

The Effect of Data Collection Modality on Students' Foreign Language Survey

Yann-Yann Shieh

American Institutes for Research, Washington DC

This paper was prepared for presentation at the annual meeting of the American Association for Public Opinion Research, May 2003, Nashville, TN. Opinions expressed in this paper are those of the author, and no endorsement of the American Institutes for Research should be inferred.

KEY WORDS: Survey data collection, Mode effect

Abstract: Surveys are an invaluable tool for helping diverse research institutes, government agencies, and business organizations to gather information needs. Selection of data collection techniques is generally based on cost, completion time and response rate. Traditionally, surveys have been done using paper-pencil methodologies, which can be costly to administer in terms of time, labor, and materials. Due to the popularity of computer usage, the Internet has dramatically increased the ease and speed of survey administration and data collection, as well as decreasing associated costs, making surveys faster, easier and cheaper than telephone or paper-pencil collecting methods. Researches have shown that paper-pencil and Internet/web methods have often been considered to produce similar results, however, there is some evidence that people may respond differently depending on modality of administration. In addition, because of the important role that surveys play in today's society, it is crucial that the psychometrics of any administered survey be evaluated to assure its technical soundness. The purpose of this study is to investigate the effect of data collection modes on the language survey administered via paper and pencil, Internet/web and PDA (palm), as well as to evaluate the psychometric structure of high school students' foreign language self-assessment survey. Data from a self-assessment survey administered to 314 high school senior students across 10 states via paper and pencil, Internet and PDA methods were examined. The Students self-assessment foreign language survey was evaluated for the survey dissemination modalities as well as technique soundness and internal reliability. Results showed that for listening, speaking, reading, writing, and total scores, students used paper-and-pencil administration had higher score than those who used PDA, and students used Internet administration had higher score than those who used PDA.

Perspectives

Surveys are an invaluable tool for helping diverse research institutes, government agencies, and business organizations to gather information needs. Selection of data collection techniques is generally based on cost, completion time and response rate. Traditionally, surveys have been done using paper-pencil methodologies, which can be costly to administer in terms of time, labor, and materials. Due to the popularity of computer usage, the Internet has dramatically increased the ease and speed of survey administration and data collection, as well as decreasing associated costs, making surveys faster, easier and cheaper than telephone or paper-pencil collecting methods.

In the twenty-first century, the Internet is having a profound effect on the survey research. It has become possible to post assessment inventories on the Internet (Musch & Reips, 2000). The rapid development of surveys on the World Wide Web (WWW) is leading some to debate that Internet, in particular, Web, surveys will replace traditional methods of survey data collection such as paper-and pencil or face-to-face interview. Others are urging caution or even voicing skepticism about the future role Web surveys will play (Couper, 2000). The attractiveness of the World Wide Web as a medium for collection of psychological information and research has led to a surge of studies and articles (Birnbau, 2000).

According to Couper (2000), Web surveys represent a double-edged sword for the survey industry. The power of Web surveys is that they make survey data collection available to the majority public. Not only can researchers get access to numbers of respondents at dramatically lower costs than traditional methods, but members of the general population can also put survey questions on dedicated sites offering free services and collected data from potentially thousands of people. The ability to conduct large-scale data collection is no longer restricted to organization (Couper, 2000). The advantages of collection information through Web are numerous and include decreased experimenter demand (Buchanan, 2000) and social desirability

effects (Joinson, 1999), reduced missing data (Stanton, 1998), avoidance of data entry errors because responses are entered directly into the database (Pasveer & Ellard, 1998), saving money and time, possibly greater self-disclosure by participants (Davis, 1999), and making feasible the delivery of multimedia survey content to respondents in a standardized way using self-administered methods (Couper, 2000). However, these potential advantages do not guarantee the generalizability of assessments conducted using the Internet (Pasveer & Ellard, 1998).

