

# An Evaluation Of The Reliability Of Alternative Survey Media

Dr. MaryAnne Atkinson, Accounting, University of Delaware  
Dr. Scott Jones, Accounting, University of Delaware

## Abstract

*This study reports the results of an experiment that evaluated the reliability of alternative survey media. The components of reliability are compared for instruments characterized as paper-based and computer-based. The evaluation was further partitioned by subjects' preference for visual versus verbal stimuli. Results indicate that there is no significant difference in scale responses between the alternative media, but reliability was found to be low when subjects with verbal cognitive style were asked to respond in the visually oriented computer environment.*

## Introduction

Data collection for survey research, including rating experiments, is typified by the use of a questionnaire. Designs for studies involving questionnaires are frequently criticized for threats to both internal and external validity. Elaboration on these threats can be found in numerous sources, such as Campbell and Stanley (1963), and Kidder (1981). Domain specific sources also exist, for example, Abdel-Khalik and Ajinkya (1979).

The usual objectives of survey research are fact-finding, hypothesis generation, response scaling, or attitude/belief measurement. Typical business applications often involve collecting customer opinions or developing profiles, evaluating employee attitudes, and collecting process data or employee input from quality circles.

External validity is often not questioned in business surveys if the response rate is high, and there is no evidence of non-response bias (Kidder, 1981). Internal validity is typically the most serious problem for survey designs. The limiting factor to internal validity is reliability, which is the degree to which instrument scales and collected responses represent actual phenomena. Reliability differs from validity since a reliable test would yield results that are independent of the application, while a valid test measures what it purports to measure.

Reliability is sometimes referred to as accuracy or dependability (Cronbach, 1951), and may be divided into these components (Krippendorff, 1980): stability, reproductibility, and accuracy.

A test is stable if it is invariant with time, reproducible if results can be recreated under different circum-

stances, and accurate if it conforms to a known standard. These components are to some extent dependent upon instrumentation and individual perceptions. Stability is a weak form of reliability, and threats to stability stem from instrumentation inconsistencies such as code-recode variation, and measurement, description, or categorical inconsistencies. Similarly, reproducibility and accuracy may be subject to instrumentation error if there are intra- and inter-person inconsistencies in the way instructions are interpreted. These threats to reliability limit the internal validity of a test, hence instrumentation is a widely recognized limitation to internal validity, especially in test-retest designs (Campbell and Stanley, 1963).

Since reliability is at least partially dependent on the observer or subject's perception of the stimulus, it may be hypothesized that an individual's preference for visual or verbal stimuli may affect reliability.

## Visual-Verbal Preference

Stimuli that attempt to elicit responses fit along a continuum from spatial/analog to verbal/linguistic. It is thought that spatial and verbal stimuli use two different working memory systems (Baddely and Hitch, 1974). Inasmuch as this is the case, the redundant use of pictorial and printed instructions is a common human engineering practice that addresses different groups in the population (Wickens, 1987).

Survey stimuli, or questionnaire components are traditionally verbal/linguistic, and therefore may not elicit reliable responses if an individual's preference is for the spatial/analog format. We therefore hypothesize that a test instrument that involves the use of visual and

verbal tasks will elicit more reliable information.

The proliferation of business microcomputing provides a new medium for instrumentation. In general, the hardware and software required to implement a survey research test may preclude the use of that instrument in many cases, but none-the-less, the medium is becoming a more pervasive force in everyday business communications. The benefits of the computer as an appealing tool in testing are enumerated by Erdman, et al. (1983), and include the ability to make completeness tests, record data, and allow for help queries. We add to that list the ability to implement visual and verbal tasks.

