

# The Efficacy of Cues in an Expenditure Diary

NHIEN TO  
ERIC FIGUEROA  
LUCILLA TAN

In designing any survey, it is important to provide respondents with clear instructions and examples. Self-administered expenditure diaries often use cues as examples, not only to aid recall, but also to prompt the respondent as to what types of expenses to record and how those expenses should be recorded. This cognitive study investigates how cues should be used in an expenditure diary to instruct respondents to record their expenses completely and accurately.

## Background

The Consumer Expenditure Diary (CED) Survey is a nationwide survey of households used by the U.S. Bureau of Labor Statistics (BLS) to collect expenditures on small, frequently purchased items. The respondent is asked to record the household's expenses for 2 consecutive weeks. Depending on how promptly the respondent records the expenditures in the diary after incurring them, various degrees of recall are involved in the task. To aid in recall, diary forms are often organized into broad categories (e.g., "Food and Drinks for Home Consumption" or "Clothing, Shoes, Jewelry, and Accessories") and include cues that are examples of expenditure items.

Over the years, the use of cues in the CED has undergone a variety of changes. The first annual CED, implemented in 1980, was organized into five

broad expenditure categories that were repeated for each day of the week, resulting in a diary that was 23 pages long. There were 76 specific cues<sup>1</sup> on the recording pages for each day.

In 1991, a new version of the diary (the Current Diary) was introduced. In this version, the five broad expenditure categories were further divided into 42 subcategories (e.g., an "Eggs and Dairy Products" subcategory within the "Food for Home Consumption" category). As a result, there were 305 specific cues on the recording pages for each day. A field test conducted in 1991 showed that, for items mentioned in the cues, the Current Diary yielded higher reporting rates with relatively higher reporting detail than did the 1980 diary.<sup>2</sup>

Despite the Current Diary's strong performance in the field test, declining response rates and diminishing data quality during the 1990s led CED researchers to reexamine the diary and the diary-keeping task. A previous test in 1985 had revealed some disadvan-

<sup>1</sup> Specific cues are precise examples of items described with sufficient detail for coding. For example, "powdered milk" and "whole milk" are specific cues because they contain enough information to be accurately coded. By contrast, "milk" is not a specific cue, because it does not specify the type of milk.

<sup>2</sup> Silberstein, A.R., "Part-Set Cuing in Diary Surveys," paper presented at the annual meeting of the American Statistical Association, 1993.

Nhien To, Eric Figueroa, and Lucilla Tan are research economists in the Branch of Research and Program Development, Division of Consumer Expenditure Surveys, Bureau of Labor Statistics.

tages associated with the subcategories,<sup>3</sup> namely, that the amount of successful recall decreases as the number of cues increases.<sup>4</sup> Furthermore, the instrument looked intimidating: it was 66 pages long (compared with the 23 pages in the 1980 CED); and although the physical size of the Current Diary was smaller than the 1980 version (14" × 8", compared with 17" × 11"), it was still large and bulky and had a landscape layout.

In response to these factors, a joint BLS and U.S. Census Bureau<sup>5</sup> team was chartered in 2000 to design a more user-friendly diary that would encourage greater participation by simplifying the diary-keeping task, yet still solicit the reporting detail required.<sup>6</sup> The team identified nine main themes from participants' recommendations. One prominent theme was a reaction to the subcategory cues. Participants recommended that the recording task be reduced to the minimum number of major categories and not include a secondary classification task required by subcategories. The team used these themes as a basis for designing a more user-friendly diary.

### The Redesigned Diary

The Redesigned Diary has four broad categories with no subcategories. To simplify the appearance of the recording pages, specific cues were removed and placed on a flap attached to the front cover. The Redesigned Diary has an 8 1/2" × 11" portrait layout with 44 pages.

<sup>3</sup> Vitrano, F.A., et al., "Cognitive Issues and Reporting Level Patterns from the CE Diary Operational Test," in *Proceedings of the Section on Survey Research Methods*. Washington DC: American Statistical Association, pp. 262–267, 1988.

