

# THE EFFICACY OF DIARY ASSESSMENTS: USING DIARY ASSESSMENTS TO EVALUATE RESPONDENTS' LEVEL OF PERFORMANCE ON ALTERNATIVE DIARY FORMS

Leslie A. Miller and Sharon Krieger, Bureau of Labor Statistics  
Leslie A. Miller, 2 Massachusetts Ave., N.E. Rm 3650, Washington, DC 20212

KEY WORDS: Assessments, Difficulty, Effort

Diaries are often used to collect information about daily activities, such as consumer spending, since they are particularly useful for collecting small, inexpensive items (e.g., Flueck, Waksberg, & Kaitz, 1971). However, there are differences in the procedures used to collect this type of data and in the structural form of the diary instruments. Thus, research has explored various diary procedures (e.g. See Grootaert, 1986; Kemsley & Nicholson, 1960; Sudman & Ferber, 1971), and has examined respondents' level of performance on different diary forms (e.g., Tucker, Vitrano, Miller, & Doddy, 1989; Tucker, 1992).

Evaluating respondent performance on a diary that inquires about daily life is frequently more difficult than it might appear at first. For example, to know for certain how often people eat breakfast, have colds, or buy fast food would require that their every movement be recorded around the clock or at least randomly sampled. Such methods of appraisal are simply infeasible for many different diary topics and in many cases are much too costly to be undertaken. Fortunately, there are general methods for evaluating respondent performance in diaries (e.g., See Corby & Miskura, 1985; Groves, 1989).

## Comparing Data to External Sources

One way of evaluating respondents' level of performance is by comparing data from diaries to data from external records. Generally, information from external records provides the most reliable record of "truth." Respondent performance is usually considered better on the diary in which respondent reports more closely match external records.

## Conducting Reinterviews

Conducting reinterviews offers another way of evaluating respondents' level of performance. In reinterviews, information is collected from the same respondent at two different times. In general, reinterviews involve asking the same questions asked in the first interview. Respondents are considered to have performed better on the diary when they respond with the same answers during both interviews.

## Using Information Collected at the Time of the Survey

Although widely used, independent sources and reinterviews may not always be the best ways to make extrapolations about respondents' level of performance on any given instrument. Other possible methods for evaluating performance rely on information from the survey itself. Two such methods include exploring diary response patterns and using post-interview assessments. When exploring diary response patterns, researchers may opt to compare, for example, response rates and mean reported events. Respondents are usually considered to have performed better on the diary with higher response rates or on the diary with higher mean expenditures. When using post-interview assessments, researchers may choose to compare the difficulty respondents have with different diaries and the effort they put into completing the diary. It can be presumed that the diary that fostered more effort and was less difficult is the diary that respondents have performed better on.

These methods have several advantages. First, compared to the other methods, they are fairly inexpensive since the information can already be found in the data and because the assessment is conducted immediately following an interview. Second, developing new interviewing procedures is not necessary. Third, searching for external sources and generalizing from a subset of the sample are avoided.

## Current Research

The current research focuses on the value of respondent and interviewer post-interview assessments for evaluating respondents' level of performance on two diary forms used in the 1991 Consumer Expenditure (CE) Interview Program field test. First, using the assessment information from this field test, we explore the difficulty associated with the diary forms and whether respondents and interviewers perceived one form as being more difficult than the other. We also explore the level of conscientious effort respondents put into the forms and whether respondents and interviewers thought that more conscientious effort was put into one diary than the other. In addition, we explore whether respondents and interviewers agreed on the amount of difficulty associated with each diary and the amount of conscientious effort respondents put into each diary. Second, we discuss the implications

of our results for the Bureau's decision to introduce the Experimental Diary as the new Production Diary. We conclude with a discussion about whether this type of assessment is helpful/useful and we make suggestions for improvement.

### Overview of the Consumer Expenditure Diary Survey Field Test

As a part of the Bureau's CE Program, BLS samples approximately 5000 consumer units annually to keep an expenditure diary (the CE Diary Survey) for two one-week periods<sup>1</sup>.

