Electronic Versus Paper Surveys in an Upward Feedback Application: Are the Methods Equivalent?

Michael Yap
Western Kentucky University

Follow this and additional works at: http://digitalcommons.wku.edu/theses
Part of the Industrial and Organizational Psychology Commons

Recommended Citation
http://digitalcommons.wku.edu/theses/629

This Thesis is brought to you for free and open access by TopSCHOLAR®. It has been accepted for inclusion in Masters Theses & Specialist Projects by an authorized administrator of TopSCHOLAR®. For more information, please contact topscholar@wku.edu.
ELECTRONIC VERSUS PAPER SURVEYS IN AN UPWARD FEEDBACK APPLICATION: ARE THE METHODS EQUIVALENT?

A Thesis
Presented to
The Faculty of the Department of Psychology
Western Kentucky University
Bowling Green, Kentucky

In Partial Fulfillment
Of the Requirements for the Degree
Master of Arts

By
Michael K.T. Yap

December, 2001
ELECTRONIC VS. PAPER SURVEYS IN AN UPWARD FEEDBACK APPLICATION: ARE THE FORMATS EQUIVALENT?

Date Recommended 10/25/01

Elizabeth L. Shepherd
Director of Thesis

Reagan Brown

Dean, Graduate Studies and Research  Date

12/10/01
# Table of Contents

Acknowledgements........................................................................................................ ii
Abstract............................................................................................................................. iii
Literature Review ............................................................................................................. 1
Method.............................................................................................................................. 9
Results............................................................................................................................... 13
Discussion.......................................................................................................................... 18
References......................................................................................................................... 22
Appendix............................................................................................................................. 25
Acknowledgements

My heartfelt thanks and appreciation go to all those who helped me in this endeavor. Elizabeth Shoenfelt, my advisor and thesis chair, kept me moving with equal measures of gentle prompting and well-placed, well-timed boots along the way. Alan Walker, my friend and colleague, provided key technical advice, editorial expertise, and moral support throughout the process. James W. Smither also provided editorial review. Most instrumental has been Kathrin Yap, my wife, confidant, and best friend. Her edits and suggestions have made this paper much easier to read. Her ongoing moral and spiritual support over the last decade has made me a better person in so many ways.
ELECTRONIC VERSUS PAPER SURVEYS IN AN UPWARD FEEDBACK APPLICATION: ARE THE FORMATS EQUIVALENT?

Michael Kim T. Yap

Directed by: Elizabeth Shoenfelt, Reagan Brown, Jaqueline Pope

Department of Psychology

Western Kentucky University

This thesis represents the first known study to investigate the equivalency of paper vs. electronic survey data collection formats in an upward feedback application. Findings are similar to recent research utilizing employee opinion survey data. Format explained less than 1% of variance in managers’ total scores over that of ratee and rater demographic variables.
Electronic versus Paper Surveys In an Upward Feedback Application:

Are the Formats Equivalent?

Computers have become a daily business tool for many jobs. Widespread computer usage, coupled with employees' growing familiarity with the Internet and World Wide Web, has helped to make electronically delivered employee opinion surveys (e.g., computer, internet and intranet-based) more convenient, more practical and more widely used.

Electronic surveys have a number of advantages over traditional paper-and-pencil surveys. For example, organizations may realize a cost savings in materials and human resources (Kraut, 1999; Sproull, 1986; Weible & Wallace, 1998; Young et al., 2000). Cost savings are especially likely for companies that have a robust technological infrastructure and the internal expertise to develop electronic surveys. Once the survey is developed, the electronic format can be significantly faster to implement since mailing time is eliminated. Data entry time can also be greatly reduced if the information can be automatically added to a database (Bachmann, Elfrink, & Vazzana, 1996; Schaefer & Dillman, 1998). Additionally, in comparison to the paper-and-pencil format, the electronic format has meant fewer errors and less missing data (Kiesler & Sproul, 1986; Stanton, 1998). Finally, some employees prefer the electronic format, especially those who work in high-tech organizations (Church, in press; Rosenfeld, Doherty, Vicino, Kantor, & Greaves, 1989). Employees who work in these organizations may have greater comfort and proficiency with the electronic medium.

