Examining Contexts-of-Use for Web-Based and Paper-Based Questionnaires

Patricia L. Hardré, H. Michael Crowson and Kui Xie

Educational and Psychological Measurement 2012 72: 1015 originally published online 9 July 2012
DOI: 10.1177/0013164412451977

Educational and Psychological Measurement
72(6) 1015–1038

DOI: 10.1177/0013164412451977
http://epm.sagepub.com

Examining Contexts-of-Use for Web-Based and Paper-Based Questionnaires

Patricia L. Hardré[1], H. Michael Crowson[1], and Kui Xie[2]

Abstract

Questionnaire instruments are frequently administered in digital formats, largely web-based, without much systematic investigation of possible effects from these administration methods. Furthermore, little attention has been given to the contextual lack of control for extraneous factors that may influence user responses. In this study, 263 university students were randomly assigned to one of two administration formats, web-based (WBA) or paper-based (PBA), to complete a set of questionnaires in an environment of their choice. Data collection included reporting context characteristics along three parameters: location, companions, and concurrent activities (including help-seeking). Outcomes of interest included location and conditions of user-chosen contexts, instrument performance, generative data quantity and quality, independence of completion, administrative efficiency, and participant affect. Participants did choose and allow distracters in their contexts-of-use, completing the questionnaires while engaged in multiple social and asocial concurrent activities. There were generally small but significant differences in instrument performance and user response characteristics by administration method and contexts-of-use. Participant comfort and data returned were both higher in PBA than WBA. Quantity return of generative data was higher in WBA while overall quality (completeness, coherence, correctness) of generative data was not significantly different. These findings present implications of administrative methods and contextual influences that inform measurement professionals' selection and design of administrative systems and conditions for research and evaluation data collection.

[1]University of Oklahoma, Norman, OK, USA
[2]Ohio State University, Columbus, OH, USA

Corresponding Author:
Patricia L. Hardre, 820 Van Vleet Oval, University of Oklahoma, Norman, OK 73019-2041, USA
Email: hardre@ou.edu

Keywords

administrative method, comparison study, experimental design, measurement system

Web-based surveys and questionnaires are commonly used for data collection, in studies from local to national in scope, conducted by public and private entities (Shin, Johnson, & Rao, 2011). Online research is tremendously popular, based on its potential to improve access to participants and streamline data management for researchers (Hayslett & Wildemuth, 2004). These benefits vary based on the types of measurement instruments and technological systems used (Hine, 2005). Previous empirical studies have addressed some effects of digital administrative methods on the performance of psychological research instruments, including user affect and perceptual issues such as perceived anonymity and self-disclosure (McCabe, 2004; Moon, 2000; Sarrazin, Hall, Richards, & Carswell, 2002) and system-related differences such as response rates (Idleman, 2003).

However, very few have taken into account the potential effects of context beyond the system itself, including social and asocial distractions that can bias or contaminate research data. As moving research to online systems leaves the choice of broader social and asocial context factors in the hands of participants, it is essential that researchers know what effects such factors may have on data. This research investigates potential differential effects of web-based administration (WBA) and paper-based administration (PBA) of questionnaires, with particular attention to the context and conditions in which they were completed. It adds to the existing research by conceptually broadening context-of-use beyond the design of materials and systems, to include external context. Outcomes of interest include participants' choice of location and conditions for completion, instrument performance, quantitative and generative data quality, and participant affect.

Background

Computer-mediated communication reciprocally influences people's lives (Dodge, 2005; Dodge & Kitchin, 2001). Researchers and practitioners frequently use digital administration methods for surveys and questionnaires for studies on a variety of topics (Ciolek, 1998; Dillman, 2000; Granello & Wheaton, 2004; Noyes & Garland, 2008; Zutshi, Parris, & Creed, 2011). Key reasons include lower costs and increased access to participant groups (Hayslett & Wildemuth, 2004; Shin et al., 2011). Web-based questionnaires are also commonly used for workplace and instructional performance feedback (Sears, Prakash, & Chiocchio, 2003; Thompson, Surface, Marton, & Sanders, 2003), in part because participant responses can easily be de-identified (Schulenberg & Yurtzenka, 1999). As instruments and administration systems proliferate, both must be evaluated based on data quality and measurement precision (Schonlan, Fricker, & Elliott, 2002). Studies addressing system effects on humans

should include implications for affect and emotions, which may influence not only engagement and completion of measures but also actual responses.

The majority of existing studies on comparability of findings across administration methods (Dillman & Bowker, 2001; Ritter, Lorig, Laurent, & Matthews, 2004) have done so in controlled settings (e.g., Hardré, Crowson, Ly, & Xie, 2007; Sarrazin et al., 2002; Truman et al., 2003). Today, many questionnaires and surveys are completed in contexts and environments not controlled by researchers but freely chosen by users, which must be considered uncontrolled contexts-of-use. Of those published studies conducted in uncontrolled environments, most have not included any characteristics of user-chosen contexts-of-administration. For both paper and digital instruments, the context may be a public or social space, and conditions may include interacting with other people and engaging in other concurrent activities (Hardré, Crowson, & Xie, 2010). Among the range of potentially influential factors that surround research administration, context factors outside of researcher control may have a direct impact on participants during instrument completion. At this time, however, little is known about those factors or about their potential consequences for data quality.

Most researchers and evaluators take advantage of benefits from digital administration methods, with only minimal attention to the critical trade-offs that may be implicit in these choices (Dillman & Bowker, 2001; Leece et al., 2004). Most experimental research on administration systems has been carried out using convenience samples comprised of college students from a single institution (Carini, Hayek, Kuh, Kennedy, & Ouimet, 2003). More empirical studies of administrative methods are essential (Hine, 2005; Shin et al., 2011), especially on details of administrative context beyond the internal features of systems themselves (Hardré, Crowson, et al., 2010).

Characteristics of Administration Systems

This study compares traditional PBA methods to digital WBA methods in authentic contexts-of-use. Both systems are in widespread use by researchers and practitioners, despite a relative lack of empirical studies of data quality and comparability (Dillman & Bowker, 2001; Hardré, Xie, & Ly, 2005). This relative dearth of research is of particular concern in light of administrative systems' increasing diversity. Characteristics of available administration systems vary widely, and detailed information on system characteristics is necessary to understand the research. Given recent advances in portability of digital technology, both methods are now equally susceptible to a wider variety of contextual influences.

Paper-based administration. Traditional use of printed questionnaire instruments generally occurs with an individual hardcopy given to each research participant, completed using a conventional writing implement (pen or pencil), and returned to the researcher in the original form. PBA questionnaires are constrained to stable formats, exerting implicit control on users' experience.

Web-based administration. Digital questionnaires are generally completed online, by users navigating to the experimental site, logging into the interface, then reading and responding to items by clicking on or selecting from menus, or alternately typing in generative text (with a range of variations in system features). WBA questionnaires range from highly interactive and dynamic instruments with graphic displays and nested prompts to digital versions of text-based questionnaires with identical layout and presentation.

