

PART-SET CUING IN DIARY SURVEYS

Adriana R. Silberstein, Bureau of Labor Statistics
2 Massachusetts Ave., NE, Rm. 3650, Washington, DC 20212

KEY WORDS: Recall, Cues, Expenditures

ABSTRACT

Two diary instruments used cuing to obtain more accurate responses in the Consumer Expenditure Diary Survey. Cuing consisted of providing categories and examples of household expenses on the diary pages. One diary had a limited number of categories and examples; the second diary included a large number of examples presented within many categories. The two diaries' responses were found to differ, overall and by type of item. As in word retrieval experiments, survey responses may be facilitated as well as inhibited by cues.

1. INTRODUCTION

Diaries are used to record information at the time or soon after an event occurs. Diary keeping, by its very nature, is subject to time lapse - the time and distance between experiencing the activity and logging it. It is this time lapse that exposes the diary keeper to recall. Respondents to the Diary component of the US Consumer Expenditure Survey (CE) record family expenses for two consecutive weeks, with emphasis on grocery items and other small expenses. The end of each day is the most frequent time for recording expenses on the diary, but many respondents make entries more irregularly. Some form of recall is often required even when receipts are available because items may not be clearly identified.

Expenditure diary formats vary among countries, but all include some type of cuing to guide respondents in the diary keeping task. The CE Diary instrument is very structured, with spaces on each page reserved for designated expense categories, referred to here as *sets*. The instrument's overall organization has not changed since the 1972/73 Diary (a journal with seven diary days), but the layout has changed and the number of examples (*cues*) appearing on the diary pages for each day has increased. The 1972/73 diary had one page for food items divided into 5 sets and one page for nonfood items divided into 11 sets; few examples were given. The survey introduced in 1980 on a continuing basis used a

diary form with one and one quarter pages for food items listed in 7 sets and three fourths of a page for nonfood items listed in 11 sets; 92 examples were included. A longer form was introduced in 1991 for a subset of respondents; the other subset used the 1980 form. The new diary had three pages for food items divided into 17 sets and three pages for nonfood items divided into 18 sets; a total of 300 examples were included. The present study investigated effects of cuing on responses to the 1991 diaries, utilizing the application of two cuing methods in the same survey environment.

The importance of cuing has long been recognized in questionnaire design (Bradburn and Sudman, 1991). Cues interact with all four stages of the survey response process - comprehension, retrieval, judgment, and communication (Dippo and Norwood, 1992). Cues are typically in the form of examples, visual aids, or screeners. Other techniques have been designed to encourage respondents to generate personal cues during an interview (Means, et al., 1989). Research is usually carried out to develop appropriate cuing methods, especially in large repetitive surveys (see, for instance, the studies reported by Martin, et al., 1986, and by Lessler, et al., 1989), but specific results on the success of cues are seldom reported after implementation. One reason may be that these effects are not easily isolated from other response factors once cues are imbedded in a questionnaire.

The values and limitations of using cues have been studied in laboratory memory research (Nickerson, 1984; Roediger, 1982). Cuing tends to facilitate information retrieval, either by the direct recall of items mentioned or through mental associations made to other items. Cuing may also inhibit information retrieval, a finding that is inconsistent with the idea that mental associations increase recall. Item competition and interference are two explanations given for this finding. The added burden of having to mentally process the cue list is another explanation; this possibly causes diversion from the recall task.

The number of cues and how cues are presented affect the results. A large number of cues may cause overload and increase inhibition

effects (Lynch and Srull, 1982). Cues can be given *piecemeal*, in ordered lists or randomly arranged, or can be presented in organized *sets* with descriptive headings. Grouping cues into categories is regarded as more successful than presenting them piecemeal. Fiske and Pavelchak (1986) devoted an article to this issue in the context of social cognition and social affect, providing the explanation that responses are processed at the top of the categories rather than through an evaluation of individual items. This efficient use of cues may revert to the piecemeal mode if set titles and cues are inconsistent. These concepts can be easily extended to information retrieval tasks, such as reporting expenditures. Findings on consumer preferences lend support to category-based recall (Sujan, 1985). The specificity of cues within categories may influence results. General statements (e.g., camping equipment) may facilitate the recall of entire classes of items if respondents are able to "lock into" their meaning (Lynch and Srull, 1982). Specific cues (e.g., tents) give the advantage of direct item recognition.