<sup>4</sup> Roediger, H. L., "Inhibiting Effects of Recall," *Memory and Cognition*, pp. 261–269, 1974.

<sup>5</sup> BLS contracts with the U.S. Census Bureau to implement the Consumer Expenditure Diary Survey in the field.

<sup>6</sup> Davis, J., et al., "What Does It Really Mean to Be User-Friendly when Designing an Expenditure Diary?" paper presented at the annual meeting of the American Association of Public Opinion Research (2002). See also Davis, J., et al. "Creating a User-Friendly Expenditure Diary," *Consumer Expenditure Survey Anthology*, Report 967, pp. 3–17, Sept. 2003.

The Redesigned Diary was field-tested from September to December of 2002. Results from the test were mixed. The new user-friendly design was overwhelmingly preferred and supported by Census field staff. Moreover, the field-test data indicated that the Redesigned Diary was comparable to the Current Diary in response rates and overall levels of reported expenditures.

However, the data also indicated that respondents failed to record expenditures at a sufficient level of detail, causing an increase in allocation rates.<sup>7</sup> This loss of detail was attributed to the elimination of the specific cues on the recording pages. Consequently, further research into the addition of cues on those pages in the Redesigned Diary was recommended.

### Scope and methodology

The purpose of the cognitive study that was recommended was to test whether adding specific cues on the recording pages would alleviate the problem of respondents failing to record at a sufficient level of detail, while maintaining the user-friendly layout of the Redesigned Diary. To accomplish this task, alternative means of adding cues to the recording pages of the Redesigned Diary were evaluated.

#### A. Test diaries

Three formats of the Redesigned Diary were tested in the cognitive study:

1. *The No-Cues Diary*. This diary was similar to the one used in the 2002 field test and had no cues on the recording pages. (See exhibit 1.)
2. *The Margin-Cues Diary*. This diary listed cues along the left side of the recording pages. (See exhibit 2.)

<sup>7</sup> Figueroa, E., et al., "Is a User-Friendly Diary More Effective? Findings from a Field Test," paper presented at the annual meeting of the American Statistical Association, 2003. Although allocations are often used to account for item nonresponse, in the diary the term refers to an expenditure that does not identify individual items at the required level of detail (e.g., a respondent reports "groceries, \$150," rather than the specific items purchased). This type of entry requires additional processing to assign the aggregate expenditure to target items.

3. *The Header-Cues Diary*. This diary listed cues along the top of the recording pages. (See exhibit 3.)

*Selection of cues:* Because space on the recording pages was limited, the number of cues had to be minimal, making the selection of cues an important task. The cues were selected on the basis of four criteria:

1. *Analysis of the 2002 field-test data*. A comparison was made between the mean expenditures of the Redesigned Diary and the Current Diary. Because research has shown that cues improve the reporting of an item, items for which reported expenditures were significantly lower in the Redesigned Diary compared with the Current Diary were identified, and a subset of those items was selected as cues. Examples include white bread, oranges, and whole chicken.
2. *Items commonly reported without adequate detail*. Certain items are commonly entered into the CED with insufficient detail, requiring data adjustment. For example, entries of "gas" must be allocated to either gasoline or utility gas. Similarly, entries of "books" must be allocated to either schoolbooks or other books. To encourage more specific reporting of items, cues such as "gasoline," "utility gas bill," "textbooks," and "cookbook" were selected.
3. *Problems identified in the two food categories "Food and Drinks Away from Home" and "Food and Drinks for Home Consumption."*
  - *Drinks without a meal*. Team members were concerned that linking "Food and Drinks" together in the titles would discourage the reporting of drinks without a meal. To encourage such entries, cues such as "beer at happy hour" and "soda from vending machine" were selected.

- *Delivery and takeout meals.* Due to the wording of these two food entries, the reporting of items such as pizza delivery and Chinese takeout is confusing to respondents. Both entries should be reported as “Food Away from Home,” but are often entered as “Food for Home Consumption,” because respondents usually consume these foods in the home. To encourage entering these items in the correct section, cues of “pizza delivery,” “Chinese takeout,” and “carryout lunch” were placed on the “Food Away from Home” recording pages.