Benefits of Digital Administrative Systems

WBA systems are widespread and increasing in use (McCabe, 2004; Wolfe, Converse, Airen, & Bodenhorn, 2009). They offer more revision flexibility than PBA and enable better efficiency both in data collection and in data management and analysis (Hardré et al., 2005; Schonlan et al., 2002). WBA has the advantage of delivering questionnaires to participants at key moments over long-term projects and at salient measurement junctures (Hardré, Nanny, Refai, Ling, & Slater, 2010). WBA may produce higher rates of response than paper (Bälter, Bälter, Fondell, & Lagerros, 2005; McCabe, 2004), but response *across* WBA survey studies varies widely, from 7% to 44% (Schonlan et al., 2002). A number of recent studies favor PBA over WBA for unit/overall (vs. item-level) response rates (Jacob, 2011; Shin et al., 2011; Wolfe et al., 2009). Web-based options may save researchers time to administer and reduce turnaround time, especially for geographically distributed populations (Goree & Marszalek, 1995; Schonlan et al., 2002). Some participants still choose to print off and complete digitally delivered questionnaires by hand, though this tendency varies by subgroups (Schonlan et al., 2002) and appears to be diminishing over time. In addition, with recent system and software advances, responses to WBA questionnaires are easily tracked, and individually prompting users to complete them is easier than ever. Current systems allow researchers to separate identity from responses invisibly, ensuring anonymity of information while tracking who has responded.

Differential Effects by Administration Method

Various studies have found differential effects by administration systems. These differences are alternately attributed to the study's instruments, task, or content; to system configurations and constraints; or to individual or group user characteristics. Digital administrations may increase negative affective responses (George, Lankford, & Wilson, 1992; Schulenberg & Yurtzenka, 1999) or increase perceived anonymity and facilitate disclosure of sensitive personal information and high-risk behaviors (Ferriter, 1993; Moon, 2000; Robinson & West, 1992; Tourangeau, 2004; Whittier, Seeley, & St. Lawrence, 2004). In some studies, WBA has demonstrated better test–retest reliabilities (Truman et al., 2003) and more positive response patterns across groups (Carini et al., 2003). In contrast, some other studies have found no significant

response difference between digital and paper attitude questionnaires, with participants either anonymous or identified (e.g., Sarrazin et al., 2002).

Across instrument types, WBA may require more cognitive load (mental effort) from participants than equivalent PBA (Noyes & Garland, 2008). In health care, digital instruments have produced differential responses by user groups, for patients (e.g., Sarrazin et al., 2002), but even more for physicians (e.g., Bälter et al., 2005). On workplace surveys and performance appraisals, study results for PBA and WBA are mixed (Thompson et al., 2003). These differences are often attributed to computer anxiety or aversion (Sears et al., 2003), but even these effects vary across multiple studies (Chua, Chen, & Wong, 1999) and covary with other characteristics (Tseng, Tiplady, MacLeod, & Wright, 1998).

Studies of affect and anxiety in PBA and WBA systems have demonstrated similar results in controlled studies (e.g., Hardré et al., 2007) and in field observations (Fletcher, Erickson, Toomey, & Wagenaar, 2003). Affect is also related to users' willingness to participate and to data quality (Hardré, Crowson, & Xie, 2010). People often respond to computers as entities with social identities or as extensions of the people behind or within digital systems (Dodge, 2005; Moon, 2000; Nass, Fogg, & Moon, 1996). This tendency to anthropomorphize digital systems and features is enhanced by personified or animated entities such as coaches and avatars. How specific systems, and design of the interface within those systems, reveal (or mask) information about people also influences responses (Bosnjak, Tuten, & Witmann, 2005; Joinson, 2005).

Trade-offs for Digital Administration

Differences in computer displays, functions, and browsers may influence participant responses through their effects on instrument presentation (MacIsaac, Cole, Cole, McCullough, & Maxka, 2002). Individual anxiety (vs. comfort) and system display characteristics can induce greater mental fatigue (Clariana & Wallace, 2002) or require higher cognitive load that detracts from users' task engagement and perceptual acuity (Noyes, Garland, & Robbins, 2004). WBA participants in uncontrolled contexts often fail to complete instruments (Dolenko, 1998; Porter & Whitcomb, 2003; Spink, Bateman, & Jansen, 1998), resulting in data loss that mars research outcomes (Hine, 2005; Schonlan et al., 2002). Response rates vary significantly across delivery methods (e.g., Bälter et al., 2005; Leece et al., 2004; Nichols & Sedivi, 1998) and are influenced by system characteristics, recruitment methods, and incentives used (Cobanoglu & Cobanoglu, 2003; Porter & Whitcomb, 2003). Digital administration may bias samples toward particular subgroups by age, education, technology skill, or social stratum (Bandilla, Bosnjak, & Altdorfer, 2003; Hayslett & Wildemuth, 2004; Noyes et al., 2004; Zhang, 1999). Authentication capability and security controls vary among systems, presenting risks and requiring diligent attention from researchers (Hine, 2005; MacIssac et al., 2002). Tracking features offer potential to map participation and interactions in virtual spaces during online data

collection (e.g., MacEachren, 1995; Yook, Jeong, & Barabási, 2002). Given their ready access and speed of processing, any technical glitches in digital systems can cripple a project faster than they can be discovered (Wilt, Condon, & Revelle, 2011). Most studies have focused on internal system features and few have included any external contexts-of-use, though these may prove equally influential on data-based outcomes.

Broadly Framing Contexts-of-Use

Participants receiving WBA are typically recruited via email or online links (Dillman, 2000; Schonlan et al., 2002), leaving them to choose any context and conditions in which to engage the system. With wireless networks widely and readily available, contexts for completing online questionnaires may include any indoor or outdoor space, public or private. Such spaces may contain any number of potential sources of distraction and interruption, from direct or indirect sources, intentional or unintentional. Many of these sources can skew, bias, or contaminate data. Consequently, they may invalidate the measures of important constructs and a study's results. Rarely are these contextual characteristics assessed or even considered in terms of their potential effects on data quality. The diversity of digital contexts-of-use is an emergent difference commensurate with the increased portability and connectivity of wireless digital tools.

A compelling reason to use online data collection is its ease and access from anywhere (Hine, 2005), but open access leaves equally open context choice with potential disruption and contamination. Research on user context choices and context features can provide richer and more nuanced information that may be valuable for evaluating the quality, accuracy, and meaningfulness of research data (Hardré, Crowson, et al., 2010; MacKay, 2005). In addition, online communication is a more fluid part of daily life and less ''virtual'' (other than normal) than it was just a decade ago (e.g., Miller & Slater, 2000; Thompson et al., 2003). Studying WBA data collection as authentically embedded in life-as-context yields more appropriately contextualized information about how research using digital systems makes sense in practice.

Findings on reliability and performance differences for research instruments by administration method are sensitive to both advances in technological capability and differences in individual study designs. Studies are needed that investigate previously unexamined issues that emerge with technological advances (Dillman & Bowker, 2001; Fox & Schwartz, 2002; Granello & Wheaton, 2004), such as context-of-use variables.

Data Quality

Data quality is a critical issue in research administration, one that has been defined and measured according to various criteria for method comparisons. Data loss is a

negative factor in overall data quality, whether it results from instrument noncompletions (unit response data; e.g., Denscombe, 2008; Greenlaw & Brown-Welty, 2009; Sax, Gilmartin, & Bryant, 2003) or from skipped items (e.g., Kiesler & Sproull, 1986; Shin et al., 2011; Wolfe et al., 2009). Data loss is important because it affects overall sample size or results in exclusion of cases in analysis (reducing sample size on longitudinal or repeated measures).