### Prior Studies

Comparisons between computer-based and paper-based instrumentation methods have been made in several other studies. Sproull (1986) compared data collected through electronic mail with that collected by paper survey. She noted no substantive differences in mean responses, a higher response rate, shorter elapsed time, and a lower missing data rate for the computer-based questionnaire. It was noted that the questions were attempting to elicit factual and objective data, and that the results may not be the same for emotional or attitudinal data since the pressure to respond in socially desirable ways may not be present in a computerized medium. Research by Erdman, et al. (1983) seems to indicate that the pressure to respond in a socially desirable manner was greater for the computer-based instrument than the pencil and paper based instrument. Their evidence suggests that the methods yield similar results except the computer-based responses reported lower incidence of substance abuse. The subjects did indicate that the computer-based survey was significantly more interesting than the pencil and paper survey. Steinbart (1989) documented a difference in mean responses to faculty evaluations for paper and computer-based formats. He concluded that the computer responses were representative of attitudes, but not comparable to the paper-based questionnaire. It should be noted that the computerized surveys in these studies involved tasks that were largely verbal, and in a sense may not use the spatial/analog aspect of working memory. In essence, any observed differences may be attributable to a "novelty effect", rather than a substantive difference in the way that the stimuli are perceived. The conclusions of these studies are tenuous and not entirely consistent, and none of the studies directly tests the reproducibility component of reliability, but instead address only the accuracy issue. Prior studies may also be limited in that there is no explicit attention given to the visual aspects of the task. Sproull (1986) suggests that it is difficult to convey visual stimuli using the computer medium. This of course is not generalizable to all computer interfaces.

### Hypermedia Instrumentation

Hypermedia is a technology that involves the simultaneous use of multiple media, and hence provides the computer environment necessary to implement a multiple-task survey instrument. The concept was defined by Kahn and Meyerowitz (1988) as "non-sequential writing". The idea was to provide a rich learning environment composed of text, graphics, sound, and video. The user could access these media and combine them in innovative ways to explore a variety of new dimensions related to a concept. These computer-based systems are thus designed to provide both verbal and visual stimuli, since words, symbols, and pictures can be presented in the hypermedia environment.

Current generations of computers make this technology possible. There are several different application systems that let a user develop hypermedia environments. Reports on the successful implementation of hypermedia systems are included in Yankelovich, Landow, and Heywood (1987), and Yankelovich, Haan, Meyerowitz, and Drucker (1988).

Based on the characteristics discussed above, the following three hypotheses were tested. Stability was not directly tested because it represents the weakest form of reliability (Krippendorff, 1980). Furthermore, the questionnaire that was used did not involve a test-retest design.

H<sub>1</sub>: The reproducibility of the hypermedia instrument is equal to the reproducibility of the paper instrument:

$$\alpha_H = \alpha_P$$

where alpha is Cronbach's coefficient of equivalence, H is the hypermedia instrument, and P is the paper instrument. Hypothesis H1 is tested by comparing the alpha values of the respective treatments. Alpha is a measure of reliability proposed by Cronbach (1951):

$$\alpha = \frac{n}{(n-1)} * \left( 1 - \frac{\sum_{i=1}^n V_i}{V_t} \right) ;$$

where:  $V_i$  = variance of individual item scores,  
 $V_t$  = variance of tests scores; and,  
 $n$  = number of items on the testing instrument.

Alpha is the corrected ratio of inter-item covariance to total covariance, and interpreted as the correlation between two tests given at the same time. Since the percentage of variance attributable to error diminishes as the number of items increases (Cronbach, 1951), it is worth noting that it is the number of scale items and the relative importance of the first principal component that has the greatest influence on alpha, not the number of subjects.

H<sub>2</sub>: The accuracy of the hypermedia instrument is equal to the accuracy of the paper instrument:

$$\text{Mean}_H = \text{Mean}_P$$

where Mean is the mean score of all subjects obtained from the respective instruments. Hypothesis H2 is tested by comparing the means derived from the hypermedia instrument to the paper instrument. The instrument that we use was developed and validated by O'Connell (1985), consequently comparisons for accuracy are made against the paper-based results that are obtained.

H<sub>3</sub>: The cognitive style of the individual does not affect the reliability of the instrument:

$$\begin{aligned} \alpha P_{*visual} &= \alpha P_{*verbal} \\ \alpha H_{*visual} &= \alpha H_{*verbal} \end{aligned}$$

where, for example,  $\alpha P_{*visual}$  is alpha for visual subjects in the paper treatment. Hypothesis H3 is tested by dividing subjects into a dichotomous partition of preference for visual versus verbal tasks as identified from the results obtained. Individual alpha values for each partition are computed.

### The Measurement Device

The instrument used in this study was developed and validated by O'Connell (1985). The scale measures an individual's preference for visual and verbal cognitive style, and is a portion of a larger instrument developed to measure several aspects of an individual's cognitive style. Individual cognitive style has been an important variable in many areas of business research. Some examples can be found in marketing (Foxall and Goldsmith, 1988 and Foxall et al., 1990), management (Velthouse, 1990), information systems (Igarria and Parasuraman, 1989 and Vasallo and Lanasa, 1990/1991), and accounting (Gul, 1987).