On the other hand, the potential risk of Web surveys is that it is difficult to distinguish the good from bad with the proliferation of Web surveys. The value of surveys that could be done on the Web is limited by the willingness of people to do them (Couper, 2000). Potential problems associated with using the Internet for assessment include (a) the lack of controlled environment that allows responding to measures at whatever time and in whatever setting suits the respondent, also allowing for repeat or mischievous responding (Buchanan, 2000); (b) important differences in the layout of questionnaire items depending on the respondents' browser software and settings (Baron & Siepmann, 2000); and (c) potential violations of privacy and issues of data security (Cho & LaRose, 1999). Studies that have explored the comparability of computer (non-Web) administrations with paper-and-pencil data collection have raised concerns about whether negative attitudes toward computers affected responses to computerized instruments (Schulenberg & Yitzenka, 1999; Tseng, Macleod, & Wright, 1997). Several studies have found subtle differences between data collected in the two modalities (Miles & King, 1997; Potosky & Bobko, 1997; Schwartz, Mullis, & Dunham, 1998; Webster & Compeau, 1996). As Buchanan (2000) pointed out, these problems suggest that equivalent reliability and validity cannot be assumed for online Internet-collected and traditional paper-and pencil assessment.

Purpose of Study

To date, few studies have explored differences between online Internet and traditional survey responses (Krantz & Dalal, 2000), but those that have (e.g., Pasveer & Ellard, 1998; Stanton, 1998) have found 'remarkable congruence (Krantz & Dalal, 2000, p.35). This accords with the results of a number of studies comparing paper-and-pencil versus computer (non-Web) administration of instrument. Given the multiple potential differences between on-line Internet and paper-and-pencil assessments, the

degree of equivalence found thus far is both surprising and worthy of further investigation. Thus, the purpose of this study is to investigate the effect of data collection modes on the language survey administered via paper and pencil, Internet/web and PDA (palm), as well as to evaluate the psychometric structure of high school students' foreign language self-assessment survey.

Methods

Data Source: Data for this study were from American Institutes for Research (AIR). In October 2002, AIR conducted a small-scale tryout of five components Foreign Language assessment among high school seniors in schools around the country. The Language Survey sections of the LSBQ (Language Survey/Background Questionnaire) were administered in three modes: paper-pencil, Internet, and PDA versions.

Participants: AIR's target sampling design called for approximately 300 students (15 students at each of 20 schools) to complete the LSBQ. We recruited a large proportion of students to complete both the LSBQ for efficiency so that we could assess the relationship between performance on the Language Screener. Only high schools that offered Spanish and agreed to allow 15 seniors to participate in the interview were scheduled to be visited as part of the small-scale tryout. The final numbers of participants were 314 administered by three modes across 10 states. Table 1 shows the participants description.

Procedures: Once the schools and students were recruited, the students were administered language survey using three modes: paper-and-pencil, Internet, and PDA versions. For the paper-and-pencil version, participants were provided with an instrument that they could complete under supervision of examiners. For the Internet version, participants were provided with a World Wide Web address where they can complete the self-assessment language survey. For the Personal Digital Assistant (PDA) version, PDAs were distributed to the participants then instructions were given to them. Instructions for the PDAs took longer than instructions for the paper-and-pencil and Internet versions.

Language Assessment Measures: The foreign language self-assessment survey contains five parts: (a) listening which consists of 13 items; (b) speaking which consists of 12 items; (c) reading which consists of 11 items; (d) writing which consists of 10 items; and (e) proficiency which consists of 16 item. For reading, speaking, reading and writing, students

were asked to answer the questions with ‘Yes’ and ‘No’ responses. If students answered ‘Yes’ in the item, students will get one point on that item. For proficiency measure, only students who answered the question right will get a point on the item. Originally there were 16 items on the proficiency measure, due to the computer equipment problem, only 15 items were read. Therefore, we only used 15 items of proficiency in this report.

