# Perceptions of Confidentiality in Survey Research: Development of a Scale

Thomas J. Whelan
Training Industry, Inc.

J. William Stoughton
SHL Talent Measurement by CEB

Lori Foster Thompson
North Carolina State University

## Abstract

As methods of survey administration have evolved over time, confidentiality perceptions and their potential impact on data collection have remained an issue for researchers and practitioners alike. This study develops and provides validity evidence for a measure of survey confidentiality perceptions in a sample of 812 respondents. The final 10-item measure is shown to demonstrate unidimensionality, as well as divergent validity evidence that perceived confidentiality is a distinct construct from anonymity perceptions. The resulting scale can be used in organizational survey research to provide deeper insight into the social and cognitive processes underlying privacy and response behavior.

For organizational researchers, many constructs of interest to practitioners and scientists require the use of self-report surveys to collect data. According to a sampling of nearly 600 published studies by Sackett and Larson (1990), roughly half included some type of self-report measure. Although self-report measures are often, "arguably the best method for assessing the variables of interest… at other times, however, they are used simply as a matter of convenience" (Sackett & Larson, 1990, p. 441). Thus, research on issues associated with self-report data is vital to improving data collection techniques, as the quality of applied research ostensibly depends on the soundness of the methods used to conduct the research.

Survey privacy has been a topic of empirical inquiry in the organizational literature that has spanned decades of research (e.g., Dunnette & Heneman, 1957;

Whelan, T. J., Stoughton, J. W., & Thompson, L. F. (2015, April). *Perceptions of confidentiality in survey research: Development of a scale*. Poster presented at the 30th annual SIOP meeting, Philadelphia, PA

Booth-Kewley, Larson, & Miyoshi, 2007). The underlying premise of how anonymity affects respondent behavior revolves around the notion that when a respondent is anonymous, they will be more forthcoming in their responses to questions asked of them on a survey instrument. Confidentiality is thought to affect respondent behavior by providing protection from any consequence by maintenance of response integrity, and therefore respondents can answer items with impunity even though their identity is known. However, while these statements have long been accepted as fundamental axioms of survey research, researchers continue to investigate how anonymity or the lack thereof, can influence the candor with which respondents answer items.

A relatively unexplored domain in survey research is in the subjective privacy perceptions held by respondents. The openness with which respondents complete surveys is vital if the results of the survey are to be useful, particularly if the survey content contains any sensitive or potentially embarrassing questions (Feigelson & Dwight, 2000). Organizational members are often promised to be safe from managerial retaliation if they reveal sensitive information, in order to encourage honest responses. However, privacy assurances can only have the desired effect on truthfulness if respondents believe the assurances are true. The purpose of this study is to develop a scale that can be utilized to measure survey respondents' perceptions of confidentiality when completing an organizational survey, thereby giving researchers a tool to measure these perceptions and investigate their potential impact on the data quality of a survey.

## Computer Privacy

As survey methodology increasingly turns to a Web-based medium, a discussion of privacy in surveys cannot afford to ignore characteristics of the survey

presentation. One of the difficulties with studying how respondents feel about Web-based survey is the fact that an individual's privacy concerns about the Internet may or may not be grounded in reality. Most people do not have accurate knowledge of privacy technology and the vulnerabilities of operating systems; at the same time, they may have erroneous beliefs about what are and are not safe practices (Graeff & Harmon, 2002; Jensen, Potts, & Jensen, 2005). The result could be that sensitive data collected by an organization could be jeopardized by a survey designer's flawed anticipation of how data are used, who has access to data once they are collected, and the effects this may have on respondents (Adams & Sasse, 2001). This is especially important considering the breadth of information that an organization might collect, as well as the amount of sensitive or identifying information the organization maintains on its employees.