Like other survey sponsors, the BLS is always concerned about underreporting. Out of a concern that underreporting was contributing to the response error on the CE Diary, BLS undertook a research program in the early 1980's to develop an alternative diary which provided respondents with more cues.

This research was basically conducted in three phases. The current paper focuses on this final phase of research. For this final phase, in 1991 and 1992, an Experimental Diary (Diary B) and the current Production Diary (Diary A) were used in production<sup>2</sup>. The Production Diary was given to 80% of 6319 CU's, while the Experimental Diary, which is longer, more structured, and provides more cues, was given to 20% of the CU's. The goal of this field test was to evaluate which diary provided better measures of consumer spending patterns so that one could be selected to remain in production.

Following a brief household characteristics interview, all CU's were asked to keep the diary they were selected to receive for two one-week periods. In addition, during the pick-up of the second week diary each CU was asked a series of questions about various aspects of the diary keeping process. Interviewers were also asked to answer some general questions following each second week diary pickup.

#### Results

In presenting the results, we first provide an overview of the types of participants in the field test followed by participants' patterns of response. In the second section of the results, we focus on the difficulty respondents had with each of the diaries. We provide a brief description of the response distributions at the question level for both the respondent and interviewer difficulty questions. Then, we create a difficulty scale for both respondent and interviewer questions in order

to compare the difficulty associated with each diary at a more general level. For each diary form, we correlate difficulty measures from respondent and interviewer reports to see the extent to which respondents and interviewers agree. In the third section of the results, we focus on the conscientious effort respondents put into the diaries. We provide an analysis of conscientious effort in the same manner described for respondent difficulty.

### Types of Participants and Patterns of Response

Diaries were placed with 6319 consumer units (CU's); Diary A with 5133 CU's and Diary B with 1186 CU's for both one-week periods.

All 6319 CU's were characterized as being either 1) two-week participants, 2) one-week participants, or 3) non-participants. Two-week participants kept the diary for both weeks. One-week participants kept the diary for one of the two weeks. Non-participants had diaries placed with them, but were either away both weeks or away for one week and non-respondents for the other week.

Of the 6319 CUs, 5458 (86.4%) were two-week participants, 742 (11.7%) were one-week participants, and 119 (1.9%) were non-participants. For Diary A, 86.1% were two-week participants, 12.1% were one-week participants, and 1.8% were non-participants. For Diary B, 87.8% were two week participants, 10.0% were one week participants, and 2.2% were non-participants. The types of participants were not significantly different between the two diaries  $X^2(2, N=6319) = 4.7, p > .10$ .

Two-week and one-week participants were further characterized by their patterns of response and their recall status during each week, and by whether the respondent and the interviewer assessment questionnaires were completed. For each week, the diary could have been completed by the CU without any recall (i.e. the CU completed each day of the diary), with partial recall (i.e., the CU left some information blank in the diary and the interviewer had the respondent recall these items during diary pickup), or with total recall (the CU did not enter any information into the diary and the interviewer had the respondent recall all of their expenditures during diary pickup).<sup>3</sup> Furthermore, for each CU, both the respondent and interviewer assessment questionnaires may have been completed, only one may have been completed, or no assessment may have been completed.

The following analyses of difficulty and conscientious effort include only those *two-week participants* who completed the diary *without total*

---

<sup>1</sup>A CU is a person or a group of related persons in a sample household who are either related by blood, legal marriage or adoption, or who as a person is independent of all other persons in the household for payment of their major expenses.

<sup>2</sup>For a complete description of the two diaries and the first two phases, see Tucker, Vitano, Miller, and Dobby (1989).

---

<sup>3</sup>In some cases, the recall status is unknown, but prior research has shown that this group consists of primarily partial recallers.

recall and for which a respondent or an interviewer assessment was completed.

### Difficulty

The assessment questions asked of respondents and interviewers that tapped difficulty are shown in Tables 1 and 2 respectively. Don't know responses were not considered in the following analyses.