Table 1
Participants by Administration Mode and States

STATE	Mode			Total
	Paper-Pencil	Internet	PDA	
AR	2	2	1	5
CA	9	12	10	31
CT	8	9	9	26
DC	4	4	3	11
FL	5	5	5	15
GA	9	2	18	29
IL	4	5	4	13
MI	15	13	13	41
OR	9	11	12	32
TX	4	7	6	17
UT	4	5	5	14
VA	4	10	7	21
WI	18	22	19	59
Total	95	107	112	314

Analysis and Results

Scale Analysis

The psychometric characteristics of each of the language assessment scales were examined with the data collected via the three modes of administration (paper-and-pencil, Internet, and PDA). These analyses were based on the study participants who responded to every item of a given assessment scale. The estimated properties (e.g., means, standard deviation, coefficient alpha, and interscale correlation) of each measure show the some similarity across the three modes of delivery. The number of study participants responding to every item on each scale and summary psychometric properties are reported in Table 2 and Table 3. For example, for the reliability (coefficient alpha), PDA administration tended to have higher coefficient alpha than paper-and-pencil and Internet administration except for Total scores. In terms of inter-correlation among measures, Internet administration tended to have a higher correlation

between listening and reading measures ($r=0.71$) than paper-and-pencil administrations. PDA administration tended to have a higher correlation between speaking and reading measures ($r=0.77$) than paper-and-pencil and Internet administrations. Paper-and-pencil administration tended to have a higher correlation between writing and proficiency measures ($r=0.50$) than PDA and Internet administrations.

Table 2
Scale Statistics for Paper-and-Pencil, Internet and PDA of Foreign Language Assessment

Mode	N	Low-est Observed	High-est Observed	Mean	SD	α
Pencil Speaking	95	2	16	11.65	3.20	0.85
Reading	95	2	11	7.02	2.51	0.81
Writing	95	2	10	5.97	2.35	0.81
Proficiency	95	5	15	12.35	2.40	0.71
Total	95	14	50	32.17	9.14	0.87
Inter- net Listening	107	2	13	8.09	2.82	0.80
Speaking	104	1	16	12.41	3.39	0.88
Reading	104	2	11	7.62	2.45	0.80
Writing	104	0	10	6.63	2.52	0.82
Proficiency	103	2	15	11.89	2.54	0.60
Total	107	5	50	34.00	10.64	0.87
PDA Listening	112	1	13	4.81	3.21	0.86
Speaking	112	1	16	9.81	3.83	0.89
Reading	112	1	11	5.60	2.69	0.84
Writing	112	1	10	5.18	2.34	0.82
Proficiency	112	5	15	12.64	2.26	0.70
Total	112	4	50	25.40	10.60	0.84

MANOVA Scale Score Analyses

Analyses assessing the comparability of the scale scores derived from the measures in the paper-and-pencil, Internet, and PDA were conducted on those individuals who provided responses to at least 90% of the items underlying each of the language scale measure. Scoring was conducted under the assumption of a consistent response pattern to the items underlying each scale. One-way multivariate analysis of variance (MANOVA) was conducted on fix dependent variables (listening, speaking, reading, writing, proficiency, and total scores) to investigate the effect of modality. The value of Wilk’s Λ , the

approximate F value, and probability associated with the approximate F are reported for the mode effect.

An effect size measure (eta square) is also reported. The result is showed in Table 4.

Table 3
Inter-correlation among Language Scale by Mode Administration

Mode	Speak- ing			Read- ing			Writ- ing			Profici- ency			Total		
	Paper- Pencil	Inter- net	PDA	Paper- Pencil	Inter- net	PDA	Paper- Pencil	Inter- net	PDA	Paper- Pencil	Inter- net	PDA	Paper- Pencil	Inter- net	PDA
Listen- ing	0.65 (95)	0.67 (104)	0.65 (112)	0.67 (95)	0.71 (104)	0.66 (112)	0.69 (95)	0.68 (104)	0.58 (112)	0.39 (95)	0.30 (103)	0.17 (112)	0.87 (95)	0.79 (107)	0.83 (112)
Speak- ing				0.60 (95)	0.61 (104)	0.77 (112)	0.72 (95)	0.73 (104)	0.72 (112)	0.60 (95)	0.58 (103)	0.26 (112)	0.88 (95)	0.88 (104)	0.91 (112)
Read- ing							0.56 (95)	0.78 (104)	0.79 (112)	0.40 (95)	0.30 (103)	0.24 (112)	0.82 (95)	0.86 (104)	0.90 (112)
Writ- ing										0.50 (95)	0.41 (103)	0.40 (112)	0.86 (95)	0.90 (104)	0.86 (112)
Profici- ency													0.56 (95)	0.47 (103)	0.30 (112)

Note. N count is shown in the parenthesis.