A survey of 1200 Internet users by Turow and Hennessy (2007) found that the majority of Internet users believe that institutions take measures to ensure information privacy, yet many people also believe institutions and organizations will disclose personal information to third parties. Though most people may not fully understand the provisions protecting the integrity of their responses, they may harbor low perceptions of the confidentiality afforded by the survey or have suspicions regarding the computer upon which a survey is presented, in turn leading to response distortion or nonresponse. Accordingly, any discussion of survey privacy must either explicitly account for the influence of computerized surveys, or transcend the discussion of media entirely and focus on the underlying social and cognitive processes of privacy perceptions. As such, the scale being developed in the current study has been designed specifically to apply equally to paper surveys, computerized surveys (delivered via web or mobile devices), or interviews.

## Defining Perceived Confidentiality

For sensitive and/or personal information, protecting the privacy of respondents by keeping data confidential is of utmost concern to researchers (Ruebhausen & Brim, 1966; Wolf, Zandecki, & Lo, 2004). In practice, this often means that information about individuals is protected from inadvertent disclosure to others by physical means such as a locked cabinet, legal provisions such as signed confidentiality statements, or methodological means such as the use of coded files (Easter, Davis, & Henderson, 2004). As the Internet continues to facilitate data collection through online survey administration, maintaining the confidentiality of records can become problematic. For instance, information from several sources can be combined and used to possibly identify respondents through online databases with greater ease compared to paper records (Leahey, 2007). Yet, confidentiality assurances may matter to respondents only when the content of the survey is considered to be sensitive (Singer, von Thurn, & Miller, 1995). Therefore, there may be instances when an individual completing a survey is not actively concerned about their information being disclosed.

Confidentiality is often poorly-defined in both research as well as in ethics guidelines. For instance, the American Psychological Association's ethical principles and code of conduct (2002) state, "Psychologists have a primary obligation and take reasonable precautions to protect confidential information obtained through or stored in any medium, recognizing that the extent and limits of confidentiality may be regulated by law or established by institutional rules or professional or scientific relationship" (sec. 4.01). However, no definition of what constitutes confidentiality is stated elsewhere in the guidelines. As with anonymity, research on confidentiality has often relied on experimental designs to provide a privacy manipulation. The

underlying logic in this privacy research has been that assignment to conditions will yield different outcomes on variables of interest, and that any observed differences can be attributed to the experimental condition. However, the cumulative results of this type of research have yielded mixed findings (e.g., Singer et al., 1995).

An alternative approach is to consider privacy as the perceptions of a respondent of his or her survey environment. Despite the fact that researchers such as Scott (2005) have proposed that confidentiality and anonymity are "either/or" concepts, this may not translate into how survey respondents think about the differences between the two ideas. What may be likely is that confidentiality is a necessary (but arguably not sufficient) condition for anonymity. That is, for an individual's identity to be completely compromised (i.e., his or her information becomes publicly known), they cannot be anonymous; conversely, an individual's identity may inadvertently become known, but it may still be held confidential by some given entity or individual and not made public.

For this study, perceived confidentiality is defined as being comprised of three facets: (1) the perception of the relative identifiability of a respondent, (2) aspects of informational trust in an institution or institutional representative, and (3) how likely it is that a respondent feels his or her personal information will be protected from disclosure beyond their consent.

With a definition of perceptions of confidentiality established, we developed a measure that can be used in survey research to more accurately gauge the social and cognitive processes that comprise confidentiality as a type of privacy perception. Further, this study will demonstrate that perceived confidentiality is a distinct latent construct from a previously established measure of anonymity perceptions (PANON; Whelan & Thompson, 2009).

# Method

## Participants

Participants for this study were 812 undergraduate students from a large Southeastern university who participated in this study to receive course credit. With regard to gender, 58.1% of the sample was female, and the average age of participants was 19.24 years (*SD* = 3.40).

## Item Generation

A pool of potential items was generated by the author to reflect the social and cognitive facets of the definition of confidentiality outlined above. A total of thirteen items for confidentiality perceptions were created to reflect the construct domain. After the initial pool of items was generated, the items were reviewed separately with two faculty members of the psychology department from which the sample was drawn for revisions and item clarification. The final set of twelve items was decided upon and administered to participants as described below.