4. *A balanced representation of items.* One specific cue from each subcategory in the Current Diary was selected:

- “cigarettes” from “Tobacco and Smoking Supplies”
- “prescription drugs” from “Medicines, Medical Supplies, and Services”

An effort was made to emphasize items that are currently known to be underreported.

*Specificity of the cues:* Cues were restricted to specific items (e.g., skim milk) that do not require allocation because they contain sufficient detail. Cues for items requiring allocation (e.g., milk) were excluded from consideration. It was thought that cuing for sufficient detail would instruct respondents to record expenditures with similar specificity. A BLS study of the CED in the early 1990s noted that cued items have higher reporting rates when the cues are specific (e.g., chuck roast vs. beef).<sup>8</sup>

*Order of the cues:* Most cues are grouped with similar items (e.g., wine, beer, and liquor) to emphasize the variety and specificity desired. Pairs of

<sup>8</sup> Dippo, C.S., and Norwood, J.L., “A Review of Research at the Bureau of Labor Statistics,” in *Questions about Questions*, ed. J.M. Tanur: Russell Sage Foundation, NY, pp. 271–290, 1992.

cues selected to encourage more specific reporting of items were placed next to one another to illustrate the importance of distinguishing similar items (e.g., “gasoline” and “utility gas bill” were placed next to each other to avoid an entry such as “gas”).

## B. Participants

Participants for this study were recruited from a database maintained by the BLS Office of Survey Methods Research and through an advertisement placed in a local newspaper. Sixty-one individuals were recruited through these methods, together with an additional 5 BLS employees, for a total of 66 participants, all from the Washington, DC, area. Thirty-four participants were women, and while no information on race or ethnicity was collected, observationally, there appeared to be a balance among African-Americans, Caucasians, and Hispanics. The average age of the participants was 42, with subjects ranging from 17 to 77 years. The completed education level of the participants ranged from 11th grade to doctorate. The average education level of the participants was 16 years, equivalent to a college degree. About one-third of the participants ( $n = 24$ ) were employed part time, one-third ( $n=19$ ) full time, and the remaining participants were unemployed ( $n = 9$ ), self-employed ( $n = 6$ ), and retired ( $n = 3$ ). The average self-reported income was \$37,000. The median income was \$31,000, with reports ranging from \$800 to \$100,000.

Twenty-four participants were single, 19 were married, 13 were divorced, and 3 were widowed. Of those from whom data were collected, half had children ( $n = 28$ ) and half did not. The median number of children per participant was one, and the ages of the children ranged from 1 to 42 years, with the average being 22 years.

## C. Study design

1. *The recall task.* Each participant was provided a diary and asked to enter all of his or her household’s expenses for the previous week. Since respondents in the field would

be able to use receipts, checkbooks, and other records to help them complete the diary, any participant who had such records available was allowed to use them. Diaries were distributed among three groups of participants, with 21 participants receiving the No-Cues Diary, 23 receiving the Margin-Cues Diary, and 20 receiving the Header-Cues Diary.<sup>9</sup>

2. *The recognition task.* After completing the diary-recall task, participants were given a comprehensive list of commonly purchased and frequently forgotten items and were asked to check off all items, including those they had recorded in the diary, that they or anyone in their household had purchased during the past week.

*Recall versus recognition.* Research on memory has revealed that, when given a recall task and a recognition task, participants are able to remember more items with the recognition task<sup>10</sup> (Standing et al., 1970, and Sternberg, 1999). Therefore, it was thought that participants in this study would identify more of the purchases made by their households when using the recognition list than had been reported by completing the diary (a pure recall task). The items that were checked on the recognition list, but not recorded in the diary during the recall task, would provide some measure of underreporting (how many items respondents forgot when completing the pure recall task of recording in the diary).

