P		Internet			P&P		Internet		
	M	SD	*M*	SD		*M*	SD	*M*	SD	
1	3.28	1.05	3.18	1.01	0.36	3.39	1.29	3.09	1.23	0.03
2	3.59	1.16	3.59	1.08	1.00	3.45	1.26	3.61	1.23	0.22
3	3.39	1.08	3.64	0.93	0.08	3.55	1.12	3.62	0.93	0.58
4	4.00	0.98	3.84	1.00	0.14	4.13	0.98	3.91	1.01	0.02
5	3.67	1.17	3.72	1.06	0.64	4.04	1.19	3.79	1.06	0.01
6	3.40	1.21	3.52	1.05	0.39	3.52	1.33	3.59	1.14	0.60
7	3.07	1.23	3.08	1.22	0.32	2.59	1.04	2.64	0.98	0.26
8	3.08	1.10	3.11	1.03	0.42	2.70	1.09	2.68	1.03	0.66
9	3.17	1.14	3.15	1.13	0.66	2.67	1.14	2.69	1.15	0.57
10	3.82	1.22	3.82	1.19	1.00	3.89	1.00	3.75	1.05	0.04
11	3.05	1.20	3.13	1.18	0.11	2.73	1.10	2.91	1.18	0.08
12	3.63	0.89	3.90	0.96	0.02	3.68	1.13	3.79	0.89	0.38
13	4.48	0.62	4.48	0.62	1.00	4.54	0.54	4.46	0.63	0.21
14	3.51	1.23	3.41	1.19	0.11	3.89	1.11	3.93	1.01	0.69
15	3.59	1.13	3.48	1.19	0.21	4.05	1.06	3.98	1.13	0.40
16	3.64	1.11	3.67	1.14	0.66	4.00	1.10	4.09	0.94	0.17
17	3.54	1.13	3.58	1.12	0.62	3.96	0.94	4.05	0.91	0.28
18	3.72	1.19	3.74	1.14	0.83	4.20	0.86	4.20	0.86	1.00

Note: P&P, pencil-and-paper administration.

paper-and-pencil and Internet scale scores for males; *Supervision* (Ms = 3.55 and 3.58, SDs = .81 and .79, respectively), $t(60) = -.38$, $p < .71$, *Promotions* (Ms = 3.17 and 3.21, SDs = .88 and .86, respectively), $t(60) = .67$, $p < .56$, and *Work Itself* (Ms = 3.76 and 3.72, SDs = .88 and .87, respectively), $t(60) = .78$, $p < .44$. Similarly, no significant scale mean differences were found for data obtained from females; *Supervision* (Ms = 3.67 and 3.60, SDs = .95 and .87, respectively), $t(55) = 1.16$, $p < .25$, *Promotions* (Ms = 3.22 and 3.17, SDs = .82 and .72, respectively), $t(60) = .71$, $p < .47$, and *Work Itself* (Ms = 4.10 and 4.11, SDs = .78 and .73, respectively), t (60) = .78, $p < .44$. Table 2 summarizes these results. Furthermore, Bonferroni adjusted *t*-tests on the means of the items demonstrated no statistically significant cross-mode item-level differences for either males or females (Table 3). Thus, Hypotheses Two and Three were partially supported.

4. Discussion

In this study, participants using the Internet to respond to an attitudinal measure responded in identical ways to their responses on a paper-and-pencil attitudinal measure, regardless of gender. Results of the present study support the conclusion of measurement equivalence and equivalent scale and item means across the two subscales measured for both males and females. Given the recent surge in organizations, clinicians, and researchers involvement in Internet data collection (Buchanan & Smith, 1999a, 1999b; Church, 2001; Thompson et al., 2003), the issue of measurement equivalence becomes increasingly salient.