Table 4
MANOVA Results and Follow-up ANOVA Results

Effect	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
MODE	Wilks' Lambda	0.730	10	606	0.000	0.145

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
MODE	LISTEN	664.534	2	332.267	39.541	0.000	0.205
	SPEAK	391.935	2	195.968	15.975	0.000	0.094
	READ	225.961	2	112.981	17.309	0.000	0.101
	WRITE	108.063	2	54.032	9.381	0.000	0.058
	SPROF	30.420	2	15.210	2.649	0.072	0.017
	TOTAL	4998.590	2	2499.295	25.402	0.000	0.142

The MANOVA result showed statistically significant main effect for mode of administration. The follow-up ANOVA results showed that statistically mean scores differences among mode administration on listening, speaking, reading, writing, and total scores expect proficiency score. Multiple comparisons using Tukey method showed that for listening, speaking, reading, writing, and total

scores, students used paper-and-pencil administration had higher score than those who used PDA, and students used Internet administration had higher score than those who used PDA. The results of multiple comparisons were presented in Table 5.

Summary and Conclusion

The psychometric, distributional, and MANOVA analytic results evidenced some differences among the responses of study participants in the paper-and-pencil, Internet, and PDA versions of studies. Overall, the mode effects when present were showed in the current study. There are systematic patterns with regard to statistically significant scale mean differences—students with paper-and-pencil administration had higher assessment scores than those with PDA administration; students with Internet administration had higher assessment scores than those with PDA administration; and there were no means scores difference between students with paper-and-pencil and Internet administrations on language self-assessment survey. The findings in this study suggest the importance of the continued evaluation of mode effects. However, when conducting this self-assessment language survey, several difficulties were observed and encountered. First, Students were most likely to talk to one another in paper-and-pencil mode. Students seemed less engaged in this mode. Second, when conducting Internet administration, some students started but could not complete the entire assessment because the computer froze. Third, When PDAs were distributed before instructions had been given, students immediately turned on the PDAs and began to play with them. Some students expressed envy of those who were assigned to PDAs.

In recent years, many researchers and testing or survey companies have shifted from paper-and-pencil to computer administration of psychological measures or assessment. The advent of Internet or computer related technology and the ability to post questionnaires on the Web or PDA further extend opportunities for assessing and research psychological function. Although this study was a small-scale tryout, the results of this study represent an important contribution. As Krantz and Dalal (2000) point out, few studies to date have explored differences between Web and traditional survey responses. As with the current study, we have expanded our data collection modes not only to the paper-and-pencil and Web modes but also the PDA mode. Our findings open the door for researchers to use the information from the self-assessment language survey that has been collected online, paper-and-pencil, and PDA, and prepare for the future development of large-scale language survey.

References

Baron, J. & Siepmann, M. (2000). Techniques for creating and using Web questionnaires in research and teaching. In M. H. Birnbaum (Ed.),

Psychological experiments on the Internet (pp.235-266). San Diego, CA: Academic Press.

Birnbaum, M. H. (2000). *Psychological experiments on the Internet*. San Diego, CA: Academic Press.

Cho, H., & LaRose, R. (1999). Privacy issues in Internet surveys. *Social Science Computer Review*, 17, 421-434.

Couper, M. P. (2000). Web surveys. *Public Opinion Quarterly*, 64, 464-495.

Davis, R. N. (1999). Web-based administration of a personality questionnaire: Comparison with traditional methods. *Behavior Research Methods, Instrument, and Computer. Assessment*, 31, 365-374.

Joinson, A. (1999). Social desirability, anonymity and Internet-based questionnaires. *Behavior Research Methods, Instrument, and Computer. Assessment*, 31, 433-438.