Another criterion for quality of quantitative data is reflected in instrument stability (e.g., internal consistency, measurement equivalence) within or between subjects (e.g., Hardré et al., 2007; Truman et al., 2003). Studies have alternately found better test–retest reliability for digital (vs. paper) self-report questionnaires (Truman et al., 2003) or little difference (Hardré et al., 2007; MacIsaac, 1999; Webster & Compeau, 1996). For quantitative subscales, the number of significant correlations among study variables is another indicator of data quality, because subscale correlations inform further analysis along with model and theory development (Hardré et al., 2007). The quality of qualitative data (such as generative data from open-ended questionnaire items) may be judged in part based on length of the response (e.g., Comley, 1996; Denscombe, 2008; Mehta & Sivadas, 1995), because longer responses generally provide the richer and more nuanced information that qualitative researchers seek. Along with length, another criterion for the quality of generative data is the substance or completeness of response; that is, whether participants addressed the question asked, because it indicates cognitive engagement in the task of responding (Hardré, Crowson, et al., 2010; Shin et al., 2011). One study explicitly compared the quality of generative responses from PBA and WBA, finding them longer in paper (but not significantly so; Denscombe, 2008). More research is needed to illuminate the dynamic of alternative administration methods and inform researchers' method selection by study designs, research questions, and data types (Carini et al., 2003). Since visual design, presentation, and contextual characteristics of questionnaires may influence participants' responses through interacting with administrative system characteristics (Dillman & Bowker, 2001; Schonlan et al., 2002), researchers designing comparison studies need to minimize format differences (Booth-Kewley, Edwards, & Rosenfeld, 1992).

The Present Study

In earlier studies comparing administration methods, we first controlled for all context features to the highest degree possible (Hardré, Crowson, Xie, & Ly, 2007), then varied context and diversified users to illuminate the possible effects of authentic contexts-of-use (Hardré, Crowson, & Xie, 2010). In the present study, we use a different range of measures and users and refine the instruments to address more details in user-chosen contexts. In the present study, we control for item presentation and ordering across administration methods rather than using the interactive options of the digital system. Exerting this degree of control diminishes the contextual ''authenticity'' of the WBA survey but enhances the clarity of our comparison

design. We use questionnaires assessing different psychological constructs, with different overall instrument design characteristics, all of which previously demonstrated consistent performance in both administration methods. Previous research demonstrated the importance of affect, comfort, and anxiety on administration (Sears et al., 2003; Schulenberg & Yurtzenka, 1999), so we included measures of all three factors.

Method

Participants

Our sample was composed of 263 traditional and nontraditional, undergraduate and graduate students in education and psychology courses at two public research universities (one in the Southwestern United States, one in the Southeast). Students received course credit for participation (with credit determined by course instructors). Anonymity was maintained on all responses. The age range was 18 to 54 years (mean = 22). The breakdown of the sample by academic level was as follows: freshmen, $n = 3$ (1%); sophomores, $n = 46$ (18%); juniors, $n = 109$ (41%); seniors, $n = 3$ (1%); and graduate students, $n = 3$ (1%). This sample included 62 (23%) males and 200 (74%) females (reflecting gender percentages in their academic programs). Participants reported their ethnicity as follows: Anglo American/White, $n = 185$ (68%); African American/Black, $n = 50$ (18%); American Indian/Alaskan Native, $n = 5$ (2%); Mexican American/Latino, $n = 4$ (2%); and Multiracial, $n = 14$ (5%).

Procedures

Two identical sets of questionnaires were created in two different formats, one in traditional paper-based (PBA) and the other in a digital survey administration system (Survey Monkey®) to be delivered via a web link (WBA). Each format included the same demographics and questionnaire instruments using multiple response formats (Likert-type continuous response scales, selection, dichotomous, and open-ended), as well as parallel structure, format, presentation, and layout. Each format had two versions (A and B) with the questionnaire sets order-inverted to control for order effects.

Study participants at two institutional sites were recruited from classes and were asked for their consent to participate in the study. Those agreeing to participate were randomly assigned in equal numbers to one of the two conditions (WBA or PBA). Participants assigned to the PBA condition were given the paper questionnaire packets with instructions and a return envelope. Participants in the WBA group were given instructions to access the online questionnaire set via a website link. They were allowed to complete the questionnaires in any location or conditions of their choice. Included in the survey were items asking about the context and conditions under which the survey was completed. The questionnaire sets were designed to take 45 to 60 minutes to complete, in either format. Participants were given 1 week to submit their responses.

Controls

Random assignment addressed concern over selection effects. Including the dual order-inverted forms in each condition addressed questionnaire order effects. Using identical questionnaires across both administration formats controlled for instrumentation effects. Maintaining anonymity on all data controlled for socially desirable responding. Including both positively and negatively worded items (recoded in analysis to produce coherent subscales) controlled for acquiescence responses. The single-event design addressed effects from maturation and history. Administering questionnaires during the same time period on both sites controlled for timing effects.

Data Processing

Data from the WBA output were downloaded from the system database and converted to SPSS© format. Variable names were edited to align with subscale identifiers. PBA data responses were hand-entered into SPSS in a template identical to the WBA data. Qualitative item responses were independently scored by two researchers using a standard rubric, compared for consistency, and entered into the master data sets. Both complete data sets were checked against the original sources for accuracy. They were merged into a comprehensive data set for analysis.

Instruments

All instruments were previously tested, paper-based, questionnaire instruments reflecting a sampling of constructs common in educational and social sciences research, including beliefs, values, perceptions, and motivation. Quantitative item responses were provided using Likert-type numeric scales, varying in scale ranges, anchor points, and item layout. The instruments have previously been demonstrated as valid for assessing their respective target constructs and have shown adequate item and subscale performance in similar participant groups. Four open-response items were included to assess differences in generative text responses.

Computer Competence Scale (CCS). Participants reported their perceived competence for using web-based and non-web-based computer applications. The CCS (Hardré, Crowson, et al., 2010) used a Likert-type scale (1 = *not at all competent* to 7 = *very competent*). Sample item: ''How competent, overall, do you feel about using the web?''

Technology Interaction Profile (TIP). The TIP assesses a set of behaviors and characteristics that predict how individuals respond to technology-based systems: computer anxiety and social behaviors with technology. The multidimensional TIP uses both 1- to 7-point Likert-type and frequency response scales. Sample items: ''Do you have any anxiety, in general, about using computers?'' (1- to 7-point Likert-type scale) and ''How often do you text or chat on technology tools during class?'' (5 responses from ''More than once a day'' to ''Never'').

Indices of Conscientiousness and Persistence. We indexed conscientiousness and persistence using items taken from the International Personality Item Pool (located at http://ipip.ori.org/), a scientific collaboratory for personality researchers (Goldberg et al., 2006). Conscientiousness was measured using 13 items, whereas Persistence was measured with 8 items. All items relied on a 1- to 7-point Likert-type response scale, with anchors of 1 = *very inaccurate* and 7 = *very accurate*. Sample items: ''I generally pay close attention to details'' (Conscientiousness); ''I finish things despite obstacles in the way'' (Persistence).