## Procedure

Participants in this study volunteered to complete a Web-based survey on academic dishonesty. Respondents to the survey were able to access the survey materials and complete them from any location they chose where they had Internet access on a computer. Privacy assurances were made to each individual who participated in this study, namely that their responses would be anonymous to the researcher and any information they provided would be kept confidential and would not be shared with any third parties except in aggregate form.

After acknowledging that they had read online informed consent materials, participants completed an Academic Dishonesty Inventory (Lucas & Friedrich, 2005) to provide sensitive item content for which respondents should feel some concern

about survey privacy. Next, participants responded to items relating to anonymity and confidentiality perceptions. Lastly, demographic information was collected for each participant, and the name and student ID of each participant was requested for the purposes of assigning course credit for participation before participants were directed to online debriefing materials.

The measures in this survey were presented in the order described above so that demographic information and names would be requested after submitting responses to all measures of interest to prevent raising respondents' suspicions about the privacy assurances that were made. It should also be noted that within each section of the survey, the items within each measure were presented in random order when possible to counteract the potential for item-level order effects (see Feldman & Lynch, 1988).

## Measures

**Anonymity perceptions**. The six-item PANON scale was administered to measure perceptions of anonymity in Web-based surveys (Whelan & Thompson, 2009). Response options were presented on a 5 point Likert-style scale ranging between 1 ("strongly disagree") and 5 ("strongly agree"), with higher values reflecting greater reported levels of perceived anonymity. See Table 1 for a complete list of items and factor loadings.

**Confidentiality perceptions**. Twelve items were administered to measure perceptions of confidentiality in Web-based surveys. Response options were presented on a 5 point Likert-style scale ranging between 1 ("strongly disagree") and 5 ("strongly agree"), with higher values reflecting greater reported levels of perceived confidentiality. See Table 2 for a complete list of retained items and factor loadings.

**Academic dishonesty** (26 items, *KR20* = 0.82). Lucas and Friedrich's (2005) Academic Dishonesty Inventory (Lucas & Friedrich, 2005) was administered to measure socially undesirable behaviors for which there feasibly could be some potential consequence if a respondent's identity was tied to their responses. Responses to this behavioral inventory were dichotomous such that 1 = "yes," and 0 = "no". Although this measure will not be used in subsequent analyses, it should be noted that 87.4% of the sample endorsed at least one item. An example item is, "Paraphrased material from a book without acknowledging the source."

## Results

### Data Cleaning & Preparation

The survey responses were examined for missing data, and, participants were removed from the database if they did not respond to any of the items contained in one or more of the measures collected in this study. Overall, 57 participants were eliminated from the dataset for a total of 755 usable survey responses..

To prepare the data for testing using a series of confirmatory factor analyses (CFAs), the dataset was randomly split into an analysis group ($N$ = 378) and a holdout group ($N$ = 377). The holdout sample will be used only for purposes of testing the final two-factor CFA model as determined by any necessary model respecifications. The measurement models and subsequent CFA models were tested using Mplus version 5 software (Muthén & Muthén, 2007).

### Confirmatory Factor Analysis

A CFA model for the PANON yielded adequate fit, $\chi^2(9) = 14.90$, $p = 0.094$, CFI = 0.99, TLI = 0.98, RMSEA = 0.042, SRMR = 0.027. The items for the PANON, in addition to factor loadings, standard errors, and item-level reliability estimates, are presented in Table 1. The items designed to measure perceived confidentiality were

also subjected to a CFA with all twelve items loading onto a single latent variable. The CFA model yielded moderately poor fit, $\chi^2(54) = 136.94$, $p < 0.001$, CFI = 0.98, TLI = 0.97, RMSEA = 0.064, SRMR = 0.026.

### Model Respecification & Reliability Estimation

Based on the unsatisfactory results for tests of factor structure as described above, all confidentiality items were subjected to respective analyses via the Tetrad3 software using the Purify module (Spirtes et al., n.d.). For the confidentiality items, items 10 and 12 were flagged for deletion. The measurement model for confidentiality was respecified, omitting the variables as suggested. The resultant model yielded adequate fit, $\chi^2(35) = 63.54$, $p = 0.002$, CFI = 0.99, TLI = 0.99, RMSEA = 0.046, SRMR = 0.022. The $\chi^2/df$ ratio of 73.40 ($p < .01$) indicates significantly improved model fit after respecification. The final list of items for this measure, in addition to factor loadings, standard errors, and item-level reliability estimates, is presented in Table 2.