#### Brief description of response distributions.

Six questions asked of respondents and one question asked of interviewers inquired about difficulties or problems respondents had with each of the diaries. Concerning those questions that were asked of respondents, we figured that CU's found the diary less difficult when they a) thought the instructions and examples provided enough information about how to make entries in the diary (respondent question 1), b) did not find it difficult to decide where to record items in the diary (respondent question 2), c) did not find it difficult to decide what information to write down for of the expenditures (respondent question 3), and d) did not find it more difficult to remember to keep the diary the longer they had it (respondent question 4). Although we included respondent question 5 as a difficulty question, we were unsure as to whether spending more time on the diary was a measure of difficulty or a measure of engagement. As for interviewer questions, we figured that the diary was less difficult for those CU's who interviewers thought had no problems completing the diary overall (interviewer question 1).

Overall, respondents' answers to both diaries indicate that the majority of respondents did not have any major difficulty with either diary. The majority of respondents found the instructions and examples to be adequate, did not have difficulty deciding where to record items, did not have difficulty deciding what information to write down, and did not find it more difficult to remember to keep the diary the longer they had it. Interviewers' answers to both diaries also indicated that the majority of respondents did not have any problems completing the diary.

Comparison of two forms. Respondent questions 1, 3, and 4 and interviewer question 1 had relatively the same distributions of responses for both diaries. Only respondent questions 2, 2a, 3a and 5 had response distributions that were noticeably different between the two diaries. Respondent question 2 indicates that respondents who kept the Experimental Diary had more difficulty deciding where to record some expenditures than respondents who kept Diary A. As shown by respondent question 2a, for those who reported that they had difficulty deciding where to record expenditures, more Diary B respondents

reported that they had difficulties with three or more types of expenditures than Form A Respondents. As shown by question 3a, of those respondents who reported it was difficult to decide what information to write down, more Diary B respondents reported problems with three or more expenditure categories. Lastly, respondent question 5 suggests that respondents who kept Diary B spent more time recording expenditures than respondents who kept Diary A.

Creation of a difficulty measure. In order to further substantiate whether the degree of difficulty differed between the two diaries, we created a difficulty measure. Respondent questions 1, 2, 3, and 4 were used to create a difficulty measure. Questions 2a and 3a were not included since they were dependent upon the answers of previous questions. Question 5 was not included, and was analyzed separately, since we were unsure as to whether more time spent on a diary meant that the diary was more difficult, or that the respondents were more engaged in the diary task.

Responses to these questions were scored and standardized, with the highest score assigned to the answer suggesting the most difficulty. A difficulty score was created for each respondent by summing their responses and dividing by the number of questions they were asked. The mean scores for each diary were compared. Interviewers' answers to the one difficulty question they were asked were compared as well.

The results suggest that according to respondents, Diary B was more difficult than Form A,  $t(1061) = 7.16, p < .0001$ . According to interviewers, Diary B was also more difficult than Diary A,  $t(1073) = 1.99, p < .05$ . A Pearson correlational analyses suggested that respondent and interviewer difficulty scores for Diary A ( $r = .20$ ) and Diary B ( $r = .31$ ) were positively correlated ( $p < .0001$ ).

Furthermore, the results suggest that respondents spent more time completing Diary B Diary than Form A  $t(1072) = 5.7, p < .0001$ .

### Conscientious Effort

In order to compare the conscientious effort respondents put into Diary A with the conscientious effort respondents put into Diary B, we selected from the assessment questionnaire those questions that inquired about the conscientious effort respondents put into each diary. The conscientious effort questions asked of respondents and interviewers are shown in Tables 3 and 4 respectively. Don't know responses were not considered in the following analyses. Some of the responses were collapsed.