Need for Cognition Scale–Short Form (NFC). The 18-item NFC (Cacioppo, Petty, & Kao, 1984) scale assesses a person's interest and enjoyment in engaging in effortful cognitive activity (Cacioppo & Petty, 1982). Items for this scale rely on a 1- to 7-point Likert-type scale, with anchors of 1 = *strongly disagree* and 7 = *strongly agree*. Sample item: ''I really enjoy a task that involves coming up with new solutions for problems.''

Need for Structure Scale (NFS). The 11-item NFS (one item is typically not included in the scoring) assesses an individual's desire to live in an orderly way, that is, with a high degree of order, structure, and predictability (Neuberg & Newsom, 1993; Thompson et al., 2003). This scale presents items on a 1- to 7-point Likert-type scale (anchored from 1 = *strongly disagree* to 7 = *strongly agree*). Sample items: ''I enjoy having a clear and structured mode of life'' and ''I don't like situations that are uncertain.''

Comfort measure. This 10-item measure assesses students' self-reported affect for the administration method. Participants responded using a 7-point Likert-type scale with anchors of 1 = *very much untrue* and 7 = *very much true*. Sample items: ''I felt comfortable completing the questionnaires for this study'' and ''I felt free to respond honestly on the questionnaires for this study.'' Negatively worded items were recoded and summed with positively worded items to generate a unidimensional administration method comfort score, with higher scores reflecting *more comfort* for the condition.

Time on task measure. User start and finish times were used to calculate time-on-task (how long participants took to complete the questionnaires). The WBA condition provided tracking data, but the PBA necessarily depended on participants' self-reporting. Prompts to users on the PBA to record start time were provided at the beginning of the questionnaire set and the prompt to record finish time was the last item at the end. Times were entered into the database and time-on-task computed as number of minutes elapsed from start time to finish time.

Quality of generative data. Two dimensions were used to judge the quality of participants' generative response on the four open-ended items: quantity of response (as length in words) and quality of response (as rater scores based on a standard rubric). The measure of *quantity* was the total number of words in each participant's response over the four questions. Count was made using a strategic trim function for the Excel files (http://www.timeatlas.com), then totals were transferred to SPSS for analysis. The measure of *quality* was the composite of two sets of independent raters' scores

across the four open-ended items. The rubric operationalized quality on three key criteria: completeness (response addressed the question fully and clearly), coherence (statements supported by relevant evidence and examples without unrelated tangents), and correctness (response is stated without technical errors or shortcuts). Each criterion was scored as high = 3, mid = 2, low = 1, none = 0 (3 criteria × 2 raters = 6 scores per response × 4 items = 24 scores @ 3-pts = range of 0-72 score points).

Contexts-of-Use

Our original contexts-of-use instrument captures characteristics of the locations and conditions chosen by participants. Both groups of participants reported the location and circumstances under which they completed the questionnaires, including specific characteristics of those contexts. The instrument provides a list of selection items (distilled from previous open-ended responses on these same behaviors) and also provides for users to fill in unique responses. It includes three separate categories of place characteristics: location and use, companions and social interactions, and concurrent activities. Each category also includes additional items (as multiple selection or dichotomous response opportunities), each of which supplies specific relevant details (generating 20 unique bits of information about the user context). The contexts-of-use instrument produces a Location Profile plus two separate indices of the context, the ''Distraction Index'' and the ''Assistance Index.''

Distraction Index. Participants' responses to eight items were used to construct an index of the degree to which the environment they chose for completing the questionnaires contained explicit distracters. The items were originally coded as ''yes/no.'' Participants' ''yes'' responses were summed such that higher scores indicated greater levels of distraction. Potential distracters included having the television on, engaging in conversation, listening to music, helping others, surfing the web, engaging in social networking, or talking on the phone (total of 7 possible).

Assistance Index. Participants' responses to three items were used to construct an index of assistance sought and received from others while completing the survey. The items comprising the scale are as follows: ''Did you talk to anyone about the items on the survey while you were completing it?'' ''Did you specifically seek advice from anyone about how to answer a question?'' ''Did you change your responses on any items because of the input or opinions of others?'' The items were coded as ''yes/no.'' Responses were summed so that higher scale scores indicated greater levels of assistance sought and received.

Research Questions

Research Question 1: What were users' choices for contexts-of-use in terms of location, companions, concurrent activities?

Research Question 2: Are there differences in instrument performance characteristics (internal reliabilities, number/strength of correlations among

measured variables, factor solutions) as a function of administration method and contexts-of-use?

Research Question 3: Do items in psychological measures exhibit invariant factor loadings between paper- and web-based administration conditions?

Research Question 4: Are there differences in user response characteristics (positive response tendencies, data loss/omissions, time on task/efficiency, personal affect, comfort/anxiety) as a function of administration method and contexts-of-use?

Research Question 5: In the generative text data, are there differences in response quantity (in length of response) or quality (in clarity, coherence, and completeness) as a function of survey administration method and contexts-of-use?

Results

Choices in Contexts-of-Use

Ninety-two percent of participants across the whole sample selected from the locations and conditions provided as selections in this instrument. Few chose ''none'' or added alternate options as fill-ins, indicating that the categories distilled from the previous generative data fit the range of choices well. We first present the whole-group frequency data here for their descriptive utility and then distill these factors to compute the relevant indices for statistical comparison by administrative method.

Location and conditions. With respect to *location*, 56% of users reported completing the survey while at home, whereas an additional 17% completed them in class (while instruction was occurring). A small minority of users completed the questionnaires at work (6%) and at school but not in class (6%), and only 3% completed them at someone else's home. Just 1% or less of participants completed the questionnaires at food services businesses or in outdoor spaces.

As to *concurrent activities and companions*, 42% of users were engaged in conversation. Thirty-two percent of users indicated having family or friends present in the room, and 36% indicated having the television on. Another 25% of users reported listening to music while completing the questionnaires, 21% talking on the phone, and 21% surfing the web. Other participants reported being concurrently engaged in social networking (17%), helping others with tasks (13%), playing games (5%), and watching movies or videos (4%).

Clearly, these percentages reflect that most users were engaged in multiple concurrent activities, some of them highly interactive, requiring significant divided attention. These data indicate that most college students completing the questionnaires in uncontrolled contexts chose locations and conditions that presented multiple social and asocial distracters and potential sources of bias or interruption.

In terms of help-seeking, only 5.5% actually sought and received help on the questionnaires explicitly, and 3.3% reported changing their answers based on input from others. These data indicate that only a small minority of participants completing

questionnaire instruments in uncontrolled contexts actively engaged in help-seeking or changed their responses based on feedback from others.

Comparisons of Context Characteristics by Administrative Method

Beyond the general profile of context characteristics, we compared these choices of context between the participants in each of the administrative conditions, to illuminate any differences.

Mean differences on Distraction Index. A one-way ANOVA revealed a nonsignificant difference, $F(1, 191) = .140$, $p = .709$, $\eta_p^2 = .001$, between our two groups on the ''Distraction Index.'' Though there were abundant sources of distraction in the user contexts overall, there was little difference observed in the degree to which participants allowed distractions based on whether they completed the questionnaires on paper or online.