There are also some disadvantages to electronic surveys. These include concerns with confidentiality, data accuracy, and problems with the technology (Kraut, 1999). For
example, organizations that have not progressed nor kept up-to-date technologically or whose core business, mission, or strategy do not require sophisticated technical infrastructure may lack employee confidence in an electronic data collection format. Concerns with rater anonymity may lead raters to be less honest or forthright with their evaluations. Furthermore, without a strong technological infrastructure in place, employees may have concerns (possibly legitimate) with issues of accurate data collection; that is, they may doubt the integrity of the system to correctly collect, tabulate, or assign their data properly. Finally, if the electronic system does not work properly, is slow, or not user-friendly, employees may become alienated with the process, leading to the survey program’s demise.

These concerns and possible negative outcomes, coupled with the increasing number of companies that are making the transition from paper-and-pencil to electronic format, raise important questions about the equivalence of scores between the two methods. For example, equivalency is important when organizations use survey data on subsequent administrations to make evaluative judgments as to whether they have made improvements from year to year; that is, organizations often compare one year’s data to the next to see if improvement in scores has been realized. Unless equivalency of electronic and paper formats is established, using paper-and-pencil data collection formats one year and an electronic format the next year make such evaluations problematic. Specifically, it may be difficult to determine if changes in scores were attributable to a real improvement or simply due to the change in the data collection format.

Another equivalency issue is faced when an organization simultaneously employs paper-and-pencil and electronic formats. Such use of dual methods frequently occurs
when an organization is in the process of evolving away from a paper-and-pencil format toward an electronic format. The organization may initially use both paper-and-pencil and electronic formats (Magnan, Lundby, & Fenlason, 2000). Other organizations may simultaneously employ both a paper-and-pencil and electronic format for the simple reason that only a portion of the workforce may have access to the Internet or the company’s intranet. Another reason companies use dual methods simultaneously is that only a portion of their employees have sufficient skill or comfort level to complete a computer-based survey. Hardware and software differences may also make it difficult to use a solely electronic format. For example, an employee’s personal computer software, configuration or set-up to the company’s intranet or the Internet may be incompatible with the data collection software.

Even though establishing the measurement equivalence of paper-and-pencil and electronic surveys is important and many organizations are currently employing both formats, relatively little research has been done on this topic. The research that has been done has focused on only one previously mentioned major equivalency concern: the degree of anonymity that respondents feel. For example, in a study conducted by Rosenfeld at al. (1996) respondents completed surveys using one of three methods: paper-and-pencil, computerized, or computerized with answers linked to other databases within the organization. Respondents were also either identified or anonymous. Only the respondents who were both identified and believed that their answers would be linked to other databases scored higher in impression management. The researchers concluded that paper-and-pencil surveys are equal to electronic surveys when participants believed they were anonymous. Similarly, Booth-Kewley, Rosenfeld, and Edwards (1993) found that
US Navy recruits responded with more social desirability when they were identified. When they remained anonymous, recruits responded similarly to computer-based and paper-and-pencil survey. Finally, Lautenschlager and Flaherty (1990) found that students displayed more social desirability when they believed their responses could be linked to their identity. The researchers surveyed students using three methods: individual computer, individual paper-and-pencil, and group paper and pencil. In addition, half the students were asked to include their name on the questionnaire, while the other half remained anonymous. The researchers found that both the identified respondents and the computer-based respondents displayed increased social desirability. These studies provide strong evidence of the role rater anonymity may play in impacting survey responses and may indicate one possible source of response differences between methods.

Booth-Kewley et al. (1993) reviewed the available research on differences between paper-and-pencil and electronic surveys. They found that results should be similar, as long as similar response formats were used. Similarly, King and Miles (1995) found that the two methods were equivalent, although the paper-and-pencil versions produced more socially desirable responding. Stanton (1998) also determined the two methods to be equivalent. Church (in press) completed a study in which he examined method effects for two administrations in a single organization. He concluded that method effects accounted for only a very small percentage of unique variance above that explained by demographic variables. Finally, Young et al. (2000) found small and inconsistent differences in item means and standard deviations between paper-and-pencil and web-based surveys. On the whole, the above research seems to indicate that paper-and-pencil and electronic data collection formats will likely lead to very similar results, especially if raters believe
their anonymity is preserved. However, some have leveled criticism at this research for several reasons. For example, Young et al. (2000) noted that one limitation of previous research is that it was conducted prior to wide acceptance and utilization of the World Wide Web or corporate intranets. Thus, the research emphasized disk or mainframe-based surveys. Although there are similarities between these methods and Web-based surveys, some important differences exist. For example, unlike disk or mainframe processes, Web-based (to include intranet-based) formats allow the participant to access the survey from a variety of locations (including a participant’s personal office space or even from home) and do not require the rater to complete the survey in a predetermined location. Young et al. (2000) also noted that the technology utilized in many earlier studies allowed the participant to view only one question at a time and did not allow editing or reviewing of previous responses. In contrast, most of the Web-based survey technologies utilized today allow the rater to review and revise earlier responses.