In terms of organizational data collection, many times a decision to implement a given organizational intervention is based on means obtained from surveys/questionnaires (Fenlason & Christianson-Demay, 2002). Because the implementation of a given intervention could be a costly, time-consuming process for an organization, the equivalence of data quality across the two modes and genders becomes highly consequential. For example, it is not outside the realm of possibility that an organization finds itself in a situation similar to that of the previously outlined scenario; gathering attitudinal data with traditional means, then initiating online surveying, and subsequently combining datasets from both administration methods. Had results of this study suggested that gender differences exist for cross-mode responding, common sense would dictate that separate analyses of data, according to administration mode, would be required. If separate analyses are ignored and the non-equivalent data are combined into one data set and reported, respondents true standing on the constructs of interest may be obscured and those responsible for human resource/ management decisions may feel compelled to implement an unnecessary, costly intervention on the basis of combining non-equivalent data sets. Being put in the position of having to carry out separate analyses could force a human resources department to devote more time and manpower to analysis and interpretation of the two separate data sets, adding to the aforementioned expense. In the end, a lack of measurement equivalence would imply that organizations should pick one and only one method of administration in the interest of fiscal responsibility. While this study

employed the use of an attitudinal measure frequently used in organizations, this scenario almost certainly plays itself out in other domains that frequently collect data from participants (i.e., clinical, educational, or research settings), and carries with it the same attendant dangers (time lost, costs, poor data quality, etc.).

Fortunately, the results from the current study demonstrate measurement equivalence in terms of factor form, factor structure, and equivalent means across gender and administration mode. These results imply that: (1) if females as a group do in- deed harbor more intense feelings of computer anxiety than males, these feelings do not seem to manifest themselves in terms of disrupted data quality, and (2) practitioners currently surveying using both methods can be assured that the combining of data sets from both administration methods is a safe, valid procedure for both males and females. Furthermore, these results suggest that any decision to implement a given intervention based on means analysis from compiled data would be based on sound psychometric properties.

As always, principles of good scientific investigation demand that studies be replicated in part or whole. As such, future replications are suggested across relevant populations (i.e., full time employees, different kinds of organizations, different education levels, white vs. blue collar occupations, etc.) to determine the reliability of these findings. While this research garnered more statistical power by way of using a repeated-measures design rather than a between-subjects design, this study nevertheless has certain limitations that may restrict its generalizability. Perhaps data obtained from employed male and female graduate students are more prone to invariant factor form and scale/item means across Internet and paper-and-pencil administrations than those whose occupation requires less education and less technological prowess. Generalizability of these findings may be an avenue for future research.

Results of the present study largely imply that there is no discernable difference in factor structure and scale or items means from data obtained from paper-and-pencil and Internet administrations over the genders. However, the presumption cannot be made that comparison of Internet and computerized administrations would lead to similar results. While many studies have demonstrated equivalence between computerized and paper-and-pencil surveys (i.e., Donovan, Probst, & Nelson, 2001; Dilalla, 1996; Mead and Drasgow, 1993; Potosky and Bobko, 1997) and other studies have found equivalence between Internet and paper-and-pencil administration (i.e., Buchanan & Smith, 1999a, 1999b; Burnkrant & Taylor, 2001; Collins et al., 2000; Etchegaray et al., 2003; Facteau & Craig, 2001; Fenlason, 2000; Magnan et al., 2000; Maurer et al., 1998; Pasveer & Ellard, 1998; Ployhart et al., 2003; Ryan et al., 1999; Spera & Moye, 2001; Stanton, 1998), a relative dearth of literature exists exploring the measurement equivalence of computerized and Internet studies. Computerized and Internet administrations differ methodologically in several important ways: Computerized administration is usually done under very controlled conditions; for example, in a quiet room at a designated time in the presence of an administrator and with a brief instruction period before administration. In contrast, Internet administration differs from computerized and paper-and-pencil administrations in that Internet administration of an online questionnaire may be accessed

from any number of locations at virtually any time, relatively free of controls. As such, Internet surveying, if done in the participant's home or office, would most likely be unmonitored by an administrator. In such a case, the participant may be less likely to read instructions properly or may ignore disclaimers more frequently than in monitored, face-to-face situations (Barak & English, 2002). Given these differences in administration conditions, future research should explore the measurement equivalence of Internet and computerized administration methods to address this issue.

In conclusion, data collected over Internet and paper-and-pencil settings are largely equivalent in terms of factor structure and obtained scale and item means for both males and females.

References

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

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*, 588–606.