Companions and assistance. We tested for differences between administration methods (coded 0 = *paper*, 1 = *online*) concerning the presence or absence of others in the survey context (coded 0 = *absent*, 1 = *present*) while participants were completing the questionnaires. A chi-square test of independence revealed a significant, $\chi^2(1) = 4.361$, $p = .037$, relationship between friendship presence and administration method. Across methods, the proportion of participants with friends present was significantly greater than the proportion without friends present, paper-based, $\chi^2(1) = 6.125$, $p = .013$; online, $\chi^2(1) = 29.971$, $p < .001$. Between methods, the presence of friends was strongest for online (with 73% responding ''yes'' to the presence of friends) as opposed to paper-based administration (with just 60.9% reporting friends present).

Next, we tested for differences regarding participants' interaction with other people regarding the survey items themselves (''Assistance Index''). We found no statistically significant difference, $F(1, 242) = .761$, $p = .384$, $\eta_p^2 = .003$, in mean scores between groups on the amount of assistance they received while completing the surveys.

Instrument Performance Characteristics

With respect to instrument performance, we examined internal consistency and measurement equivalence. First we computed the internal reliabilities of the measures in the administration methods using structural equation modeling.[1] Following this we assessed whether there were significant differences in internal consistency estimates between paper- and web-based conditions. Finally, we tested for measurement equivalence between conditions using multigroup confirmatory factor analysis (Brown, 2006).[2]

Effects of administration condition on internal consistencies. We calculated internal consistencies for our measures of need for cognition, conscientiousness, personal need for structure, persistence, and comfort for participants in the paper-based and

online survey administrations. We used structural equation modeling (via AMOS™ 17.0) to estimate the reliability coefficients for our measures given that Cronbach's alpha provides a lower-bound estimate of reliability ''for more realistic congeneric measures'' (Fan, 2003, p. 37), such as those included in this study. The reliability coefficients we obtained by condition are as follows: conscientiousness (paper-based α = .839; online = .846); persistence (paper-based α = .731; online = .733); need for cognition (paper-based α = .759; online = .880); need for structure (paper-based α = .785; online = .810); comfort (paper-based α = .869; online = .870). To facilitate comparisons between conditions, we used the AMOS 17.0 bootstrapping option (500 bootstrapped samples) to generate bias-corrected 95% confidence intervals (CIs) for the estimates associated with each measure by condition. The only statistically significant difference observed between conditions in terms of internal consistency estimates were those associated with the need for cognition measure (95% CI for internet group [.830, .914]; 95% CI for paper group [.575, .884]).

Multiple-Group Confirmatory Factor Analyses: Invariance Tests

We used confirmatory factor analysis to assess the degree to which the items comprising our measures of need for cognition, personal need for structure, persistence, conscientiousness, and comfort with administration method functioned equivalently (i.e., were invariant) between the paper-based and online administrations. According to Kline (2005), the ''evaluation of measurement invariance typically involves the comparison of the relative fits of the χ^2_D statistic of two-factor models, one with cross-group equality constraints imposed on some of its parameters and the other without constraints'' (p. 295). As such, the invariance-testing process begins with a test of the difference between a fully constrained model (in which all factor loadings are constrained to equality between groups) and a fully unconstrained model (in which all loadings are freely estimated). In those cases where a significant overall difference in fit is observed, the next task is to go about testing each loading ''to determine the extent of partial measurement invariance'' (p. 295).

Fit statistics for our fully constrained and unconstrained multiple-group models, along with results of the $\Delta\chi^2$ tests, are included in Table 1. Using conventional standards for judging model fit (see Schumacker & Lomax, 2004), it is clear that the fit statistics for the fully constrained and unconstrained models between conditions and across measures were low, indicating poor fit to the data. No significant differences in factor loadings were apparent between the online and paper-based administration conditions for our measures of conscientiousness, persistence, personal need for structure, and comfort. On the other hand, we noted a statistically significant difference between the fully constrained and unconstrained models for the measure of need for cognition. Follow-up tests of the individual factor loadings revealed statistically significant ($p < .01$) differences between conditions for the loadings associated with Items 5, 6, and 15.

Table 1. Fit Statistics for CFA Models

	χ^2	*df*	CFI	RMSEA	$\Delta\chi^2$
Conscientiousness (C)	451.8***	143	.703	.095	—
Conscientiousness (UC)	441.1***	130	.702	.099	10.7
Persistence (C)	85.1***	47	.921	.057	—
Persistence (UC)	80.5***	40	.916	.065	4.6
Need for cognition (C)	661.1***	288	.749	.074	—
Need for cognition (UC)	616.9***	270	.767	.074	44.2**
Need for structure (C)	260.5***	99	.767	.081	—
Need for structure (UC)	250.1***	88	.766	.086	10.4
Comfort (C)	606.2***	80	.613	.168	—
Comfort (UC)	581.4***	70	.624	.177	24.8

Note. CFA = confirmatory factor analysis; CFI = comparative fit index; RMSEA = root mean square error of approximation; C = factor loadings fully constrained between conditions; UC = factor loadings fully unconstrained between conditions.
***$p < .001$. **$p < .01$.

Mean differences on psychological constructs. One-way ANOVAs were computed to test for mean differences between administration methods on our measures of need for cognition, conscientiousness, persistence, and need for structure. A statistically significant difference in means was only observed for the need for cognition, $F(1, 238) = 8.623$, $p = .004$, $\eta_p^2 = .035$, variable. Participants in the paper-based method ($M = 4.389$, $SD = 0.956$) scored higher on need for cognition than online participants ($M = 4.040$, $SD = 0.886$).

Number and magnitude of correlations. We conducted a series of tests, using Kristopher Preacher's online calculator (Preacher, 2002) to determine whether there might be any significant differences between administration methods in the *correlations* among our need for cognition, conscientiousness, persistence, need for structure, comfort, and assistance measures. All tests of differences in correlations among our measures of these factors were nonsignificant ($ps > .05$), indicating good consistency in the correlations observed between methods.

Exploratory factor analysis: Number of factors. Our next set of analyses was designed to assess whether administration method might lead researchers to retain different numbers of factors during factor analysis. We factor analyzed the 29 items comprising the Need for Cognition and Need for Structure scales in the paper-based and online groups using principal axis factoring. We used the 95th percentile of eigenvalues from 500 randomly generated correlation matrices as our criterion for retaining factors for each of the two samples. In each condition, a four-factor solution was best supported by the data. The factor correlations by condition following Promax rotation are presented in Table 2. Based on these results, it appears that although the absolute numbers of factors identified in the data were equivalent between administration conditions, there were considerable differences observed in the pattern of factor intercorrelations.

Table 2. Factor Intercorrelations Following Promax Rotation

	F1	F2	F3	F4
F1	—	.526	.087	−.145
F2	−.142	—	−.049	−.173
F3	.581	−.258	—	.310
F4	.016	.047	.077	—

Note. Factor correlations for paper-based condition above primary diagonal. Factor correlations for online condition below primary diagonal.

User Response Characteristics

Among user response characteristics, we examined possible differences in user comfort, acquiescence responding, and missing data between the two administration methods.