There are twenty-two items in the scale, which measures an individual's visual/verbal preference. The score on the scale is a total of the responses to each item. Item responses range from one (highly visual) to seven (highly verbal). The scale's external validity was tested through a comparison with an alternative measure of cognitive style. In addition, the internal validity was measured through the use of Cronbach's alpha. O'Connell reported alpha values ranging from 0.73 to 0.81, suggesting a high degree of internal validity for the scale. Results indicate that the scale appears to measure the construct it was intended to measure.

### Administration of the Scale

The test scale was administered in two ways: one group was given the scale in paper form, and another

completed the scale in a hypermedia form. Assignment was random, but not counterbalanced. The paper form was a replica of the test administered by O'Connell (1985). The subjects were primarily practicing accountants and financial executives.

The hypermedia form of the test was written in Apple Computer's HyperCard. Figure 1 shows the screen that asks the first two questions in the scale. Similar screens presented the remaining twenty questions. Subjects were given an on-line tutorial explanation of the use of the mouse to point to the button corresponding to their choice, how to click the button to indicate their response, and how to move to subsequent screens. The responses and instructions were symbolic and involved an extensive use of pictures. Subjects could change their answers, but omissions or multiple responses were not possible. When finished, each subject was given a brief description of his or her cognitive style based on their responses. They were also given the opportunity to enter comments about the questionnaire. The tasks involved in the hypermedia form can be described as a combination of visual and verbal. The individuals had to read the questions, but the instructions, response recording tasks (clicking a button), and tasks involved in moving among questions (clicking on arrows) are spatial/analog. Consequently, these individuals were exposed to both visual and verbal stimuli.

The paper-based form was administered to the subjects in the traditional manner, with check marks placed above the selected response points. Individuals had the opportunity to make written comments, but immediate feedback was not provided. Because few visual tasks are involved in answering the questions, the paper-based instrument is entirely verbal.

### Results

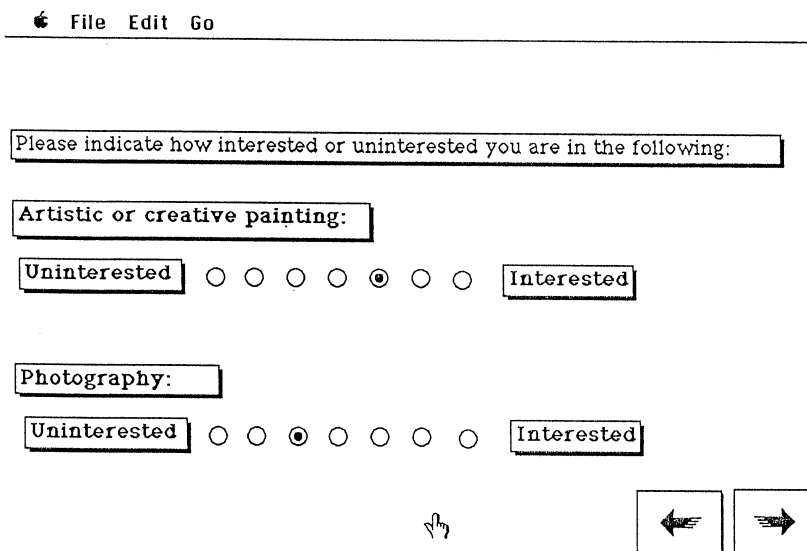
The results are presented in two tables. Table 1 shows the reliability coefficients, mean responses and test statistics for between means differences. Table 2 shows the reliability coefficients partitioned based on subjects' scores on the visual/verbal scale.

Table 1 indicates that the alpha values for the hypermedia and paper treatments are 0.75 and 0.70, respectively. O'Connell (1985) reported values between 0.73 and 0.81. Accordingly, there appears to be weak support for concluding that the hypermedia environment is more reliable, since the scale values are more in line with those reported by O'Connell. But since the differences in alpha values between the treatments is quite small, hypothesis H1 is supported: there appears to be no significant differences in reliability between the treatments.