Others (e.g., King & Miles, 1995) have cautioned that researchers and practitioners should not assume measurement equivalence based solely on a lack of mean differences in responses, even if survey items are identical between the two formats. Rather, they suggest that data should be considered equivalent only if they have the same underlying content domains as determined through similar outcomes on a principle components analysis of the two databases. In a similar vein, Young et al. (2000) suggested that one weakness in past employee survey research is that most authors focus on mean differences in scale scores rather than on the response chosen. Thus, extreme responding in one area of a survey could result in fewer neutral responses, yet could still produce the same mean score. For example, paper-and-pencil survey respondents may be
more likely to use the extreme points of the response scale because they believe their ratings are completely anonymous. In contrast, electronic format raters may tend to use the midpoint of the scale (and be less willing to offer extreme scores) due to a concern for their anonymity. In this situation just described, the mean scores may be equal, but only when the response distributions of the two methods were examined would this situation be detected.

In spite of the cautions that have been raised in using mean scores, organizations continue to widely utilize them as the sole criteria when evaluating equivalence between data collection formats. Additionally, many organizations use mean difference scores for other purposes as well -- for example, evaluating whether efforts aimed at improving employee attitudes have been successful as measured by differences in mean scores between survey administrations. Thus, many organizations may be erroneously concluding that their efforts were successful (or unsuccessful) since difference in means scores may be due in large part to differences in data collection formats rather than any initiatives that the company may have implemented.

Of even greater concern is that many organizations utilize organizational survey data to make individual administrative decisions such as who to outplace, promote, or select into "high potential" status (London & Smither, 1995). Such use of data is especially likely when the survey program is an upward feedback or multisource feedback program. In fact, Dalton (1996) reported that an increasing number of companies are moving toward an administrative use of multisource feedback. Thus, issues of data collection format equivalence may be especially salient in upward feedback or multisource feedback applications. In this context, Bracken and Timmreck (2000) have cautioned that
some organizational climates may show resistance to certain technologies. They advocate organizations that employ multiple technologies to collect upward feedback data systematically examine whether any possible method bias may have been introduced. They suggest examining issues such as lower or higher scores, lower response rates, incomplete questionnaires, or errors in responding (such as a rater marking all items in the exact same way). Obviously, if systematic error has been introduced into the ratings due to employing multiple technologies (or moving a program from one technology to another), the validity of the program may be compromised. Of particular concern would be a situation in which such a change introduces adverse impact or even differential validity for certain groups of employees.

In spite of these issues, to date, no research has empirically investigated measurement equivalency in an upward feedback or multisource feedback application. The lack of research is especially disconcerting given upward feedback and multisource feedback instruments are held to the same legal standards as any other organizational decision-making tool such as selection tests or performance evaluations (Bracken, 1994)

The present study was designed to add to the existing knowledge of survey format equivalency. Specifically, this study was undertaken to fill the void of empirical evidence regarding the equivalency of data collection formats when used in an upward feedback application. Unlike most of the earlier researchers, the current investigator examined format equivalency utilizing participants from a single organization. By doing so, the researcher minimized confounding variables that might have otherwise been introduced (e.g., differences in survey administration other than format, corporate culture, technological infrastructure).
The organization involved in this study uses the data collected in its upward feedback program for both developmental and administrative purposes. As such, instances arise in which leaders’ scores are used as a determining factor when personnel decisions are made. For example, the data are used in making designations to the company’s high-potential list and are used as an element in senior leaders’ annual personal bonus plans. Such decisions affect both the careers and the lives of the parties involved. Although the equivalency of survey methods is assumed, it has not been empirically investigated. The present work will test this assumption.

The study was conducted in a large regional bank. Approximately half the raters evaluated their leader using a paper-and-pencil format, while the other half used an electronic format. Based on the review of the literature, the following hypotheses were proposed and tested. First, based on employee opinion survey research that on the whole demonstrates equivalence across data collection formats, the following hypothesis was offered:

**Hypothesis 1a**: No significant difference will be observed between the overall means for those who responded via paper-and-pencil vs. electronic formats.