After establishing the scalability via a measurement model of acceptable fit, reliability estimates for each scale were calculated using the $R_{max}$ framework proposed by Drewes (2000), which utilizes the item-level reliability estimates to construct a scale-level index of maximum reliability. For the confidentiality perceptions scale, the $R_{max}$ value was 0.95. These indices show that the scale displays adequate reliability within their respective measurement models.

### Factor Correlations

After establishing the reliability of the perceived confidentiality measure, a two-factor CFA was conducted for the anonymity and confidentiality scales (see Figure 1 for a path diagram of this CFA model). The CFA model yielded moderately

poor fit, $\chi^2(103) = 287.49$, $p < 0.001$, CFI = 0.947, TLI = 0.938, RMSEA = 0.069, SRMR = 0.060. Modification indices for the CFA model tested above suggested that the fifth anonymity item cross-loaded on the confidentiality perceptions scale; as such, the CFA model was respecified to allow this relationship.

The respecified CFA model yielded adequate fit, $\chi^2(101) = 176.96$, $p < 0.001$, CFI = 0.978, TLI = 0.974, RMSEA = 0.045, SRMR = 0.038. The interfactor correlation between anonymity perceptions and confidentiality perceptions for the respecified model was $r = .626$. This indicates a strong relationship between the concepts of anonymity and confidentiality, as would be expected, while simultaneously demonstrating that both constructs have a distinct unidimensional latent factor structure.

### Cross-Validation

The holdout sample of 377 survey respondents was tested using CFA to replicate the model parameter estimates of the analysis sample. A two-factor CFA was conducted for the anonymity and confidentiality scales, with item 5 cross-loading on the confidentiality factor as before. The results of the CFA using the holdout data were similar to the previous sample with the model indicating adequate fit, $\chi^2(101) = 193.77$, $p < 0.001$, CFI = 0.975, TLI = 0.970, RMSEA = 0.049, SRMR = 0.041. The correlation between the anonymity perceptions and confidentiality perceptions factors was $r = 0.587$ for this sample. The results of this CFA suggest that the factor structures and interfactor relationship of the anonymity and confidentiality measures are relatively stable across samples.

## Discussion

This study developed and collected preliminary validity evidence for a measure of confidentiality perceptions. The systematic development of this scale provides a theoretical framework and empirical means through which to assess perceptions of confidentiality in survey research. As defined earlier, confidentiality perceptions are comprised of three facets: (1) the perception of the relative identifiability of a respondent, (2) aspects of informational trust in an institution or institutional representative, and (3) how likely it is that a respondent feels his or her personal information will be protected from disclosure beyond their consent. The analyses conducted for this study provide evidence that the measure constructed based on the above definition is unidimensional, and that while strongly correlated, the constructs represented by these scales can be considered to tap distinct privacy domains.

## Limitations and Future Research

A notable limitation of this study is the use of undergraduate students as a study sample. In an applied setting, finding an effect for the perceptions of a survey respondent might be dependent on other variables not typically present in university research, such as the use of computer monitoring by an employer.

Another potential limitation of this study concerns the nature of the previous research that has been drawn from, particularly in regards to explaining disinhibited behavior and computers. Much of the research considering anonymity and disinhibition does so in the context of computer-mediated communication, or CMC (e.g., Joinson, 2003; Kiesler, Siegel, & McGuire, 1984; Postmes, Spears, & Lea, 1998). As such, CMC by its very nature could include significantly more social cues than a survey, since CMC is typically interactive in some form or another (Joinson, 1998).