#### Brief description of response distributions.

Five questions asked of respondents and four questions

Table 1

**Respondent Difficulty Questions for Diary A and Diary B Comparison**

#	Question	Response	Diary A	Diary B
1	Did the instructions and examples provide enough information about how to make entries in the diary?	YES	3028 (98%)	737 (97%)
		NO	77 (2%)	25 (3%)
2	Was it difficult to decide where to record some items in the diary?	YES	374 (13%)	245 (32%)
		NO	2722 (87%)	517 (68%)
2 a	For which types of expenditures was it difficult?	ONE TYPE	290 (79%)	180 (75%)
		TWO TYPES	52 (14%)	34 (14%)
		THREE OR MORE	26 (7%)	26 (11%)
3	Was it difficult to decide what information to write down for some of your CU's expenditures?	YES	239 (8%)	74 (10%)
		NO	2872 (92%)	686 (90%)
3 a	For which types of expenditures was it difficult?	ONE TYPE	189 (78%)	53 (77%)
		TWO TYPES	43 (18%)	9 (13%)
		THREE OR MORE	11 (4%)	7 (10%)
4	Was it more difficult to remember to keep the diary the longer your CU had it?	YES	577 (19%)	156 (20%)
		NO	2490 (81%)	606 (80%)
5	How much time was spent during the past week making entries in the diary?	30 MINUTES OR LESS	2202 (69%)	448 (57%)
		31-60 MINUTES	772 (24%)	252 (32%)
		61-90 MINUTES	165 (5%)	51 (7%)
		91-120 MINUTES	50 (2%)	29 (4%)
		> 120 MINUTES		

Table 2

**Interviewer Difficulty Questions for Diary A and Diary B Comparison**

#	Question	Response	(Diary A)	(Diary B)
1	Do you think the respondent had problems completing the diary?	YES	318 (10%)	98 (13%)
		NO	2757 (90%)	655 (87%)

asked of interviewers inquired about the conscientious effort respondents put into each of the diaries. As for respondent questions, we figured that CU's put more effort in the diary task when a) other members reported their expenditures daily to the primary recording person (respondent question 1), b) CU's made entries in the diary on a daily basis (respondent question 2), c) CU's had other members reporting their expenditures (besides the person responsible for recording expenditures) (respondent question 3), d) CU's did not have to make entries in the diary for expenditures purchased earlier in the week (respondent question 4), and e) CU's used receipts and other records to help keep the diary (respondent question 5). As for interviewer questions, we figured that interviewers thought CU's put more effort in the diary task if CU's a) seemed cooperative (interviewer question 1), b) used receipts or other expenditure records (interviewer question 2), c) made a conscientious effort to complete the diary every day (interviewer question 3), and d) did not have items to recall at the end of either week (interviewer question 4).

Overall, respondents' answers to both diaries indicate that many respondents made a conscientious effort in completing the diaries, however their effort was not ideal. Overall, only a little more than one half of the respondents reported that they entered expenditures daily. Furthermore, while many multiple member CU's reported that other CU members' expenditures were recorded in the diary, and these expenditure were reported to the diary keeper daily, about 20% reported that other CU members' expenditure were not recorded in the diary, and about 40% did not report daily. About 25% of respondents reported that it was necessary to make entries for expenditures made earlier during the week. Lastly, the majority of respondents reported using receipts and other records.

Interviewers' answers to both diaries indicate that many respondents made a conscientious effort to complete the diaries, but there is again some indication that this effort was not ideal. Although interviewers perceived that the majority of respondents were cooperative and made a conscientious effort to complete the diary every day, interviewers reported that more than 20 percent of the respondents had items to recall at the end of either week. Furthermore, according to interviewer reports for both diaries, slightly more than half of the respondents who had to recall items at the end of either week used receipts or other records to aid recall.

Comparison of two forms. All respondent and interviewer questions had relatively the same distributions of responses for both diaries.

Creation of an effort measure. In order to further substantiate whether one diary fostered more effort than the other diary, we created a conscientious effort measure. All six questions asked of respondents were used to create this conscientious effort measure. Responses to these questions were scored and standardized, with the highest score assigned to the answer suggesting the most conscientious effort. A conscientious effort score was created for each respondent by summing their responses and dividing by the number of questions they answered. The mean scores for each diary were compared. The same was done for the questions asked of interviewers.