As has been suggested by others (e.g., Young et al., 2000), method equivalence should not be investigated by testing only for mean differences. Instead, they have suggested that the response distributions on the scale be compared and tested for equivalence. Therefore, the following hypothesis was offered:

**Hypothesis 1b**: The response distributions across the rating scales will not differ significantly by response format.
Similarly, others (e.g., King & Miles, 1995) have suggested that the underlying content domains between the two formats should be examined. Therefore the following hypothesis was proposed:

**Hypothesis 1c:** The underlying content domains will not differ as examined by a separate principle components analyses for the two response formats.

Based on the findings of Kiesler and Sproul (1986) and Stanton (1998), who found fewer errors and less missing data in systems that utilized an electronic format, the following hypotheses were offered:

**Hypothesis 2a:** Fewer invalid response patterns (i.e., all maximum or all minimum scores) will be observed in the electronic format than in the paper-and-pencil format.

**Hypothesis 2b:** The electronic format will yield a better overall response rate (i.e., fewer items left blank) than will the paper-and-pencil format.

**Method**

**Overview**

This study was conducted in an organization that operates in a mature, highly competitive business environment. Competitors are able to easily match products, services and pricing. In this business environment, the organization considers its leaders to be its true competitive advantage. As a result, corporate strategy emphasizes developing strong leaders. The data collected in this study were a part of a major ongoing organizational development effort designed to identify, measure, and develop leadership potential within the organization. The program was originally designed to be developmental. However, as mentioned, the organization has subsequently used the
results to make administrative decisions. Accordingly, a copy of the target manager’s feedback report is provided to his or her direct supervisor.

**Survey Development**

The survey instrument was developed based on a series of employee and management focus groups designed to identify behaviors believed to be associated with effective leadership, productivity, and implementation of strategic business objectives. The instrument consists of 31 behaviorally based items (see Appendix) and incorporates a 5-point Likert-type scale. Responses range from 1 = strongly agree to 5 = strongly disagree. For this survey, it is important to note that the lower the score the better. Such “reverse scoring” is common for surveys conducted within institutions of this type. A sixth response point “Too new to rate” was added to the scale due to managers’ concern that new employees may provide a rating of 3 (sometimes agree, sometimes disagree) when they were lacking insufficient information. In these analyses, this response will be treated as missing data. Yearly principle components analyses with a varimax rotation consistently revealed a single component (using a decision rule of Eigenvalues greater than 1, item loadings .40 or higher on only one component, and scree plot examinations) accounting for over 60% of the variance. The single component is interpreted as an overall leadership component.

When referring to electronic surveys, Young, Daum, Robie and Macie (2000) noted that there is a technical difference between the terms “computer-based” and “Web-based” surveys. Computer-based questionnaires are typically administered from a stand-alone computer or e-mailed. Web-based surveys require the computer to be linked to
either the Internet or a company’s intranet. The present study made use of the corporate intranet as the delivery channel for the electronic survey.

Participants

The survey was administered within a major division of the organization. This division employs approximately 6,700 individuals, mostly within a single southeast state. In 1999, 5,438 employees (raters) were involved in the study, each of whom was part of a work group of three or more (in order to protect rater confidentiality). Of these 2,629 (48%) employees rated their manager using the on-line format, while 2,809 (52%) utilized the paper format. A number of demographic variables were captured in the survey process. These demographic items were presented at the end of an employee opinion survey that was administered simultaneously with the upward feedback survey. Instructions informed raters that providing demographic information was entirely voluntary and would be used only to determine if differences exist between large groups of individuals (such as region). The average age of raters (coded 1 = less than 20, 2 = 20-25, 3 = 26-35, 4 = 36-50, and 5 = 50+) was 3.44, or about 37. Raters had, on average, less than 3 yrs of job tenure (mean = 3.88; coded 1 = less than 6 months, 2 = 6 months to 1 year, 3 = more than 1 year, but less than 3 years, 4 = more than 3 years, but less than 5 years, 5 = more than 5 years, but less than 10 years, and 6 = more than 10 years).