Cues are given to stimulate recall and reporting in the CE Diary, and they necessarily include only a fraction of items. The term *part-set cuing* (Nickerson, 1984) is applicable to describe this response process. It is more general than the often used term *part-list cuing*, since it includes situations where cues are lists of unrelated items and where they are items from well known sets or categories (e.g., Clothing). Cues are given as further explanations of clearly understood titles, in the latter case. Most sets in the new diary are coherent groupings, but some sets are collections of disparate items (e.g., Housewares and Small Household Appliances), and a few pertain to residual items (e.g., Other Shoes).

2. CUING STUDY DESIGN

2.1 The 1991 Diaries

The background of the 1991 Diaries is found in the 1985 Diary Operational Test (Tucker, 1992) and previous analytical studies. Comparisons of data from the 1972/73 and 1980/81 surveys indicated positive effects of cuing in the context of a general reporting decline (Silberstein, 1983). Consequently, more cues were included in two experimental diary formats for the 1985 test: a *Specific Diary* with preprinted detailed expense categories (to be simply checked off) and a

Nonspecific Diary showing category subdivisions and examples. Increased cues were most beneficial in nonfood categories of the specific diary (Vitran, et al., 1988). The nonspecific format was more successful overall, and became the foundation for the new diary form.

The purpose of the 1991 test was to evaluate improvements to the national estimates in a production mode: 20% of the sample received the new form (Diary B), 80% received the 1980 form (Diary A). Both diaries were kept for two consecutive weeks by respondents solicited throughout the year across 101 PSUs. The new form was adopted for the whole sample beginning in 1993, because it provided better estimates. Overall findings can be found in Tucker (1993).

Following is a description of similarities and differences in the two instruments. Both diaries contained five parts for each diary day: 1. Food for home consumption, 2. Food and beverages purchased as gifts, 3. Food away from home, 4. Apparel (clothing, shoes, and jewelry), and 5. Other expenses. Diary B had three times as many cues as Diary A, but the allocation of cues to food and nonfood was the same, 1/3 and 2/3, respectively. The diary length and its layout were major differences between the two diaries. The greater length of Diary B was a natural outcome of using a more specific categorization. Category titles and examples were printed horizontally, creating boundaries for each set and strongly suggesting that entries be made according to the organization presented; respondents had to search through six pages for the right place to enter expenses. The greater burden of Diary B was contrasted with insufficient space in Diary A, especially for nonfood items. The page layout in Diary A gave the impression of a free-form instrument in which expenses could be entered anywhere on the appropriate page, since category titles and examples appeared on the side of the form.

2.2 Method of Analysis

The configuration of Diary B was used to summarize data from both diaries, since more detailed comparisons resulted and design differences were in one direction (Diary A had less structure and cues). The *Census Item* classification was chosen instead of the UCC classification because it was more detailed and closer to actual diary examples. Each code was classified according to whether it was cued, and if

so, in which diary (and set) it appeared. (The appropriate set was identified also for noncue items.) The 522 specific codes were thus subdivided into the following groups:

DIARY B ONLY (157)	BOTH DIARIES (68)
DIARY A ONLY (14)	NEITHER DIARY (283)

A separate coding scheme was used for diary entries that were not specific (e.g., *paper products*) or combined several items (e.g., *groceries*). A decline in these types of entries signified improved reporting of individual expenses.

Responses from diaries completed during the diary weeks were analyzed (5797 diaries for A, and 1388 for B). These diaries represented 68 percent of all 1991 diaries, counting each week separately. The remaining responses were diaries completed during pickup with partial or total recall.

Significance testing of differences between mean vectors was carried out with simultaneous two-tailed T^2 tests. Pooled estimates of the variances from the two diaries were computed with the method of balanced repeated replication (44 replicates). Reporting rates (r) were used for detailed level comparisons; r is the percentage of diaries with a nonzero expenditure for a specific item or set. Greater values of means and reporting rates are usually interpreted as better estimates, since expenses tend to be underreported (Gieseman, 1987).

One-tailed tests were employed for detailed item level comparisons, because overall findings indicated the B Diary obtained higher estimates in general. These tests evaluated the relative differences between A and B for cue items and noncue items. The Rank Sum Test (Wilcoxon) was used. The variable ranked was the difference in reporting rates ($r_B - r_A$). The main alternative hypothesis was that greater differences between the two diaries would be found for codes with cues in B. Tests were conducted by size class of reporting rates (using r_A) and by commodity. The first test controlled for the variation in reporting rates; this may be very small for items infrequently purchased in a week's time since rates are often close to zero.