No differences were seen in the amount of effort put into Forms A and B in either the respondent reports or the interviewer reports. Respondent and interviewer reports about Diary A ( $r=.30$ ) and Diary B ( $r=.22$ ) were positively correlated ( $p<.0001$ ).

Overall, our results indicate that the diaries were not that difficult for respondents, but nonetheless, Diary B was more difficult for respondents than Diary A. However, looking back at the questions that were asked of respondents, and were included in the difficulty measure, it is very likely that question 2 was the primary contributor to these results. Overall, both respondent and interviewer measures support the notion that Diary B was more difficult than Diary A, but their reports were only weakly correlated.

Our results also indicate that respondents were fairly effortful reporters and this did not differ between the diaries. Once again, overall, respondent and interviewer measures support this, but their reports were only weakly correlated. Lastly, our results indicate the respondents using Diary B spent more time recording expenditures than respondents using Diary A. But, we are not sure whether this implies that Diary B was more difficult than Diary A, or that Diary B engaged respondents more than Diary A.

#### Comparison of Assessment Results with Other Measures of Performance

Following the 1991/1992 field test of Diary A and Diary B, a committee consisting of representatives from the BLS decided on the criteria to be used in evaluating the data quality of these two diaries. The primary measure of data quality included comparisons of weekly mean expenditures for the two diaries, the reporting rate on each diary, and the response rates for the two diaries.

Results indicated that there was a substantial increase in the means for most categories of expenditures when Diary B was used. Furthermore, the results indicated that Diary B had a slightly greater mean number of records per week. There was no difference in the response rates. Based on these

criteria and the results, Diary B was selected to become the new diary to be used in production.

It would have been nice if in the current research we had found that Diary A was more difficult for respondents than Diary B, and/or that a great deal more effort was put into Diary B than was put into Diary A. In fact, although our results suggest that Diary B was more difficult and took more time for respondents than Diary A, the results are not so definitive. First, it appears as if respondents really only had more difficulty on Diary B deciding where to record items. Given the fact that this diary contains more pages per day and more pages overall, this should be expected, and should not be considered a strong indication of difficulty. Furthermore, neither diary caused respondents any great difficulties and respondents put a decent amount of effort into both diaries.

In actuality, the results of these analyses nicely compliment the Bureau's decision to select the Diary B as the new Production Diary. They offer insight on the few things respondents found difficult about both diaries and the one thing they found particularly more difficult about Diary B. First, the results suggest that given the length of Diary B, we should somehow make it easier for respondents to decide where to record items. Second, we should take steps to make it easier for respondents to decide exactly what information needs to be written down. Third, we should consider ways to help respondents remember to keep the diary over the two-week period. Such improvements may, in fact, decrease the amount of time respondents spend recording expenditures in the diary.

#### Discussion

Overall, the assessment questionnaires were useful in pointing out what steps we could take in order to improve Diary B. However, the research reported in this paper has some limitations. First, although we suggested that one of the appeals of post-interview assessments was that generalizing from a subset of the sample to the entire sample was avoided, unfortunately, this was not the case for the current research. Interviewer and respondent assessments were not collected from all respondents. More careful monitoring could have ensured that both a respondent and an interviewer assessment be collected from all CU's.

Second, we made assumptions about the assessment questions. For example, while we were not so quick to assume that more time spent on a diary means more difficulty, we did assume that those who recorded their expenditures more frequently were

Table 3

**Respondent Conscientious effort Questions for Diary A and Diary B Comparison**

#	Question	Response	(Diary A)	(Diary B)
1	How frequently did other CU members report their expenditures to (primary diary keeper)?	DAILY OTHER	1236 (65%) 659 (35%)	313 (64%) 179 (36%)
2	When were entries usually recorded on the diary?	DAILY OTHER	1859 (60%) 1235 (40%)	453 (60%) 299 (40%)
3	Besides (the person responsible for recording expenditures), were expenditures of other persons in this CU entered in the diary?	YES NO	1827 (81%) 439 (19%)	472 (83%) 96 (17%)
4	Was it ever necessary to make entries in the diary for expenditures made earlier during either week?	YES NO	707 (32%) 1535 (68%)	171 (30%) 399 (70%)
5	Were receipts or other expenditure records used to help keep the diary?	YES NO	2779 (87%) 410 (13%)	685 (88%) 95 (12%)