Procedure

The survey has been administered annually in this organization since 1991. Up to 1998 all data had been collected via “scan-able” forms. As mentioned, for the first time in 1999 the organization’s technical capabilities made it possible to collect data over the company’s intranet. As such, all managers with at least three direct reports were
contacted by phone to determine the type of administration they wanted to utilize: either a group survey administration that was facilitated by a representative of the human resources department or electronic surveys via the corporate intranet. Manager’s choice could be influenced by a variety of factors, including whether employees had intranet access available at their workstation; if not, whether business need would allow employees to travel to an electronic survey facility; convenience; or other preference of one method over another. Work units located away from major centers or in rural areas were visited by personnel representatives who administered the survey in person if the paper-and-pencil format was used. Employees were able to select another survey administration other than the one scheduled by his or her manager to accommodate business or personal schedules.

Although this type of administration is quite time-consuming, it does have the advantage of substantially increasing overall participation and response rates. In fact, due to administrative thoroughness, the organization achieved near 100% participation rates for managers and response rates of over 90% for subordinates for the previous five-year period.

Surveys were returned to the human resources department for tabulation and report generation. All managers involved with the program received a 5-page report that summarized their individual data and presented comparisons with larger groups within the organization. Data were presented to managers in terms of item means, standard deviations, range (minimum and maximum score), and valid N aggregated across subordinates. A leadership “total score” was also computed (based on the mean across the means of the individual items) and presented to managers. Managers were also provided with detailed, group-facilitated instruction on interpreting their data, facilitating a
feedback session with their employees, and developing action plans. This training and orientation are mandatory for all first-time participants.

Results

Coefficient Alpha for the 31-item instrument was found to be .98. As can be seen in the Appendix, mean scores and standard deviations between the two data collection formats were somewhat dissimilar. Specifically, the mean score for the paper-and-pencil format was 2.04, while the mean score for the on-line format was 1.74. Similarly, the mean standard deviation for the paper-and-pencil format was 1.00, whereas the mean standard deviation for the on-line format was .85.

To evaluate Hypothesis 1a, an independent groups t-test was conducted to compare the mean differences between the two data collection formats. Contrary to expectations, results revealed that mean differences between the paper and on-line version were significant (t = -14.4, p < .01). Examination of the mean scores revealed that, on average, those employees who rated their manager using the on-line data collection format (M = 1.74, SD = .85) rated their leader significantly more favorably than did those who rated their manager via the paper-and-pencil format (M = 2.04, SD = 1.00). Of course, factors other than response format may influence differences in mean scores between the groups. For example, it could be that employees in lower levels of the organization may rate their managers more harshly and at the same time be less likely to utilize the on-line format (Kraut, 2001). Similarly, employee age and/or tenure and especially organizational level could be related to both one’s willingness (or comfort level or ability) to complete the survey on-line and the favorability of his/her ratings.
In order to further investigate mean differences and to control for these possible confounding variables, a hierarchical regression model was used where individual raters’ Total Scores (computed by averaging across the 31 items for each individual) served as the dependent variable. On the first step a block of ratee (i.e., target or focal manager) demographic variables was entered. These variables included managers’ work unit size (determined by the number of individual subordinates the manager had rate them), most recent annual review ratings, and previous year’s upward feedback ratings. On step 2 a block of rater demographic variables was entered, including the raters tenure, organizational level and age. Organizational level was determined by the focal managers’ (i.e., ratees’) salary grade. On the third step the data collection format (i.e., Method) was entered. Thus, one goal was to determine whether or not Method could explain unique variance above and beyond that of ratee and rater demographic variables. On steps four through nine the interaction terms were entered between each demographic variable and Method. Table 1 presents the findings of this analysis. The F-test associated with the change in $R^2$ was significant after adding Method. However, the change in $R^2$ was only .005, indicating that Method explains less than one percent of the variance in managers’ Total Scores above that of rater and ratee demographic variable. Interestingly, it appears that ratee demographic variables were more salient in predicting managers’ Total Scores than were either rater demographics or data collection format.
Table 1

Hierarchical Regression Analysis for Data Collection Format Equivalency
(Paper-and-Pencil vs. On-line)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Increase in $R^2$</th>
<th>F Change</th>
<th>Sig. of F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ratee Demographic Block</td>
<td>.122</td>
<td>148.6</td>
<td>.000</td>
</tr>
<tr>
<td>Step 2&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rater Demographic Block</td>
<td>.001</td>
<td>1.65</td>
<td>.18</td>
</tr>
<tr>
<td>Step 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Method</td>
<td>.005</td>
<td>18.50</td>
<td>.000</td>
</tr>
</tbody>
</table>

Note. Dependent variable is centered 1999 Leadership Survey total score.
<sup>a</sup>Step 1, ratee demographic block, includes group size, last annual review score, and the previous year’s (1998) upward feedback score.
<sup>b</sup>Step 2, rater demographic block, includes tenure, age, and level (manager’s salary grade used as proxy).