Diary B represented the experimental condition of greater cognitive stimuli; Diary A provided the control condition. The range and consistency of sets and the specificity and prominence of cues were considered in the formulation of the following research hypotheses.

$H_0 - \Delta = 0$: lack of cuing effects

H_1 - Positive effects of cuing
Cuing provides higher reports

H_{11} Direct effects

Expenses are remembered by cue recognition

H_{12} Indirect effects

Expenses are remembered through associations

H_2 - Inhibiting effects of cues

Forgetting can be cue induced

H_3 - Set/cue prominence

Prominently placed sets/cues have more power

H_4 - Overload effect

Large number of cues have diminishing returns

H_5 - Consistent categorization

Provides cue strength

H_6 - Specificity of cues

No prediction

3. FINDINGS

The expenditure means from Diary B were at least 10 percent higher in most categories or sets than means from Diary A. These positive results did not emerge for food away from home and some nonfood categories. Overall, nonfood expenses obtained greater gains than food expenses. Differences between the two diaries were often wider for items listed as cues in *B Only*. The increase for these items in food at home, for instance, was 23 percent compared to the 14 percent overall increase. (Table 1)

Comparisons of reporting rates provided similar relationships to the ones found for the means. The average rank scores were generally higher for items in *B Only*, suggesting greater changes for these items compared to noncue items (Table 2). This outcome was consistent by reporting rate size, except for items with rates lower than 1% (shown only for nonfood). Significant effects were found for three of the five commodity groups analyzed - Home Food, Personal Care/Housekeeping, and Miscellaneous. (Food Away was not tested since most of the cues were in both diaries.) These findings provided support to the notion of direct cuing effects (H_{11}), and the null hypothesis was rejected.

The hypothesis of indirect cuing effects (H_{12}) postulated that cuing facilitates the respondents' ability to extrapolate, or *carry over*, from items shown to items not shown. The set of Entertainment and Sports expenses, for instance, showed similar increments in reporting rates for cue items and noncue items. This provided a

glimpse of the effects. Increments in reporting might also have been obtained for other expenses connected with a given (e.g., sporting) event, but such information could not be reconstructed from the data analyzed. Mental associations do not occur solely within a prescribed categorization, but involve related items in separate parts of the instrument - this information is lost as entries are made in the diary. The way in which items are remembered by respondents could not be discerned even if individual diaries were examined, due to the structured instrument.

The hypothesis of inhibition effects (H_2) implied that remembering certain items from a list tends to be reduced in the presence of cues. Conclusions from memory research point out that scanning a cue list may sometimes impede even simple recall tasks. Greater effort is necessary in responding to surveys. After scanning the cue list, respondents have to decide whether events listed have occurred and determine whether events not listed should be reported. Another facet of surveys is the nonrandom choice of cues. There may be a tendency to select items at the extremes, that is, some quite typical (and recognizable by respondents) and some infrequent (for which respondents need to be reminded).

These issues suggested a test for a special set of items, rather than for all items. The following criteria were used to select the items: a. specific and well identified items (relatively frequent and popular expenses), b. with sufficiently specific corresponding codes, and c. of similar importance (according to reporting rate level). The expectation was that similar items would have a similar chance of being reported. Results of this test provided support to the hypothesis of inhibition effects. Only 4 of the 20 noncue items had greater reporting rates in B, 6 items were nearly equal in A and B, and 10 items had lower rates in B. The group of 20 cue items had opposite results: 15 items had greater reporting rates in B, and 5 items in B were nearly equal or lower than in A. The group of items not presented as cues included popular items such as pizza, crackers, and greeting cards; these were reported less in Diary B, which generally had reporting increases. (Table 3)

A large number of sets and cues can be considered burdensome to respondents and eventually less effective. Items cued in both instruments might be expected to yield relatively higher reporting in Diary A, compared to other items, because of greater prominence than in

Diary B; this consideration would make the null hypothesis more likely. A test (H_3) comparing cue items in both instruments to noncue items showed no significance for Food, as expected, but some significance for Nonfood. (Table 4)

Another aspect of cue prominence is the length of the cue list within a category. Long lists are not read in their entirety and words are missed even by attentive readers. This issue is related to the categorization used. Most food categories in B were narrowly defined and included no more than seven cues, which is an upper limit suggested in the literature. Nonfood categories included a wide variety of items and a minimum of 12 cues. Residual categories (*other*), in particular, had more cues because needed more explanations. Best results were obtained for sets where cues were consistent with the domain projected by the title. An additional result from this format, especially for food, was an increased rate in reporting individual items.