To test Hypothesis H2, the differences in mean

Figure 1

Example Screen



responses between the treatments is compared. Since the scale has been validated by O’Connell, the results obtained from the paper treatment would be the standard by which the accuracy of the hypermedia treatment should be judged. The parametric Student’s t is not significant at any reasonable level, while the non-parametric Mann-Whitney is only marginally significant at 0.10. This evidence supports H2, suggesting that the hypermedia form is at least as accurate as the paper form. This result is inconsistent with some prior research (Erdman, et al., 1983; Steinbart, 1989), which concluded that the responses do differ between computer and paper based instruments, but are consistent with Sproull (1986), in which no differences were observed. It should be noted that the means observed are larger than those reported by O’Connell, which is likely attributable to the nature of our sample. The majority of subjects were current or former practicing accountants. O’Connell (1985) noted that certain individuals, including accountants, would be dominated by a preference for verbal stimuli and would score higher on the scale than the general public. Similar indications that accountants would favor a verbal rather than visual cognitive style is evident in Amernic and Beechy (1984).

The results shown in Table 2 are based on a partition of observed scores. Individuals scoring above the mean score were placed in the "verbal" group, while those scoring below the mean were placed in the "visual" group. In both the paper and hypermedia treatments, four subjects were verbal and six were visual. A partition based on the median score was also evaluated. Except for group size, the results are comparable, and are not shown.

The highest reliability is observed for verbal subjects in the paper treatment(0.78) and visual subjects in the hypermedia treatment(0.82). Reliability for both subject partitions is similar for the paper treatment(0.78 vs 0.70), but markedly different for the hypermedia treatment(0.82 vs. 0.51). Based on these alpha values, H3 must be rejected for the hypermedia instrument, but is supported for the paper instrument. The differences in reliability coefficients in the hypermedia partitions, and similarity in reliability coefficients in the paper partitions may be attributed to a learned adaptation. It is feasible that subjects preferring visual stimuli have adapted to the common verbal-based paper questionnaires, and therefore respond in a way that is similar to those subjects characterized as verbal. On the other hand, individuals with a verbal cognitive style may not have

Table 1  
Reliability Coefficients and Mean Responses

	Questionnaire Form	
	Hypermedia	Paper
Cronbach’s alpha	0.75	0.70
Scale mean	92.40	87.40
Scale standard deviation	12.55	13.74
Number of Items	224.00	22.00
Number of Subjects	104.00	10.00

Pooled standard deviation = 13.16; t for significant difference between means = 0.85(p > 0.10); and Mann-Whitney for difference between medians = 33(p = 0.10).

adapted to the relatively new computer environment, and therefore provide less reliable results when asked to respond in that context.

Table 2  
Reliability Coefficients for Partitioned Subjects

Partition	k	Questionnaire Form	
		Hypermedia	Paper
Verbal Subjects	4	0.51	0.78
Visual Subjects	6	0.82	0.70

k = number of subjects in partition; n = 22.

### Conclusions

Contrary to some prior studies, no differences in mean score were observed in alternative survey media. The reliability of the computer medium was found to be dependent on the cognitive style of the individual. This phenomenon may be short term as verbal individuals may adapt to the more visual computer environment. The reliability of the paper medium was not found to be dependent upon cognitive style, perhaps due to adaptive behavior. The results suggest that when the implementation advantages of a computer recommend its use in a survey study (e.g. convenience, demand for complete data, need to provide on-line help, and the desire to avoid face to face interviews), it may be as reliable as a paper-based survey. It must be cautioned that reliability should be checked, especially if the individual may be biased towards a verbal cognitive style, since the computer-based survey may not be reliable under those conditions.

### Suggestions for Future Research

It would be a useful extension of this study to investigate survey reliability for other groups of subjects, including students, since a significant amount of survey research uses students as subjects. Our subject group was practicing accountants and financial executives. Different results might be obtained for students or other subject groups.

Future research might also consider the reliability of computer-based surveys for different types of computers. Our study was conducted on an Apple Macintosh computer and the survey incorporated some of this computer's graphical features. Perhaps different results may be found using computer systems featuring different user interfaces.