Krantz, J. H., & Dalal, R. (2000). Validity of Web-based psychological research. In M. H. Birnbaum (Ed.) *Psychological experiments on the Internet* (pp/35-60). San Diego, CA: Academic Press.

Miles, E. W., & King, W. C. (1997). Gender and administration mode effects when pencil-and-paper personality tests are computerized. *Educational and Psychological Measurement*, 58, 68-76.

Musch, J., & Reips, U. (2000). A brief history of Web experiment. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 61-87). San Diego. CA: Academic Press.

Pasveer, K. A., & Ellard, J. H. (1998). The making of a personality inventory: Help from the WWW. *Behavior Research Methods, Instrument, and Computer. Assessment*, 30, 309-313.

Potosky, D., & Bobko, P (1997). Computer versus paper-and-pencil administration mode and response distortion in noncognitive selection tests. *Journal of Applied Psychology*, 82, 293-299.

Schulenberg, S. E., & Yutzenka, B. A. (1999). The equivalence of computerized and paper-and-pencil psychological instrument: Implications for measures of negative affect. *Behavior Research Methods, Instrument, and Computer. Assessment*, 31, 315-321.

Schwartz, S. J., Mullis, R. L., & Dunham, R. M. (1998). Effects of authoritative structure in the measurement of identity formation: Individual computer-managed versus group paper-and-pencil testing. *Computer in Human Behavior*, 14, 239-248.

Tseng, H., Macleod, H. A., & Wright, P. (1997). Computer anxiety and measurement of mode change. *Computer in Human Behavior*, 13, 305-316.

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

Table 5
Summary of Tukey Multiple Comparisons Results

Dependent Variable	(I) Mode	(J) Mode	Mean Difference (I-J)	Std. Error	Sig.
LISTEN	paper-pencil	Internet	-.5708	.4124	.349
		PDA	*2.7138	.4043	.000
	Internet	paper-pencil	.5708	.4124	.349
		PDA	*3.2846	.3957	.000
	PDA	paper-pencil	*-2.7138	.4043	.000
		Internet	*-3.2846	.3957	.000
SPEAK	paper-pencil	Internet	-.7843	.4982	.257
		PDA	*1.8401	.4885	.000
	Internet	paper-pencil	.7843	.4982	.257
		PDA	*2.6244	.4781	.000
	PDA	paper-pencil	*-1.8401	.4885	.000
		Internet	*-2.6244	.4781	.000
READ	paper-pencil	Internet	-.5615	.3634	.270
		PDA	*1.4228	.3564	.000
	Internet	paper-pencil	.5615	.3634	.270
		PDA	*1.9843	.3488	.000
	PDA	paper-pencil	*-1.4228	.3564	.000
		Internet	*-1.9843	.3488	.000
WRITE	paper-pencil	Internet	-.6238	.3414	.161
		PDA	*.7898	.3347	.048
	Internet	paper-pencil	.6238	.3414	.161
		PDA	*1.4137	.3276	.000
	PDA	paper-pencil	*-.7898	.3347	.048
		Internet	*-1.4137	.3276	.000
SPROF	paper-pencil	Internet	.4542	.3409	.377
		PDA	-.2955	.3343	.650
	Internet	paper-pencil	-.4542	.3409	.377
		PDA	-.7497	.3272	.057
	PDA	paper-pencil	.2955	.3343	.650
		Internet	.7497	.3272	.057
TOTAL	paper-pencil	Internet	-2.5403	1.4110	.169
		PDA	*6.7666	1.3835	.000
	Internet	paper-pencil	2.5403	1.4110	.169
		PDA	*9.3070	1.3541	.000
	PDA	paper-pencil	*-6.7666	1.3835	.000
		Internet	*-9.3070	1.3541	.000

Based on observed means.

* indicates that the mean difference is significant at the .05 level.

Address for correspondence: Yann-Yann Shieh,
American Institutes for Research, 1000 Thomas
Jefferson St. NW, Washington DC, 20007, U.S.A.
Email: yshieh@air.org. Phone: 202-342-5048.