The aspect of specificity of cues was investigated by comparing specific and nonspecific examples presented within Diary B. A test (H_6) of these items showed mixed results. Greater specificity produced some (positive) differences for more frequently purchased food items; no significant results were found for nonfood items.

4. CONCLUSIONS

This study confirmed general assumptions that cues facilitate information retrieval. Comparisons of detailed statistics from two diary formats indicated higher reporting levels for the more detailed Diary B, with relatively greater reporting for items mentioned as cues. Symptoms of inhibition aspects of cuing were also suggested by the data, since certain noncue items had relatively lower reporting. The potential interaction of cuing with the type of categorization was also discerned: Diary B was less effective when categories contained too many types of expenses.

Results of format changes were consistent across the categories of Food at Home, and reporting with greater detail was an added improvement brought about by increased cuing. More complex results were obtained in nonfood sections. Compared to food, relatively greater increments were obtained in some nonfood categories, as, for instance, Home Furnishings; although some of these changes could be traced to the cue list, overall, differences appeared to be

influenced to a greater extent by increased space and visibility of nonfood sections in Diary B. There were no improvements in terms of reporting detail for nonfood categories, and some categories did not experience a reporting increase. The test of cuing effects was not significant for expenses not often recorded over a week's time. This apparent *immunity* to cues may be explained by the fact that many infrequent expenses have large dollar values and tend to be remembered well.

This study has implications for diary design because more accurate reporting can be stimulated by the effective use of cuing. Biasing effects of cuing, however, should be analyzed continuously. Cognitive research should address the connection between items reported and items not reported, using alternative displays of cues. Recall is generally affected by recency and prominence of events. An issue specific to cues is the level of detail at which inhibiting effects operate: whether individual items or whole classes may be excluded. A related issue is whether cues tend to block the retrieval of items of the same *power*, in terms of the importance and frequency of purchases.

Greater design changes appear needed to increase reporting for some commodities. The new diary improved the reporting of Apparel only marginally and not consistently by category. Like the previous diary, Diary B did not remind (cue) the diary keeper to check about purchases by other family members, and did not provide cues about holiday shopping or gifts in general. No improvements were found for Food Away from Home. The home-style diary method, common to both diaries, has intrinsic limitations when reporting food items consumed away from home. Examples such as *lunch or dinner* may not cue reporting about special events, and, again, no reminders were given about expenses of individual family members.

Additional cuing methods should be researched. These may extend beyond the words included as examples within each set. Cuing techniques that could provide larger associations than within set associations may be more effective in some cases. What constitutes *related items* may be an individual matter; they may be items purchased together (for a meal, a trip, etc.) or items purchased for the same cluster (person, room, etc.). Diary designs allowing respondents to recall and list some expenses according to personal shopping patterns or to special shopping clusters should be considered.

REFERENCES

- BRADBURN, N.M., and SUDMAN, S. (1991), "The Current Status of questionnaire Research," in *Measurement Errors in Surveys*, Eds. P.P.Biemer, R.M.Groves, L.E.Lyberg, N.A.Mathiowetz, and S.Sudman, Wiley, N.Y., 29-40.
- DIPPO, C.S., and NORWOOD, J.L. (1992), "A Review of Research at the Bureau of Labor Statistics," in *Questions About Questions*, Ed. J.M. Tanur, Russel Sage Foundation, N.Y., 271-290.
- FISKE, S.T., and PAVELCHAK M.A. (1986), "Category-Based Versus Piecemeal-Based Affective Responses," in *Handbook of Motivation and Cognition*, Eds. R.M.Sorrentino and E.T.Higgings, The Guilford Press, N.Y., 167-203.
- GIESEMAN, R.W. (1987), "The Consumer Expenditure Survey: Quality Control by Comparative Analysis," *Monthly Labor Review*, 3, 8-14.
- LESSLER, J., TOURANGEAU, R., and SALTER, W. (1989), "Questionnaire Design in Cognitive Research Laboratory," *National Center for Health Statistics, Vital and Health Statistics*, 6 (1).
- LYNCH, J.G.Jr., and SRULL, T.K. (1982), "Memory and Attentional Factors in Consumer Choice: Concepts and Research Methods," *Journal of Consumer Research*, 9, 18-37.
- MARTIN, E., GROVES, R., MATLIN, J., and MILLER, C. (1986), "Report on the Development of Alternative Screening procedures for the National Crime Survey," Bureau of Social Science Research, Washington, DC.
- MEANS, B., NIGAM, A., ZARROW, M., LOFTUS, E.F. and DONALDSON, M.S. (1989), "Autobiographical Memory for Health-Related Events," *National Center for Health Statistics, Vital and Health Statistics*, 6 (2).
- NICKERSON, R.S. (1984), "Retrieval Inhibition from Part-Set Cuing: A Persistent Enigma in Memory Research," *Memory and Cognition*, 12 (6), 531-552.
- ROEDIGER, H.L.III, and NEELY J.H. (1982), "Retrieval Blocks in Episodic and Semantic Memory," *Canadian Journal of Psychology*, 36 (2), 213-242.
- SILBERSTEIN, A.R. (1983), "Diary Reporting Levels: 1972/73 and 1980/81," Memorandum to C.A.Jacobs, December 16.
- SUJAN, M. (1985) "Consumer Knowledge: Effects on Evaluation Strategies Mediating Consumer Judgments," *Journal of Consumer Research*, 12, 1-16.
- TUCKER, C. (1992), "The Estimation of Instrument Effects on Data Quality in the Consumer Expenditure Diary Survey," *Journal of Official Statistics*, 8 (1), 41-61.
- TUCKER, C. (1993), "The Effects of Format Changes on Reporting in the 1991 Consumer Expenditure Diary Survey," *Proceedings of the Section on Survey Research Methods*, American Statistical Association.
- VITRANO, F.A., HUBBLE, D.L., and VACCA, E.A. (1988), "Cognitive Issues and Reporting Level Patterns from the CE Diary Operational Test," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 262-267.