### \*\*\*References\*\*\*

- Abdel-khalik, A. Rashad and Bipin Ajinkya. *Empirical Research in Accounting: A Methodological Viewpoint*. Sarasota, Florida: American Accounting Association, 1979.
- Amernic, Joel H. and Thomas H. Beechy. "Accounting Students' Performance and Cognitive Complexity: Some Empirical Evidence." *The Accounting Review*. 1984, Vol 59 (2), 300-313.
- Baddely, A.D. and G. Hitch. "Working Memory." in G. Bower, Ed., *Recent Advances in Learning and Motivation*. (Vol 8) New York: Academic.
- Campbell, Donald T. and Julian C. Stanley. *Experimental and Quasi-Experimental Designs for Research*. Boston: Houghton Mifflin Company, 1963.
- Cronbach, Lee J. "Coefficient Alpha and the Interval Structure of Tests." *Psychometrika*. 1951, Vol 16 (3): 297-334.
- Erdman, Harold; Marjorie H. Klein; and John H. Greist. "The Reliability of a Computer Interview for Drug Use/Abuse Information." *Behavior Research Methods & Instrumentation*. 1983, Vol 15 (1): 66-68.
- Foxall, Gordon R. and Goldsmith, Ronald E. "Personality and Consumer Research: Another Look." *Journal of the Market Research Society (UK)*. 1988, Vol 30 (2): 111-125.
- Foxall, Gordon R.; Payne, Adrian F.; Taylor, James W. and Bruce, Grady D. "Marketing and Non-Marketing Managers." *Marketing Intelligence & Planning (UK)*. 1990, Vol 8 (1): 21-26.
- Gul, Ferdinand A. "Field Dependence Cognitive Style as a Moderating Factor in Subjects' Perceptions of Auditor Independence." *Accounting & Finance*. 1987, Vol 27 (1): 37-48.
- Igbaria, Magid and Parasuraman, Saroj. "A Path Analytic Study of Individual Characteristics, Computer Anxiety and Attitudes Toward Microcomputers." *Journal of Management*. 1989, Vol 15 (3): 373-388.
- Kahn, Paul and Norman Meyrowitz. "Guide, HyperCard and Intermedia: A Comparison of Hypertext/Hypermedia Systems." *IRIS Technical Report Number 88-7*.
- Kidder, Louise H. *Research Methods in Social Relations*. Fourth Edition. New York: Holt, Rinehart and Winston, 1981.
- Krippendorff, Klaus. *Content Analysis: An Introduction to its Methodology*. Beverly Hills: Sage Publications, 1980.
- Logan, Elisabeth. "Cognitive Styles and Online Behavior of Novice Searchers." *Information Processing & Management (UK)*. p62 1990, Vol 26 (4): 503-510.
- O'Connell, Thomas R. Development and Preliminary Validation of an Instrument Measuring Four Cognitive and Affective Psychological Dimensions. Unpublished Ph.D. Dissertation, University of North Dakota, 1985.
- Sproull, Lee S. "Using Electronic Mail for Data Collection in Organizational Research." *Academy of*

- Management Journal*, 1986. Vol 29 (1): 159-169.
17. Steinbart, Paul John. "A Comparison of Computer-Administered and Paper Questionnaires: Effects on Student Evaluations of Teaching Effectiveness." *Issues in Accounting Education*. 1989, Vol 4 (1): 120-125.
  18. Vasallo, Helen G. and Lanasa, John M. "The Effects of Cognitive Style on the Design of Expert Systems." *Review of Business*. 1990/1991 Vol 12 (3): 37-42.
  19. Velthouse, Betty A. "Creativity and Empowerment: A Complementary Relationship." *Review of Business*. 1990 Vol 12 (2): 13-18.
  20. Walter, Mark. "IRIS Intermedia: Pushing the Boundaries of Hypertext." *The Seybold Report on Publishing Systems*, Vol 11 (21).
  21. Wickens, Christopher D. "Information Processing, Decision-Making, and Cognition." in G. Salvendy, ed. *Handbook of Human Factors*. New York: John Wiley & Sons, 1987.
  22. Yankelovich, Nicole; Bernard J. Haan; Norman K. Meyrowitz; and Steven M. Drucker. "Intermedia: The Concept and the Construction of a Seamless Information Environment." *Computer*, January, 1988.
  23. Yankelovich, Nicole; George Landow; and Peter Heywood. "Designing Hypermedia 'Ideabases' - The Intermedia Experience." *IRIS Technical Report Number 87-4*.