User comfort. We compared our two groups (paper-based administration vs. online administration) on our measure of comfort with completing the survey. A one-way ANOVA revealed a statistically significant, $F(1, 245) = 6.029$, $p = .015$, $\eta_p^2 = .024$, difference in level of expressed comfort between groups. Participants in the paper-based method ($M = 4.465$, $SD = 1.082$) exhibited higher levels of comfort than those in the online method ($M = 4.104$, $SD = 1.227$).

Acquiescence. To test for differences in the tendency for persons in the two administration methods to demonstrate ''yay-saying,'' we computed an index of acquiescence by summing items (without reverse-coding those that are typically negatively scored) from across our need for cognition, need for structure, conscientiousness, and persistence measures. A one-way ANOVA revealed no significant difference in the tendency to ''yay-say,'' $F(1, 226) = 0.109$, $p = .741$, $\eta_p^2 = .000$.

Differences in Patterns of Missing Data

We created a summative index of the number of missing values (''MISVAL'') occurring in the data for the need for cognition, need for structure, persistence, and conscientiousness measures. Given the strong departure from normality on MISVAL, we related scores on this measure to administration condition via logistic regression. MISVAL was included as the predictor of administration condition (coded 0 = *paper-based*, 1 = *online*). The model containing MISVAL predicted group membership significantly better than a null model with no predictors, Model $\chi^2(1) = 4.08$, $p = .043$. The Nagelkerke R^2 value was .02. The odds ratio of 1.024 indicated that for every one unit increase on the MISVAL variable, the odds of membership in the online group changed slightly by a factor of 1.024. Respondents were correctly classified into either paper-based or online groups 55.1% of the time. In other words, individuals in the online method exhibited a slightly but significantly greater tendency to exhibit missing values.

Measurement Efficiency

We used time-on-task data to compare the efficiency of measurement administration for the two methods. We calculated the amount of time it took (in minutes) for each participant to complete the questionnaires, then compared the two groups by method. On average, participants took significantly longer to complete the paper-based (M = 64.25, SD = 39.51) than the online instruments (M = 47.64, SD = 41.63), $F(1, 227)$ = 9.593, $p < .001$, partial $\eta^2 = .041$.

Qualitative Data Characteristics

To examine the quality of the generative data that participants provided in response to the open-ended items, we used two dimensions, quantity and quality.

Quantity of generative data. We conducted a one-way ANOVA to test the effects of survey administration condition on the quantity of generative data, as number of words provided in the open responses. The word count across the four open response questions was averaged (justified by rs among the four counts ranging from .718 to .909) to provide an average word count for each participant. Administration condition accounted for statistically significant, $F(1, 270) = 6.098$, $p = .014$, $\eta_p^2 = .022$, between-group variation in number of words provided. On average, more words were provided by participants online (M = 1713.207, SD = 3048.649) than on paper (M = 808.316, SD = 2265.304).

Quality of generative data. Our next set of analyses was designed to address the question of whether participants completing the paper or online surveys differed significantly in the *quality* of their essay responses. Two of the researchers independently scored each of the 1,052 open-ended participant responses on three dimensions of quality (completeness, coherence, and correctness) using a standard rubric. For all four essay responses, the interrater reliability was $r = .80$ or greater (ps $< .001$). Moreover, the correlations across raters and across the four essay responses were all high, with the lowest being $r = .682$ ($p < .001$) between essay responses two and three for Rater 1. Given this degree of consistency, we averaged essay ratings across essays and raters to produce a global index of ''Essay Quality.'' The difference in ''Essay Quality'' between paper-based and online administrations was not statistically significant, $F(1, 270) = 2.410$, $p = .122$, partial $\eta^2 = .009$.

Discussion

This research was designed as a systematic investigation of issues surrounding the capture of data across different methods of administration. It was aimed at providing useful information to researchers and practitioners as they make choices about what methods are most appropriate for their work. Moreover, the research was aimed at further expanding on administrative contexts-of-use to include conditions outside of the systems themselves. This study is the first, to our knowledge, to explicitly examine social and asocial conditions as elements of administrative context. Given the

prevalence of systems and practice that enable participants to choose uncontrolled contexts for completion of questionnaire instruments, an examination of their parameters is critical.

Summary and Synthesis of Findings

Participants in both administrative methods chose a range of locations and allowed potential distracters in the contexts-of-use. Most (56%) completed the questionnaires at home (albeit with distracters). Most participants in both groups had friends present while completing the surveys and were engaged in multiple concurrent activities, indicating that users left to their own choice allowed numerous sources of potential interruption and interference.

A smaller but notable percentage (17%) completed them in class (while instructional activities were occurring). This circumstance presents potential disruption of the questionnaire completion and also of attention in class. The higher overall percentage of those in WBA with more social distractions present raises questions about whether participants perceived PBA as a more academic or isolated task and WBA as a more social or public activity.

The finding that few in either group explicitly sought help or changed responses based on input from others indicates intended independence of responding but does not remove the possibility of implicit influences that may contribute to dependent observations. The finding that participants in both groups chose to allow multiple potential distracters underscores the importance of researchers considering the factors and potential effects of external contexts-of-use characteristics for all administrative methods.

As to instrument performance effects of the administrative methods, there were statistically significant differences in the internal consistency and mean difference scores for only one instrument (NFC), with consistency higher and mean scores lower in WBA. Moreover, our multiple group tests indicated that there was some slight measurement noninvariance between administration conditions for the NFC. These instrument performance differences are minimal overall. Nevertheless, we cannot rule out the possibility that particular instrument characteristics might interact with context factors to create noninvariant measurement characteristics in different survey conditions.

The randomly assigned participants reported higher comfort in PBA than in WBA. This group of mostly university undergraduates is certainly made up largely of digital natives, yet they felt more comfortable completing the questionnaires in paper than online. There was a nonsignificant acquiescence difference, and a small but statistically significant tendency to exhibit greater data loss in the WBA condition. Participants took significantly longer to complete the instruments on paper than online. Overall, these findings slightly favor paper-based administrative methods for quantitative instrument performance in terms of data retention and overall user comfort.

For generative data, quantity produced was significantly higher in the WBA method, whereas quality was not significantly different. Thus, the online administration yielded generative data of arguably equal quality in a shorter administrative time. This finding supports the use of digital systems for open-ended response items.

Implications for Measurement Practice

These findings underscore the error in assuming that all participants seek out quiet places and complete questionnaires in solitude. Few measurement professionals would choose to administer questionnaire instruments in a crowded room or allow unrelated social interactions during research activities. Yet we tend to freely send questionnaires out online to reach larger and more distributed population samples without attention to what else is occurring while they complete those instruments.

Our findings do indicate that there are small but significant differences between the two administrative methods when questionnaires are completed in authentic, uncontrolled contexts.

For those conducting research or evaluation, differences suggest the importance of considering how the context in which measures are completed may influence one's results. Measurement professionals need to recognize that the increased portability of technology tools that enable digital access also introduces increased risk of contextual factors that present potential distractions. Recognizing the nature of the broader context-of-use dynamic should help researchers and evaluation practitioners to select administration methods by informed reason rather than by default.

Our findings indicate issues of importance in the contexts and conditions chosen by participants, for uncontrolled studies. These findings can inform judgments about designing in control-of-context based on study purpose and priorities. Where distractions or assistance are acutely important, they may guide specific user instruction, which most online studies currently do not include. These findings also raise questions that suggest additional areas for future research.