Results from the study showed that the average number of unique recognition items reported by participants was greater than the average number of unique diary (or recall) items reported. There was no significant difference

<sup>9</sup> The original sample contained 66 diaries. Due to data problems, 2 diaries from the group receiving the Header-Cues Diary were eliminated from the analysis.

<sup>10</sup> Standing, L., et al., “Perception and memory for pictures: Single-trial learning of 2500 visual stimuli,” *Psychonomic Science*, 19, pp. 73–74, 1970. Also Sternberg, R.J., *Cognitive Psychology*, 2nd edition. Harcourt Brace College Publishers, New York, 1999.

across the diaries in the percentage of respondents underreporting. (See table 1.)

*3. Followup questionnaire and debriefing.* After completing both the recall and recognition tasks, participants were given a questionnaire about their experience with the diary. There was a separate questionnaire for each diary format. The questions were designed to identify the various features of the cues, including the location, format, and the actual cues that were selected.

Finally, before concluding the session, each participant received a 5-minute debriefing in which he or she had the opportunity to provide further comments.

## Findings

### A. Qualitative findings

#### *Observational findings*

Because the goal of the study was to examine the impact of adding cues to the recording pages of the Redesigned Diary, it was important to identify any problems participants had that appeared to be a direct result of the cues. This goal was achieved by observing the participants and noting the questions they asked as they completed the tasks and then reviewing each diary for errors.

One of the main problems found was with the Margin-Cues Diary. A few participants circled the margin cues instead of entering the description in the space provided. This problem may have stemmed from the visual layout of the vertically formatted cues in the Margin-Cues Diary, compared with the horizontally formatted cues in the Header-Cues Diary. Apparently, when cues are listed vertically, some participants are more likely to view them as a comprehensive list of expenses to circle than when they are listed horizontally.

When recalling their purchases, some participants asked what they should do if they didn't buy something that was listed. Others asked what they should do if they purchased something that was *not* listed. These questions

suggested that some participants did not fully understand the purpose of the cues and thought of them as comprehensive lists from which they had to choose. This type of confusion could lead to overreporting of cued items and underreporting of noncued items.

#### *Findings from the followup questionnaire*

Because the cues were designed to help participants recall items they may have purchased, one question asked whether the participants used the sample items (on the flap of the No-Cues Diary, listed along the side of the recording pages in the Margin-Cues Diary, and listed along the top of the recording page in the Header-Cues Diary) to help them remember their purchases. Among participants using the No-Cues Diary, 50 percent reported that they found the sample items helpful in remembering purchases. Almost 70 percent of participants using the Margin-Cues Diary said the cues along the side of the recording pages were helpful, and 86 percent of respondents using the Header-Cues Diary reported that the sample cues along the top of the recording pages were helpful.

In addition, the majority of the participants indicated that cues were helpful for determining which purchases to record, how to record purchases, and in which section to record purchases.

#### *Findings from the debriefing*

The debriefing questions provided additional feedback about the participants' experience with the diary, so any comments they made regarding the cues were seen as particularly useful. Many participants stated that the examples were very helpful. Although the term "examples" may have been used to denote examples anywhere in the diary, some participants specifically referred to the cues listed along the top of the page or cues along the side of the recording page.

### B. Quantitative findings

A one-way analysis of variance

(ANOVA) was performed to test differences between the three diary forms on the following factors:<sup>11</sup>

- Overall level of expenditures
- Total number of items reported
- Number of unique diary items (items recorded only with the recall task)
- Number of unique recognition items (items checked only with the recognition task)
- Percent of reported items requiring allocation
- Percent of items that matched the cues verbatim

#### *Comparing diary items*

The only significant difference found among the three types of diaries was the average proportion of items matching the cues printed on the recording pages verbatim. (See table 1.) Compared with the No-Cues diary, the Margin-Cues Diary and the Header-Cues Diary both had more than twice the proportion of items matching the cues (7 percent, as opposed to 19 and 20 percent, respectively). This difference suggests that the participants were looking at the cues on the pages. However, there was no significant difference between the Margin-Cues Diary and the Header-Cues Diary (19.1 percent and 19.7 percent, respectively).