Brosnan, M. J. (1998). The impact of psychological gender, gender-related perceptions, significant others and the introducer of technology upon computer anxiety in students. *Journal of Educational Computing Research, 18*, 63–78.

Buchanan, T., & Smith, J. L. (1999a). Research on the Internet: validation of a world-wide web mediated personality scale. *Behavior Research Methods, Instruments & Computers, 31*, 565–571.

Buchanan, T., & Smith, J. L. (1999b). Using the Internet for psychological research: personality testing on the world-wide web. *British Journal of Psychology, 90*, 125–144.

Burnkrant, S. R., Taylor, C. D. (2001). *Equivalence of traditional and Internet-based data collection: three multigroup analyses*. Paper presented at the 16th annual conference of the society of industrial and organizational psychology, San Diego, CA, April, 2001.

Carmines, E., & McIver, J. (1981). Analyzing models with unobservable variables. In G. W. Bohrnstedt & E. F. Borgotta (Eds.), *Social measurement: Current issues* (pp. 65–115). Beverly Hills: Sage.

Cheung, G. W., & Rensvold, R. B. (2001). The effects of model parsimony and sampling error on the fit of structural equation models. *Organizational Research Methods, 4*, 236–264.

Church, A. H. (2001). Is there a method to our madness. The impact of data collection methodology on organizational survey results. *Personnel Psychology, 54*, 937–962.

Collins, W. C., Raju, N. S., & Edwards, J. E. (2000). Assessing differential functioning in a satisfaction scale. *Journal of Applied Psychology, 85*, 451–461.

Cooper, J., & Weaver, K. (2003). *Gender and computers: Understanding the digital divide*. Lawrence Erlbaum Associates, Inc.

Conoley, J. C., & Impara, J. C. (Eds.). (1995). *The twelfth mental measurements yearbook*. Buros Institute of Mental Measurements, Lincoln: University of Nebraska.

Ding, L., Velicer, W., & Harlow, L. (1995). Effects of estimation methods, number of indicators per factor, and improper solution on structural equation modeling fit indices. *Structural Equation Modeling, 2*, 119–144.

Donovan, M. A., Probst, T. & Nelson, N. (2001). Web-based attitude assessments: Does technology affect equivalence? Paper presented at the 16th annual conference of the society of industrial and organizational psychology, San Diego, CA, April, 2001, *Behavioral Research Methods, Instruments and Computers. 31*, 572–577.

Durndell, A., & Haag, Z. (2001). Computer self efficacy, computer anxiety, and attitudes towards the Internet and reported experience with the Internet, by gender, in an East European sample. *Computers in Human Behavior, 18*, 521–535.

Etchegaray, J. M., Sexton, J. B., Helmreich, R. L. & Thomas, E. J. (2003). *Job satisfaction ratings: Measurement equivalence across physicians and nurses*. Paper presented at the 18th annual conference of the society of industrial and organizational psychology, Orlando, LA, April, 2003.

Facteau, J. D., & Craig, S. B. (2001). Are performance appraisal ratings from different sources comparable?. *Journal of Applied Psychology, 86*, 215–227.

Fenlason, K. J. (2000). *Multiple data-collection methods in a 360 feedback program: Implications for use and interpretation*. Paper presented at the 15th annual conference of the society of industrial and organizational psychology, New Orleans, LA, April, 2000.

Fenlason, K. J. & Christianson-Demay, C. (2002). *Paper-and-pencil and web surveys: Different methods, same results?* Paper presented at the 17th annual conference of the society of industrial and organizational psychology, Toronto, April, 2002.

Frericks, L., Ehrhart, K. & O'Connell, M. (2003). *Computer anxiety and test performance: Comparing selection test formats*. Paper presented at the 18th annual conference of the society of industrial and organizational psychology, Orlando, April, 2003.

Glass, C., & Knight, L. (1988). Cognitive factors in computer anxiety. *Cognitive Therapy and Research, 12*, 351–366.

Gregson, T. (1987). Factor analysis of a multiple-choice format for job satisfaction. *Psychological Reports, 61*, 747–750.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structural analysis: Conventional versus new alternatives. *Structural Equation Modeling, 6*, 1–55.