Future Research Directions

Findings to date indicate a number of productive directions for this line of research. Given the indications here that particular measures may be more sensitive to contextual influences, comparative investigations across more and diverse instruments should be pursued. These effects for both administration methods and contexts-of-use characteristics can be tested on different types of instruments and extended to performance tasks such as problem-solving and technical simulations. It will also be important to divide even more explicitly among the range of concurrent activities and between social and asocial distractions. In addition, there may be relative controls of context implicit in the fact that these questionnaires were linked to academic tasks (as course credit), which may be diminished or eliminated for instruments

simply received in an email or downloaded from a website. These should be explored as elements of context-of-use for measurement administration as well.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. For all of these measures, we assume that they are capturing unidimensional constructs.
2. Because of some minor differences in data loss between survey conditions, the *N*s per condition for several of our measures were unequal. To facilitate comparisons between groups during our multiple-group confirmatory factor analyses, as well as to test for differences in internal consistency estimates using the bootstrapping option using AMOS™ 17.0, we created equal *N*s groups by randomly deleting cases from the larger groups so that the total number of cases in those groups were equal to those of the smaller groups. The number of cases deleted from the larger groups was generally negligible.

References

Bälter, K. A., Bälter, O., Fondell, E., & Lagerros, Y. T. (2005). Web-based and mailed questionnaires: A comparison of response rates and compliance. *Epidemiology, 16*, 577-579.

Bandilla, W., Bosnjak, M., & Altdorfer, P. (2003). Survey administration effects? A comparison of web-based and traditional written self-administered surveys using the ISSP environment module. *Social Science Computer Review, 21*, 235-243.

Booth-Kewley, S., Edwards, J. E., & Rosenfeld, P. (1992). Impression management, social desirability, and computer administration of attitude questionnaires: Does the computer make a difference? *Journal of Applied Psychology, 77*, 562-566.

Bosnjak, M., Tuten, T. L., & Whitmann, W. W. (2005). Unit (non)response in web-based access panel surveys: An extended planned-behavior approach. *Psychology & Marketing, 22*, 489-505.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*, 116-131.

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48*, 306-307.

Carini, R. M., Hayek, J. C., Kuh, G. D., Kennedy, J. M., & Ouimet, J. A. (2003). College student responses to web and paper surveys: Does mode matter? *Research in Higher Education, 44*, 1-19.

Chua, S. L., Chen, D. T., & Wong, A. F. L. (1999). Computer anxiety and its correlates: A meta-analysis. *Computers in Human Behavior, 15*, 609-623.

Ciolek, T. M. (1998). *The scholarly uses of the Internet: 1998 online survey*. Retrieved from http://www.ciolek.com/PAPERS/InternetSurvey-98.html

Clariana, R., & Wallace, P. (2002). Paper-based versus computer-based assessment: Key factors associated with the test mode effect. *British Journal of Educational Technology, 33*, 593-602.

Cobanoglu, C., & Cobanoglu, N. (2003). The effect of incentives in web surveys: Application and ethical considerations. *International Journal of Market Research, 45*, 475-488.

Comley, P. (1996). *The use of the Internet as a data collection method*. Retrieved from http://www.sga.co.uk/esomar.html

Denscombe, M. (2008). The length of responses to open-ended questions: A comparison of online and paper questionnaires in terms of a mode effect. *Social Science Computer Review, 26*, 359-368.

Dillman, D. A. (2000). Navigating the rapids of change: Some observations on survey methodology in the early 21st century. *Public Opinion Quarterly, 66*, 473-494.

Dillman, D. A., & Bowker, D. K. (2001). The web questionnaire challenge to survey methodologists. In U. Reips & M. Bosnjak (Eds.), *Dimensions in Internet science*. Lengerich, Germany: Pabst Science.

Dodge, M. (2005). The role of maps in virtual research methods. In C. Hine (Ed.), *Virtual methods: Issues in social research on the Internet* (pp. 113-127). New York, NY: Berg.

Dodge, M., & Kitchin, R. (2001). *Mapping cyperspace*. New York, NY: Routledge.

Dolenko, M. (1998). Making user-based assessments work: Evaluation statistic Canada's website. In M. E. Williams (Ed.), *Proceedings of the 1998 National Online Meeting* (pp. 105-114). New York, NY: Information Today.

Fan, X. (2003). Using commonly available software for bootstrapping in both substantive and measurement analysis. *Educational and Psychological Measurement, 63*, 24-50.

Ferriter, M. (1993). Computer-aided interviewing and the psychiatric social history. *Social Work and Social Sciences Review, 4*, 255-263.

Fletcher, L. A., Erickson, D. J., Toomey, T. L., & Wagenaar, A. C. (2003). Handheld computers: A feasible alternative to paper forms for field data collection. *Evaluation Review, 27*, 165-178.

Fox, S., & Schwartz, D. (2002). Social desirability and controllability in computerized and paper-and pencil personality questionnaires. *Computers in Human Behavior, 18*, 389-410.

George, C. E., Lankford, J. S., & Wilson, S. E. (1992). The effects of computerized versus paper-and-pencil administration on measures of negative effect. *Computers in Human Behavior, 8*, 203-209.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Clonninger, C. R., & Gough, H. G. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality, 40*, 84-96.

Goree, D. T., & Marszalek, J. F., III. (1995). Electronic surveys: Ethical issues for researchers. *College Student Affairs Journal, 15*(1), 75-79.

Granello, D. H., & Wheaton, J. E. (2004). Online data collection: Strategies for research. *Journal of Counseling and Development, 82*, 387-393.

Greenlaw, C., & Brown-Welty, S. (2009). A comparison of web-based and paper-based survey methods. *Evaluation Review, 33*, 464-480.

Hardré, P. L, Crowson, H. M., Xie, K., & Ly, C. (2007). Testing differential effects of computer-based, web-based, and paper-based administration of questionnaire research instruments. *British Journal of Educational Technology, 38*, 5-22.

Hardré, P. L., Crowson, H. M., & Xie, K. (2010). Differential effects of web-based and paper-based administration of questionnaire research instruments in authentic contexts-of-use. *Journal of Educational Computing Research, 42*, 103-133.

Hardré, P. L., Nanny, M., Refai, H., Ling, C., & Slater, J. (2010). Re-engineering a dynamic service learning environment for K-12 teachers. *Teacher Education Quarterly, 37*, 157-178.

Hardré, P. L., Xie, K., & Ly, C. (2005). Production and data management issues of digital administration of research questionnaires. *Performance Improvement Journal, 44*(5), 33-39.

Hayslett, M. M., & Wildemuth, B. M. (2004). Pixels or pencils: Relative effectiveness of web-based versus paper surveys. *Library and Information Science Research, 26*, 73-93.

Hine, C. (2005). Research-relationships and online relationships. In C. Hine (Ed.), *Virtual methods: Issues in social research on the Internet* (pp. 17-20). New York, NY: Berg.

Idleman, L. (2003, April). *Comparing responses of mail and web-based surveys*. Paper presented at AERA, Chicago, IL.

Jacob, R. T. (2011). An experiment to test the feasibility and quality of a web-based questionnaire of teachers. *Evaluation Review, 35*, 40-70.