In order to test Hypothesis 1b, a Chi-Square test of association was conducted which examined whether the response distributions differed across the 5-point scale between the two data collection formats. Contrary to expectations, results indicated that significant differences did indeed exist ($X^2 = 5,432; df = 4; p<.01$). In particular, those who rated their manager utilizing the paper version were much more likely to utilize a “3,” “4,” or “5” rating and much less likely to utilize a “1” rating (see Table 2). Interestingly, raters in both data collection formats were equally likely to utilize a “2” rating. Thus hypothesis 1b was not supported.
Table 1

Contingency Table for Chi-Square Test of Hypothesis 1b

<table>
<thead>
<tr>
<th></th>
<th>1 Strongly agree</th>
<th>2 Agree</th>
<th>3 Sometimes agree/Sometimes disagree</th>
<th>4 Disagree</th>
<th>5 Strongly disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>On-line</td>
<td>36,241</td>
<td>29,358</td>
<td>9,162</td>
<td>2,261</td>
<td>831</td>
</tr>
<tr>
<td>Paper</td>
<td>28,460</td>
<td>32,180</td>
<td>14,445</td>
<td>9,788</td>
<td>2,503</td>
</tr>
</tbody>
</table>

In order to test Hypothesis 1c and to address the issue raised by King and Miles (1995), that different data collection formats should have similar underlying content domains (in addition to having similar mean scores), separate principal components analyses were conducted on both the paper-and-pencil and electronic data capture formats. Consistent with expectations, results from these analyses revealed one underlying component for both the on-line and the paper-and-pencil data collection format. Specifically, for the paper format, results revealed one component with an Eigenvalue greater than 1. This component had an Eigenvalue of 20.6 that explained 67% of the variance. Similarly, for the on-line format, results revealed a single component accounting for 67% of the variance. Thus, Hypothesis 1c was confirmed.

Hypothesis 2a predicted that there would be significantly fewer invalid response patterns for the on-line format (i.e., all “1’s” or all “5’s”). Contrary to expectations, the on-line format had more invalid response patterns than did the paper format. Specifically, 325 raters rated their manager all “1’s” while 222 raters responded with this invalid pattern for the paper format (see Table 3). Very few raters rated their manager all “5’s” (only 4 for the paper format and 1 for the on-line version).
In order to test Hypothesis 2b, a Chi-Square test was conducted which compared the equivalence of the frequency of missing data for the two formats. Contrary to expectations, results revealed a nonsignificant difference in missing data frequencies between the two methods ($X^2 = .27; df = 1; p > .05$). Contrary to expectations, the online format did not result in significantly less missing data (see Table 4).

Table 4

| Contingency Table for Chi-Square Test of Hypothesis 2b |
|-----------------|-----------------|
|                  | Left Blank      | Not Left Blank |
| On-line          | 843             | 80,656         |
| Paper            | 877             | 86,202         |

In order to further investigate this finding, an item completion rate was computed using a method similar to that of Stanton (1998) and Church (in press). Specifically, the total number of blank, missing or skipped responses at the individual level were summed. An independent-groups t-test was conducted to determine if there were significant
differences in the average number of items left blank between the two formats. Results revealed no significant differences ($t = .24, p > .05$). On average, on-line raters left .32 items blank while paper respondents left .31 items blank. Again, contrary to hypothesis 2b, there was not less missing data for the on-line format. Finally, the average number of missing demographic variables between the two formats using the item completion rate described above (i.e., simply summing the total number of blank, missing or skipped responses at the individual level) was investigated. An independent-groups t-test revealed significant differences between the average number of demographic items left blank between the two methods ($t = -2.26, p < .05$). However, an examination of the means revealed that leaving demographic variables blank was not an issue for either method as on-line raters left only .63 demographic items blank while paper-and-pencil respondents left only .54 items blank on average. However, this finding does point to the possibility that those who utilized the on-line format may have had less confidence in the anonymity of the on-line system.