Liu, M., Reed, W. M., & Phillips, P. D. (1992). Teacher education students and computers: major, prior computer experience, occurrence, and anxiety. *Journal of Research on Computing in Education, 24*(2), 457–467.

Magnan, S. M., Lundby, K. M. & Fenlason, K. J. (2000). *Dual media: The art and science of paper and internet employee survey implementation*. Paper presented at the 15th annual conference of the society of industrial and organizational psychology, New Orleans, LA, April, 2000.

Mahar, D., Henderson, R., & Deane, F. (1997). The effects of computer anxiety, state anxiety, and computer experience on user's performance on computer-based tasks. *Personality and Individual Differences, 22*, 683–692.

Maitland, S., Dixon, R., Hultsch, D., & Hertzog, C. (2001). Well-being as a moving target: Measurement equivalence of the Bradburn Affect Scale. *Journal of Gerontology (Series B), 56*, 69–78.

Maurer, T. J., Raju, N. S., & Collins, W. C. (1998). Peer and subordinate performance appraisal measurement equivalence. *Journal of Applied Psychology, 83*, 693–702.

Meier, S. (1985). Computer aversion. *Computers in Human Behavior, 1*, 171–179.

Pasveer, K. A., & Ellard, J. H. (1998). The making of a personality inventory: Help from the WWW. *Behavioral Research Methods, Instruments and Computers, 30*, 309–313.

Paxton, A., & Turner, E. (1984). The application of human factors to the needs of the novice computer user. *International Journal of Man-Machine Studies, 20*, 137–156.

Ployhart, R. E., Weekley, J., Holtz, B., & Kemp, C. (2003). Web-based vs. paper-and-pencil testing: Are test scores comparable across applicants and incumbents?. *Personnel Psychology, 56*, 733–752.

Robie, C., Zickar, M., & Schmidt, M. (2001). Measurement equivalence between applicant and incumbent groups. An IRT analysis of personality scales. *Human Performance, 14*, 187–207.

Robinson, P. W. (1981). Multigroup designs with one independent variable. In A. D. Gnerre (Ed.), *Fundamentals of experimental psychology* (2nd ed.) (pp. 253–281). Englewood Cliffs: Prentice-Hall.

Rosen, L. D., & Weil, M. M. (1994). Computer availability, computer experience and technophobia among public school teachers. *Computers in Human Behavior, 11*(1), 9–31.

Rosen, L., & Weil, M. (1995). Adult and teenager consumer users of technology: Potholes on the Information Superhighway?. *Journal of Consumer Affairs, 29*, 55–84.

Rosznowski, M. (1989). Examination of the measurement properties of the Job Descriptive Index with experimental items. *Journal of Applied Psychology, 74*, 805–814.

Ryan, A. M., Chan, D., Ployhart, R. E., & Slade, L. A. (1999). Employee attitudes surveys in a multinational organization: Considering language and culture in assessing measurement equivalence. *Personnel Psychology, 52*, 37–58.

Smith, P. C., Kendall, L. M., & Hulin, C. L. (1969). *The measurement of satisfaction in work and retirement*. Chicago, IL: Rand McNally.

Spera, T. J. & Moye, B. G. (2001). *Measurement equivalence between paper and web survey methods in a multinational company*. Paper presented at the 16th annual conference of the society of industrial and organizational psychology, San Diego, CA, April, 2001.

Stanton, J. M. (1998). An empirical assessment of data collection using the Internet. *Personnel Psychology, 51*, 709–724.

Teasdale, S. & Lupart, J. (2001). *Gender differences in computer attitudes, skills, and perceived ability*. Presented at the Canadian society for studies in education Quebec, May, 2001.

Thompson, L. F., Surface, E. A., Martin, D. L., & Sanders, M. G. (2003). From paper topixels: Moving personnel surveys to the web. *Personnel Psychology, 56*, 197–227.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 4*, 4–70.

Williams, S., Ogletree, S., Woodburn, W., & Raffeld, P. (1993). Gender roles, computer attitudes, and dyadic interaction performance in college students. *Sex Roles, 29*, 515–525.

Yun, G. W., & Trumbo, C. W. (2000). Comparative response to a survey executed by post, email, & web form. *Journal of Computer-Mediated Communication, 6*, 1–26.