No significant differences were found on any of the other variables measured, including number of unique diary items recalled, number of unique recognition items reported, and percentage of items requiring allocation due to inadequate detail in reporting.

#### *Comparing diary expenditures*

No significant differences in expenditures were found among the three diaries.

<sup>11</sup> Where the data met the assumptions required for ANOVA. When the data violated these assumptions, the nonparametric Wilcoxon Rank Sum test was performed.

## Conclusion

The purpose of the cognitive study was to test whether adding specific cues on the recording pages would alleviate the problem of respondents failing to record at a sufficient level of detail. Although there was no significant finding that the recording page cues resulted in more detailed reporting, the qualitative results provided evidence that respondents used the cues and found them helpful both in recalling purchases and in remembering how to record purchases. Quantitative analysis showed no significant differences in the number of entries among the diaries, but there were significantly more entries in the Header-Cues and Margin-Cues Diaries that matched the recording page cues than did entries in the No-Cues Diary, suggesting that respondents noticed and used the recording page cues.

Given both the qualitative and quan-

titative evidence that respondents found the recording page cues helpful and that the vertical format of the cues in the Margin-Cues Diary might be problematic, the team recommended that the Header-Cues Diary be implemented with two modifications, to emphasize that the cues are only examples and not a comprehensive list. This change would help to decrease the potential for overreporting of cued items and underreporting of noncued items. The modifications are as follows:

1. The word “Examples” is to be added in a larger and different-color font next to the lists of cues.
2. The arrow that was used to instruct respondents to look in a different section for “Additional Examples” is to be moved to a more prominent location closer to the list of cues, to encourage respondents to utilize a

more extensive list.

The Modified Header-Cues Diary (exhibit 4) will be implemented in January 2005.

In 1980, the CED had five broad categories, which were then divided into 42 detailed subcategories in 1991. In 2005, the subcategories will be removed, leaving four broad categories. In terms of the specific cues it contains, the CED went from 76 in 1980 to 305 in 1991. The 2005 diary has 89 specific cues.

Will the combination of a user-friendly layout and a decreased number of specific cues on the recording pages have a positive impact on response rates and quality of the data? Did BLS strike the right balance between too many cues and too few? These questions will be answered after data are collected with the Redesign Diary in 2005. ■

Table 1. Comparing the sample means of the three diaries

Characteristic	No-Cues Diary	Margin-Cues Diary	Header-Cues Diary
Sample size (number of diaries).....	21	23	20
Number of entries in diary .....	42	43	42
Part 1. Food away from home .....	7	10	9
Part 2. Food for home consumption .....	21	17	16
Part 3. Clothing, shoes, jewelry, and accessories .....	3	3	4
Part 4. All other products, services, and expenses .....	11	12	13
Number of unique diary items .....	27	26	27
Number of unique recognition items .....	35	47	44
Percent of items reported need allocation <sup>1</sup> .....	5.6	6.2	5.3
Percent underreporting <sup>2</sup> .....	28.2	37.7	37.4
Percent of cued items reported .....	51.0	62.0	56.2
Percent of items that matched the cues(verbatim) <sup>3</sup> .....	7.3	19.1	19.7
Total expenditure .....	1,317	893	1,100
Part 1. Food away from home .....	39	61	96
Part 2. Food for home consumption .....	67	82	58
Part 3. Clothing, shoes, jewelry, and accessories .....	83	79	71
Part 4. All other products, services, and expenses .....	1,128	672	875

<sup>1</sup> Although “allocation” is often used to account for item nonresponse, in the diary, the term refers to an expenditure that does not identify individual items at the required level of detail. (For example, a respondent reports “groceries \$150,” rather than the specific items purchased.) This type of entry requires additional

processing to assign the aggregate expenditure to target items.

<sup>2</sup> “Underreporting” refers to the items that were checked on the recognition list, but not recorded in the diary during the recall task.

<sup>3</sup> Significant difference at  $p = 0.05$ .