Discussion

Overall, the results of this study suggest equivalency between paper-and-pencil and electronic data collection formats in this upward feedback application. While the response distributions for the two formats across the 5-point scale varied significantly and the on-line version exhibited more not fewer invalid response patterns, data collection format explained less than one percent of unique variance above what could be explained by rater (and especially) ratee demographic variables. Further, the underlying content domains were found to be the same between the two formats.
This finding should be welcome news to practitioners utilizing multiple data collection formats in an upward or multisource feedback application. Practitioners in organizations that utilize such ratings in administrative versus “purely developmental” approaches may be especially pleased with the finding. As Bracken and Timmreck (2000) pointed out, biases or systematic error introduced by utilizing multiple data collection formats may impact the validity of the program.

Additionally, several characteristics of the design of the upward feedback process described in this study may add to the comfort level of practitioners. For example, gaining intranet access required users to logon with their personal id and password, yet this requirement did not appear to result in users’ heightened concern for their own anonymity. Second, participants who completed the electronic survey were also asked to type their name on an “electronic sign-in sheet.” Although the clearly stated purpose was to track participation while keeping responses anonymous, it was an additional instance in which identifying information had to be given. Yet, participants’ concern for anonymity did not appear to heighten. Although participants who completed the paper-and-pencil survey were also asked to write their name on a separate sign-in sheet, there was a visual separation between their name and their responses (i.e., sign-in sheets went in one envelope, surveys went in another envelope). Such a visual separation did not exist for the electronic format. Further, the upward feedback program was designed, implemented and analyzed, and reported in-house. The fact that corporate representatives handled such sensitive feedback did not appear to concern participants about anonymity to the degree that it affected their subsequent ratings. Finally, the present organization is clearly several
years behind the state-of-the-art in technological sophistication. Given these program
design features, the results of equivalency between the formats are even more striking.

Study Limitations

One limitation of the present research is that it is not an experimental study.
Specifically, participants were not randomly assigned to either take the survey by paper-
and-pencil or electronically. Although the group sizes were nearly equal, managers made
arrangements for their work groups to participate. As a result, managers had an influence
on whether their employees utilized the paper-and-pencil or the web-based format. Even
though managers' most recent annual review, previous upward feedback scores, and work
unit size were all held constant statistically, it could be that other unmeasured variables
were associated with both managers' propensity to schedule their employees into one or
the other formats and their scores (e.g., comfort with technology, availability of
technology for their work areas, etc.). Future studies utilizing true random selection to
either on-line or paper format conditions, as well as measures of employees' trust in the
system, is needed in order to more fully understand the underlying dynamics at play here.

Future Research

One variable that future research should control is the level of technological skill
that exists between the large employee subgroups. For example, many employees of the
company in the present study were in production-oriented jobs that require little if any
operation of personal computers. Even for those employees who use computer
technology on the job, much of the work is accomplished via a "dumb terminal" rather
than a true personal computer (PC). Similarly, many employees do not have a computer
at home and may have been intimidated by or at least unfamiliar with even basic computer
operation skills such as using a mouse. However, other employees use PCs daily to complete their work and were likely proficient users of computer technology. Since the groups were not truly assigned to experimental conditions at random, employees may have self-selected into one group or another based on comfort with the different methods or some other differentiating factor. Perhaps better scoring managers were more invested in developing and utilizing their employees’ technological skills. Thus, perhaps more highly developed employees self-selected into the electronic survey group. On balance, the results of this study support the notion that paper-and-pencil and electronic data collection formats were equivalent in this upward feedback application.
References

Bachman, D., Elfrink, J., & Vazzana, G. (1996). Tracking the progress of e-mail vs. snail-mail. Marketing Research, Summer, 30-36.