Note - The following types of expenses were excluded from all tables because not relevant to the study: Smoking expenses, gasoline, car and transportation expenses, medical, housing, and school expenses.

Table 1. Comparison of Expenditure Means

Diary:	ALL ITEMS		Items cued in B Only	
	A	B	A	B
<i>Weekly Exp. Means:</i>				
Home Food	\$56	\$64	\$24	\$29
Food Away	29	30	1	1
Apparel	35	38	15	16
Other	87	121	21	32
<i>% Increase (B/A):</i>				
Home Food		14 *		23 *
Food Away		2		3
Apparel		9		13
Other		40 *		54 *
(* 0.05-level, simultaneous two-tailed test)				

Table 2. Test of Direct Cuing Effects ¹

	Items cued in:			
	B Only		Neither	
	n	W ₁ /n	m	W ₂ /m
<i>1. Reporting Rate Size</i>				
Food items				
Less than 5%	17	29	29	20 **
5% - 10%	19	27	23	17 **
10% and over	26	21	9	10 **
Nonfood items				
Less than 1%	31	70	109	71
1% - 3%	18	25	22	17 **
3% and over	17	22	17	13 **
<i>2. Commodity Group</i>				
Home Food	60	78	59	42 **
Food Away	2	-	2	-
Apparel	12	15	13	12
Personal Care & Housekeeping	7	13	12	8 *
Home Furnishings	21	43	52	35
Miscellaneous	26	57	71	46 *

Table 3. Test of Inhibition Effects ¹

	Items cued in:			
	B Only or Both		Neither	
	n	W ₁ /n	m	W ₂ /m
Selected items ²	20	26	20	15 **

Table 4. Test of Cuing Prominence ¹

	Items cued in:			
	BOTH		Neither	
	n	W ₁ /n	m	W ₂ /m
<i>Reporting Rate Size</i>				
Food items				
Less than 10%	7	32	52	30
10% and over	18	15	9	12
Nonfood items				
Less than 1%	1	-	109	-
1% - 3%	8	22	22	13 **
3% and over	21	22	17	16 *

¹ Rank Sum Test (Wilcoxon) $W_1 = \sum R_j$

R_j ranks of reporting rate differences for B Only or for Both; W/n average scores.

One-tailed test: $H_0: \Delta = 0$

Reject H_0 if $W_1 \geq W_{\alpha, m, n}$

(** $p < 0.05$, * $p < 0.10$)

² Selected items:

B Only or Both: cakes/cupcakes, cookies, round steak, butter, apples, oranges, sugar, corn, nondairy substitutes, peanut butter, baby foods, nuts, shorts, shaving needs, deodorants, paper towels, toilet tissue, hand tools, toys, stationary.

Neither: crackers, muffins, chuck roast, yogurt, grapefruits, grapes, cake frosting, carrots, olives/pickles/relishes, chili, mexican foods, pizza, sweaters, face/body powder, nail preparation, paper napkins, facial tissues, infants equipment, games, greeting cards.