Table 4

**Interviewer Conscientious effort Questions for Diary A and Diary B Comparison**

#	Question	Response	(Diary A)	(Diary B)
1	Did the respondent seem cooperative or uncooperative about keeping the diary	COOP. UNCOOP.	2894 (97%) 82 (3%)	711 (98%) 18 (2%)
2	Were receipts or other expenditure records used by the respondent to aid recall?	YES NO	378 (51%) 357 (49%)	91 (55%) 73 (45%)
3	Do you think the respondent made a conscientious effort to complete the diary every day?	YES NO	2541 (88%) 349 (12%)	631 (90%) 74 (10%)
4	Did the respondent have items to recall at the end of either week?	YES NO	713 (23%) 2428 (77%)	160 (21%) 617 (79%)

putting more effort into the diary. It is possible that respondents who actually waited to record their expenditures at the end of the week, when they could

have spent more time on the diary, were more effortful reporters.

Third, we used these assessment questions to create difficulty and effort measures post hoc. These measures might have been more valid if we had had a better idea of what exactly each question was measuring and had constructed the questions to fit measures that were developed prior to the research.

**Future Research**

As shown here, the respondent and interviewer post interview assessments were useful for learning about the difficulty respondents had with the diaries and about the amount of conscientious effort respondents put into the recording process. However, future research should address the above limitations of this research. Specifically, future research should attempt to develop difficulty and effort scales that have a high reliability and can be used for evaluations of other diaries sponsored by the BLS, as well as diaries sponsored by other organizations.

**References**

- Corby, C. & Miskura, S. (1985). Evaluating data quality in the economic and decennial censuses. *Proceedings of the First Annual Research Conference*, U.S. Bureau of the Census, 159-175
- Flueck, J.A., Waksberg, J. & Kaitz, H.B. (1971). An overview of Consumer Expenditure Survey Methodology. *Proceedings of the Business and Economic Statistics Section*, American Statistical Association, 238-246.
- Grotaert, C. (1986). The use of multiple diaries in a household expenditure survey in Hong Kong. *Journal of the American Statistical Association*, 81, 938-944.
- Groves, R.M. (1989). *Survey errors and survey costs*. New York: John Wiley and Sons.
- Kemsley, W.F.F. & Nicholson, J.L. (1960). Some experiments in methods of conducting family expenditure surveys. *Journal of the Royal Statistical Society, Ser.A*, 123, Part 3, 307-328.
- Sudman, S. & Ferber, R. (1971). Experiments in obtaining consumer expenditures by diary methods. *Journal of the American Statistical Association*, 66, 725-735.
- Tucker, C., Vitrano, F., and Miller, L.A., & Doddy, J., (January 1989). *Cognitive research and issues on the Consumer Expenditure Survey*, Paper presented at the meeting of the AAAS, San Francisco, CA.
- Tucker, C., Vitrano, F., and Miller, L.A., & Doddy, J. (March, 1989). *More cognitive issues and the Consumer Expenditure Diary Survey*, Paper presented at the South Eastern Psychological Association Meeting, Washington, DC.
- Tucker, C., Vitrano, F., and Miller, L.A., & Doddy, J. (May, 1989). *Cognitive issues and research on the Consumer Expenditure Diary Survey*, Presented at the American Association for Public Opinion Research, St. Petersburg, FL.
- Tucker, C. (1992). The estimation of instrument effects on data quality in the Consumer Expenditure Diary Survey. *Journal of Official Statistics*, 8, 41-61.
- Tucker, C., (May, 1988). *Estimating measurement error in surveys*. Paper presented at the 1988 Annual Conference of the American Association for Public Opinion Research, Toronto, Canada.