Appendix

Item Descriptive Statistics for Paper vs. On-line Formats

<table>
<thead>
<tr>
<th>Item</th>
<th>Paper Sample(^a)</th>
<th>On-line Sample(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>1. My manager shares with me the information I need to do my job.</td>
<td>1.90</td>
<td>.90</td>
</tr>
<tr>
<td>2. When I need it, my manager provides information about how I'm performing my job.</td>
<td>1.93</td>
<td>.91</td>
</tr>
<tr>
<td>3. My manager empowers me to create value and build loyalty for my customers.</td>
<td>1.94</td>
<td>.92</td>
</tr>
<tr>
<td>4. My manager promotes teamwork within our work unit.</td>
<td>2.01</td>
<td>1.05</td>
</tr>
<tr>
<td>5. My manager promotes teamwork between people in our work unit and people in other work units including those company-wide.</td>
<td>2.13</td>
<td>1.02</td>
</tr>
<tr>
<td>6. My manager listens to my suggestions.</td>
<td>1.98</td>
<td>.97</td>
</tr>
<tr>
<td>7. My manager keeps me informed on what the company is trying to accomplish.</td>
<td>2.12</td>
<td>1.00</td>
</tr>
<tr>
<td>8. My manager keeps me informed on what our work unit is trying to accomplish.</td>
<td>1.98</td>
<td>.94</td>
</tr>
<tr>
<td>9. My manager involves our work unit in continuously improving the way we service our customers.</td>
<td>2.06</td>
<td>.94</td>
</tr>
<tr>
<td>10. My manager encourages me to develop myself.</td>
<td>2.09</td>
<td>1.05</td>
</tr>
<tr>
<td>11. My manager makes sure I am trained to do my job.</td>
<td>2.05</td>
<td>1.00</td>
</tr>
<tr>
<td>12. My manager treats me with respect.</td>
<td>1.78</td>
<td>.97</td>
</tr>
<tr>
<td>13. My manager supports my career development even if it means my moving to another area of the company.</td>
<td>2.04</td>
<td>1.01</td>
</tr>
<tr>
<td>Item</td>
<td>Paper Sample&lt;sup&gt;a&lt;/sup&gt;</td>
<td>On-line Sample&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>---------------------------------------------------------------------</td>
<td>--------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>14. My manager helps me understand what our corporate culture is and means in my day-to-day activities.</td>
<td>2.21</td>
<td>1.05</td>
</tr>
<tr>
<td>15. My manager works with me to ensure I understand the standards/goals on which my performance review will be based.</td>
<td>2.07</td>
<td>1.04</td>
</tr>
<tr>
<td>16. My manager is accessible for discussions.</td>
<td>1.91</td>
<td>.97</td>
</tr>
<tr>
<td>17. My manager and I have discussed the knowledge, skills, and abilities that could affect my progress with the company.</td>
<td>2.24</td>
<td>1.08</td>
</tr>
<tr>
<td>18. I have confidence in the fairness of my manager.</td>
<td>2.08</td>
<td>1.14</td>
</tr>
<tr>
<td>19. My manager makes sure that I present my views during my performance reviews.</td>
<td>1.87</td>
<td>.91</td>
</tr>
<tr>
<td>20. My manager helps me understand how my job contributes to the company’s success.</td>
<td>2.07</td>
<td>.97</td>
</tr>
<tr>
<td>21. My manager makes sure that I get the recognition for my performance.</td>
<td>2.18</td>
<td>1.09</td>
</tr>
<tr>
<td>22. My manager actively promotes and models ethical business practices.</td>
<td>1.93</td>
<td>.96</td>
</tr>
<tr>
<td>23. My manager has a good understanding of what motivates me to do my best.</td>
<td>2.27</td>
<td>1.12</td>
</tr>
<tr>
<td>24. If I thought I needed to go out on a limb to deliver excellent service, I am confident my manager would support me.</td>
<td>1.87</td>
<td>.99</td>
</tr>
<tr>
<td>25. My manager works with me to help resolve conflicts between work and family/personal issues.</td>
<td>1.90</td>
<td>1.01</td>
</tr>
<tr>
<td>26. My manager coaches me to meet the challenges of my job.</td>
<td>2.16</td>
<td>1.01</td>
</tr>
<tr>
<td>27. My manager provides encouragement in a way that is meaningful to me.</td>
<td>2.20</td>
<td>1.01</td>
</tr>
<tr>
<td>Item</td>
<td>Paper Sample&lt;sup&gt;a&lt;/sup&gt;</td>
<td>On-line Sample&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>---------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>28. My manager routinely recognizes or praises my good work.</td>
<td>2.25 1.13</td>
<td>1.85 .94</td>
</tr>
<tr>
<td>29. My manager provides me with the tools to create value and build loyalty for our customers.</td>
<td>2.12 .99</td>
<td>1.75 .77</td>
</tr>
<tr>
<td>30. My manager does a good job of integrating new employees into our work unit (e.g., by providing job description and expectations of work performance, necessary supplies and equipment, celebrating their arrival, etc).</td>
<td>2.15 1.05</td>
<td>1.82 .87</td>
</tr>
</tbody>
</table>

Rating Mean

2.04 1.00 1.74 .85

<sup>Note</sup>. Dependent variable is centered 1999 Leadership Survey total score
<sup>a</sup>Paper Sample Alpha = .98.
<sup>b</sup>On-line Sample Alpha = .99.