Future research should take into account the anonymity and confidentiality perceptions of participants to test the effects of intentional and unintentional differences in experimental conditions where privacy may be a concern of survey respondents. For instance, exploring the potential for confidentiality in predicting outcomes could provide insight into the influence of contextual variables in Web-based survey research that heretofore have been considered confounds. As noted by researchers such as Ong and Weiss (2000), the relative sensitivity of items can be subjective, so measuring perceived anonymity and confidentiality may be able to assist researchers in identifying when reports of behavior are low due to privacy concerns. Lastly, research should explore what types of threats to privacy can impact anonymity and confidentiality perceptions, and how these threats may lead to differences in outcomes depending on the salience of the privacy threat.

## References


Adams, A., Sasse, A., 2001. Privacy in Multimedia Communications: protecting users, not just data. In: Blandford, A., Vanderdonkt, J., Gray, P. (Eds.), *People and Computers XV—Interaction Without Frontiers: Joint Proceedings of HCI2001 and IHM2001*. Springer, Berlin, pp. 49–64.

American Psychological Association (2002). Ethical principles of psychologists and code of conduct. Washington, DC: APA. Available online at http://www.apa.org/ethics/code2002.pdf

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*(4), 562-566.

Booth-Kewley, S., Larson, G. E., & Miyoshi, D. K. (2007). Social desirability effects on computerized and paper-and-pencil questionnaires. *Computers in Human Behavior, 23*, 463-477.

Corey, S. M. (1937). Signed versus unsigned attitude questionnaires. *Journal of Educational Psychology, 28*, 144-148.

Drewes, D. W. (2000). Beyond the Spearman-Brown: A structural approach to maximal reliability. *Psychological Methods, 5*, 214-227.

Dunnette, M. D., & Heneman, H. G. (1956). Influence of scale administrator on employee attitude responses. *Journal of Applied Psychology, 40*, 73-77.

Easter, M. M., Davis, A. M., & Henderson, G. E. (2004). Confidentiality: More than a linkage file and a locked drawer. *IRB: Ethics & Human Research, 26*, 13-17.

Feigelson, M. E., & Dwight, S. A. (2000). Can asking questions by computer improve the candidness of responding? A meta-analytic perspective. *Consulting Psychology Journal: Practice and Research, 52*(4), 248-255.

Feldman, J. M., & Lynch, J. G. (1988). Self-generated validity and other effects of measurement on belief, attitude, intention, and behavior. *Journal of Applied Psychology, 73*, 421-435.

Graeff, T. R., & Harmon, S. (2002). Collecting and using personal data: consumers' awareness and concerns. *Journal of Consumer Marketing, 19*(4), 302-318.

Jensen, C., Potts, C., & Jensen, C. (2005). Privacy practices of Internet users: Self-reports versus observed behavior. *International Journal of Human-Computer Studies, 63*(1-2), 203-227.

Joinson, A. N. (1998). Causes and implications of disinhibited behavior on the Net. In J. Gackenbach (Ed.), *Psychology and The Internet* (pp. 43-60). New York: Academic Press.

Joinson, A.N. (2003). *Understanding the psychology of Internet behaviour: Virtual worlds, real lives*. Basingstoke: Palgrave Macmillan

Kiesler, S., Siegel, J., & McGuire, T. W. (1984). Social psychological aspects of computer-mediated communication. *American Psychologist, 39*, 1123-1134.

Leahey, E. (2007). Convergence and confidentiality? Limits to the implementation of mixed methodology. *Social Science Research, 36*, 149-158.

Lucas, G. M., & Friedrich, J. (2005). Individual differences in workplace deviance and integrity as predictors of academic dishonesty. *Ethics & Behavior, 15*, 15-35.

Muthén, L. K., & Muthén, B. O. (2007). *Mplus: Statistical analysis with latent variables*. Software available at http://www.statmodel.com

Ong, A. D., & Weiss, D. J. (2000). The impact of anonymity on responses to sensitive questions. *Journal of Applied Social Psychology, 30*, 1691-1708.

Postmes, T., Spears, R., & Lea, M. (1998). Breaching or building social boundaries? SIDE-effects of computer-mediated communication. *Communication Research, 25*, 689-715.