Joinson, A. N. (2005). Internet behavior and the design of virtual methods. In C. Hine (Ed.), *Virtual methods: Issues in social research on the Internet* (pp. 21-34). New York, NY: Berg.

Kiesler, S., & Sproull, L. (1986). Response effects in the electronic survey. *Public Opinion Quarterly, 50*, 243-253.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: The Guilford Press.

Leece, P., Bhandari, M., Sprague, S., Swiontkowski, M. F., Schemitsch, E. H., Tornetta, P., . . . Guyatt, G. (2004). Internet versus mailed questionnaires: A randomized comparison. *Journal of Medical Internet Research, 6*(3), e30.

MacEachren, A. M. (1995). *How maps work: Representation, visualization and design*. New York, NY: Guilford Press.

MacIsaac, D. L. (1999). Comparing web and paper-based administration of standardized introductory physics conceptual and attitudinal instruments. Final Program and Abstracts: 72nd NARST Annual Meeting 1999, 204.

MacIsaac, D. L., Cole, R. P., Cole, D. M., McCullough, L., & Maxka, J. (2002). Standardized testing in physics via the world wide web. *Electronic Journal of Science Education, 6*, 3. Retrieved from http://ejse.southwestern.edu/article/view/7681

MacKay, H. (2005). New connections, familiar settings: Issues in the ethnographic study of new media use at home. In C. Hine (Ed.), *Virtual methods: Issues in social research on the Internet* (pp. 129-140). New York, NY: Berg.

McCabe, S. E. (2004). Comparison of web and mail surveys in collecting illicit drug use data: A randomized experiment. *Journal of Drug Education, 34*(1), 61-72.

Mehta, R., & Sivadas, E. (1995). Direct marketing on the Internet: An empirical assessment of consumer attitudes. *Journal of Direct Marketing, 9*(3), 21-31.

Miller, D., & Slater, D. (2000). *The Internet: An ethnographic approach*. London, England: Berg.

Moon, Y. (2000). Intimate exchanges: Using computers to elicit self-disclosure from consumers. *Journal of Consumer Research, 26*, 323-339.

Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies, 45*, 669-678.

Neuberg, L. S., & Newsom, J. T. (1993). Personal need for structure: Individual differences in the desire for simple structure. *Journal of Personality and Social Psychology, 65*, 113-131.

Nichols, E., & Sedivi, B. (1998). *Economic data collection via the web: A census bureau case study*. Paper presented at the meeting of the American Statistical Association, Dallas, TX.

Noyes, J., & Garland, K. (2008). Computer- vs. paper-based tasks: Are they equivalent? *Ergonomics, 51*, 1352-1375.

Noyes, J., Garland, K., & Robbins, L. (2004). Paper-based versus computer-based assessment: Is workload another test mode effect? *British Journal of Educational Technology, 35*, 111-113.

Porter, S. R., & Whitcomb, M. E. (2003). The impact of contact type on web survey response rates. *Public Opinion Quarterly, 67*, 579-588.

Preacher, K. J. (2002, May). *Calculation for the test of the difference between two independent correlation coefficients* [Computer software]. Retrieved from http://www.quantpsy.org/corrtest/corrtest.htm

Ritter, P., Lorig, K., Laurent, D., & Matthews, K. (2004). Internet versus mailed questionnaires: A randomized comparison. *Journal of Medical Internet Research, 6*(3), e29.

Robinson, R., & West, R. (1992). A comparison of computer and questionnaire methods of history-taking in a genitourinary clinic. *Psychology and Health, 6*(1-2), 77-84.

Sarrazin, M. S. V., Hall, J. A., Richards, C., & Carswell, C. (2002). A comparison of computer-based versus pencil-and-paper assessment of drug use. *Research on Social Work Practice, 12*, 669-683.

Sax, L. J., Gilmartin, S. K., & Bryant, A. N. (2003). Assessing response rates and nonresponse bias in web and paper surveys. *Research in Higher Education, 44*, 409-432.

Schonlan, M., Fricker, R. D., Jr., & Elliott, M. N. (2002). *Conducting research surveys via email and the web*. Santa Monica, CA: Rand.

Schulenberg, S. E., & Yurtzenka, B. A. (1999). The equivalence of computerized and paper-and-pencil psychological instruments: Implications for measures of negative affect. *Behavior Research Methods, Instruments, & Computers, 31*, 315-321.

Schumacker, R.E., & Lomax, R.G. (2004). *A beginner's guide to structural equation modeling* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Sears, G. J., Prakash, S., & Chiocchio, F. (2003). *The world wide web: An added dimension in 360° feedback*? (Report of the Public Service Commission of Canada). Retrieved from http://www.psc-cfp.gc.ca/ppc/world_wide_web_e.htm

Shin, E., Johnson, T. P., & Rao, K. (2011). Survey mode effects on data quality: Comparison of web and mail modes in a US National panel survey. *Social Science Computer Review*. Advance online publication. doi:10.1177/0894439311404508

Spink, A., Bateman, J., & Jansen, B. J. (1998). Users' searching behavior on the Excite web search engine. In M. E. Williams (Ed.), *Proceedings of the 1998 National Online Meeting* (pp. 375-386). New York, NY: Information Today.

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

Tourangeau, R. (2004). Survey research and societal change. *Annual Review of Psychology, 55*, 775-801.

Truman, J., Robinson, K., Evans, A. L., Smith, D., Cunningham, L., Millward, R., & Minnis, H. (2003). The strengths and difficulties questionnaire: A pilot study of a new computer version of the self-report scale. *European Child and Adolescent Psychiatry, 12*, 9-14.

Tseng, H., Tiplady, B., MacLeod, H. A., & Wright, P. (1998). Computer anxiety: A comparison of pen-based personal digital assistants, conventional computer, and paper assessment of mood and performance. *British Journal of Psychology, 89*, 599-610.

Webster, J., & Compeau, D. (1996). Computer-assisted versus paper-and-pencil administration of questionnaires. *Behavior Research Methods, Instruments & Computers, 28*, 567-576.

Whittier, D. K., Seeley, S., & St. Lawrence, J. S. (2004). A comparison of web with paper-based surveys of gay and bisexual men who vacationed in a gay resort community. *AIDS Education and Prevention, 16*, 476-485.

Wilt, J., Condon, D. M., & Revelle, W. (2011). Telemetrics and online data collection: Collecting data at a distance. In T. Little, N. Card, & B. Laursen (Eds.), *Handbook of developmental research methods* (pp. 163-180). New York, NY: Guilford Press.

Wolfe, E. W., Converse, P. D., Airen, O., & Bodenhorn, N. (2009). Unit and item nonresponses and ancillary information in web-and paper-based questionnaire administration to school counselors. *Measurement and Evaluation in Counseling and Development, 42*, 92-103.

Yook, S. H., Jeong, H., & Barabási, A.-L. (2002). Modeling the internet's large-scale topology. *Proceedings of the National Academy of Sciences, 99*, 13382-13386.

Zhang, Y. (1999). Using the Internet for survey research: A case study. *Journal of the American Society for Information Science, 51*, 57-68.

Zutshi, A., Parris, M. A., & Creed, A. (2011). Write it or click on it? Paper vs. online questionnaires for organizational research. *International Journal of Business Innovation and Research, 5*, 663-684.