Rogelberg, S. C., Spitzmüeller, C., Little, I., & Reeve, C. L. (2006). Understanding response behaviors to an online special topics organizational satisfaction survey. *Personnel Psychology, 59*, 903-923.

Ruebhausen, O. M., & Brim, O. G. (1966). Privacy and behavioral research. *American Psychologist, 21*, 423-437.

Sackett, P. R., & Larson, J. R. (1990). Research strategies and tactics in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., pp. 419-489). Palo Alto, CA: Consulting Psychology Press.

Scott, C. R. (2005). Anonymity in applied communication research: Tension between IRBs, researchers, and human subjects. *Journal of Applied Communication Research, 33*, 242-257.

Singer, E., von Thurn, D. R., & Miller, E. R. (1995). Confidentiality assurances and response: A quantitative review of the experimental literature. *Public Opinion Quarterly, 59*, 66-77.

Spirtes, P., Scheines, R., Meek, C., Richardson, T., Glymour, C., Hoijtink, H., & Boomsma, A. (n.d.). *Tetrad III* (software). Available at http://www.phil.cmu.edu/projects/tetrad/

Turow, J., & Hennessy, M. (2007). Internet privacy and institutional trust: Insights from a national survey. *New Media & Society, 9*, 300-318.

Whelan, T. J., & Thompson, L. F. (2009, April). *Development/validation of the PANON scale assessing survey anonymity perceptions*. Paper presented at the 24th annual conference of the Society for Industrial and Organizational Psychology, New Orleans, LA.

Wolf, L. E., Zandecki, J., & Lo, B. (2004). The Certificate of Confidentiality application: A view from the NIH institutes. *IRB: Ethics & Human Research, 26*, 14-18.

Table 1

*Items, Factor Loadings, Standard Errors, and Reliability Estimates for PANON*

| Item Stem | $\lambda$ | *SE* | $R^2$ |
|---|---|---|---|
| 1. I feel my responses are indistinguishable from the responses of others that have taken this survey. | 0.789 | 0.028 | 0.622 |
| 2. It would be impossible to trace my responses to this survey back to me. | 0.475 | 0.045 | 0.226 |
| 3. I feel that my responses are unidentifiable from the responses of others. | 0.794 | 0.028 | 0.630 |
| 4. My responses will blend in with the responses of other people. | 0.599 | 0.039 | 0.359 |
| 5. I feel certain that this survey is anonymous. | 0.581 | 0.040 | 0.338 |
| 6. If someone saw my responses, they would never know who it was who filled out the survey. | 0.510 | 0.043 | 0.260 |

Table 2

*Items, Factor Loadings, Standard Errors, and Reliability Estimates for Confidentiality Perceptions*

| Item Stem | $\lambda$ | *SE* | $R^2$ |
|---|---|---|---|
| 1. I feel sure that all of my answers will be kept secret. | 0.802 | 0.020 | 0.644 |
| 2. I feel sure that my answers to this survey will never be used against me. | 0.649 | 0.031 | 0.421 |
| 3. Nobody else besides the researchers will be able to see my answers to this survey without my approval. | 0.735 | 0.025 | 0.540 |
| 4. I feel certain that my answers to this survey will be kept private. | 0.882 | 0.013 | 0.778 |
| 5. I feel certain that this survey is confidential. | 0.843 | 0.017 | 0.711 |
| 6. I feel certain that my answers to this survey will not be seen by anyone but the researchers. | 0.845 | 0.016 | 0.714 |
| 7. I feel certain that all information I disclose to the researchers will be held confidential. | 0.871 | 0.014 | 0.758 |
| 8. I feel certain that information I disclose to the researchers will be held confidential unless there is a good reason for them to share it. | 0.432 | 0.043 | 0.186 |
| 9. I feel certain that my answers to this survey will not be shared with anyone else by the researchers. | 0.755 | 0.024 | 0.570 |
| 10. The researchers are committed to respecting my privacy. | 0.755 | 0.024 | 0.571 |

**Figure 1**

*Path Diagram for Two-Factor CFA Model*