

Improving Purchase Recollection

Monica Dashen¹

Bureau of Labor Statistics, Room 4915, 2 Massachusetts Ave. NE
Washington DC 20212

Key words: Improving Data Quality, Response Error, Survey Methods

I. INTRODUCTION

Retrospective reports about personal events are often collected by a survey long after the event in question. It is well known that people under-report events (the tendency to report fewer events than actually occurred) and over-report events (the tendency to report more events than actually occurred) over time. A large body of research has demonstrated that under- and over-reporting can affect the data quality (Sudman & Bradburn, 1973). Yet, little is known about how to improve the accuracy of improve recollection. Accordingly, the present work explored how well survey respondents are able to remember frequently-purchased items and examined ways to improve purchase recollection.

There are two distinct causes of under- and over-reporting. Under-reporting is generally thought to be due to forgetting particular events. The debate over the cause of forgetting is an old one. Some researchers assert simply that memory traces decay with time and the memory trace is lost forever (e.g., Ebbinghaus, 1885/1964). Other researchers assert that we have not really lost the old information. Rather, it merely becomes inaccessible for the time being. The most popular analysis of inaccessibility is the interference theory (see Crowder, 1976, for a review). Interference proponents suggest that forgetting occurs because other material interferes with the retention of the relevant information. Newly-learned information may interfere with the retrieval of old information (or vice versa). For example, a recently purchased can of motor oil might interfere with the memory of a can of gas purchased at an earlier point in time (Shiffrin, 1970).

Interestingly, recently occurring events are thought to be better remembered than those less

recent (or remote) events. The rationale behind this finding is that more recent events are considered to be “fresher” in one’s memory and therefore more likely to be remembered than their more remote counterparts.

Now let us turn to the causes of over-reporting. Intrusions (or false positives) are usually derived from schemata or telescoping. Schemata usually consist of assumptions, beliefs, and habits about someone or something (Dashen, 2000; Schwarz & Wellens, 1997). For example, a person may remember buying a newspaper, not because he has any specific memory of that particular paper, but because he is in the habit of buying a paper on the way to work and therefore he must have done so on the particular day in question. Because habits and tendencies are subject to aberrations, a reliance on schemata will lead to errors from time to time. For example, the person who always buys the newspaper on the way to work may forget that the newsstand was sold out one day, but when asked he/she may still report buying a paper (a false positive) because of a reliance on a schemata. Telescoping errors (or misremembering the date) are another source of false positives (Loftus & Marburger, 1983). For instance, an item purchased fifteen months ago would be reported by the respondent as having been purchased during the reference period.

Aim of Present Work.

The objectives of the present research are two-fold. First, I seek to examine under- and over-reporting of frequently purchased items over time (one-week, two-weeks and four-weeks). As a second objective, I investigate an alternative method involving memory enhancers used to reduce under- and over-reporting

¹ The views expressed here are those of the author alone and do not necessarily reflect the official position of the Bureau of Labor Statistics. The author assumes responsibility for any inaccuracies.

in interviews. Memory enhancers are considered to be tips for improving one's memory. One type of memory enhancer involves remembering the details surrounding the event in question. This technique is akin to the old adage of retracing one's steps to find lost keys. This procedure has been adopted by survey methodologists. For example, Means & Loftus (1991) found that when they asked people about the weather or other circumstances leading up to a medical visit, it improved the reporting performance. Other types of memory enhancers have also been shown to improve performance. (A more elaborate description of these enhancers will be described later in this paper.) Such methods have been shown to be quite effective in improving reporting performance of health-related events, and dietary intake (or food eaten) (Fisher & Quigley, 1992; Means & Loftus, 1991).

II. STUDY 1

This experiment investigated how self-reports of purchases differ over time (one-week, two-weeks and four weeks). Participants were asked to keep a diary of their own purchases for two weeks. After the diary-recording period, a surprise memory test was administered. Respondents were asked to report their own purchases in a recall test. The results of the recall test were assessed against the appropriate diary entries.

A. Methodology

A.1 Participants. Forty-eight paid volunteer participants were recruited by advertising in a local newspaper. The participants' mean age was thirty-eight, and their average educational level was fifteen years of schooling (high school and three years of college). Participants were evenly distributed across income levels. Participants received \$25.00 for participating.

A.2 Procedure. The present study consisted of two experimental sessions as described below:

Session I: Daily Diary Recordings by Participants. Each participant was instructed to keep a diary of his/her own purchases for two consecutive weeks. Participants were required to make daily recordings of the following purchases: (a) housekeeping supplies (e.g., cleaning solutions), (b) food (including groceries), (c) personal care items (e.g., shampoo), (d) clothing expenses (e.g., pantyhose), (e) household appliances, and (f) miscellaneous expenses (including gas). These

categories were chosen because they are currently being considered in the redesign of the Consumer Expenditure Diary survey (CE Diary), which is sponsored by the Bureau of Labor Statistics.²

Participants were told that a follow-up session (Session II) was required involving additional paper work, and it was arranged. Respondents were asked to return for the follow-up session either 1, 2, or 4 weeks after the end of the reference period. Respondents were randomly assigned to each of the retention intervals. No suggestion was made that the follow-up session would involve memory tests. Respondents were instructed to mail their diaries to the experimenter upon completion of the diary (day 15).

Session II: Expenditure Reports by Respondents. All participants completed a recall test. The recall test was designed to assess the quality of self-reports. In this recall test, participants were instructed to report items from the categories that appeared on the previously completed diary. The instructions included only these open-ended categories so as not to provide cues to spur the participants' memories.

B. Results/Discussion

B.1 Analytic Strategy. The data have been broken down to simplify explication of the data. Section B.2 describes the scoring of the diaries and recall tests. Section B.3 discusses the diary-keeping performance, and section B.4 describes recall test performance.

B.2. Scoring Diaries and Recall Tests. For each participant, the reported items in the recall tests were classified into three mutually exclusive categories: (a) matches (items reported in both the recall test and the diary), (b) misses (items reported only in the diary) and (c) intrusions (items reported only in the recall test). Using the match and intrusion counts, two types proportional measures of performance were calculated: match rate³ and intrusion rate⁴. The

² The Consumer Expenditure Survey provides data used to determine the market basket- a fixed set of goods tracked over time- on which the official government inflation rate is based.

³ The match rate performance was defined as: $p(\text{match}) = m/D$. Let D be the number of items in the diary and m be the number of matches between the diary and the report.

⁴ The intrusion rate performance was defined as: $p(\text{intrusions}) = 1 - (m/R)$. Let R be the number of

match rate is the fraction of the to-be-reported items (items in the diary) that was reported correctly during the recall test. Two types of match rates were calculated: unique and duplicate match rates. The calculation of the unique match score involved eliminating all duplicates prior to scoring the number of correctly reported items. To illustrate how these measures differ consider the situation where a person's entire ensemble of purchases over a two-week period consisted of a daily "newspaper," a daily can of "Diet Pepsi" and a pair of "socks". In other words, this individual made 29 purchases (14 newspapers + 14 cans of Diet Pepsi +1 socks). Suppose at the time of test he reports "14 newspapers" and "14 cans of Diet Pepsi." While the unique match rate (involving the non-duplicate scoring system) of this individual would be .67, the duplicate match rate is .966. The intrusion rate is the fraction of the reported items that was not recorded in the diary (see Smith, Jobe, & Mingay, 1991, for the indice's rationale).

B.3 Diary-Keeping Performance. The analysis of the diary-keeping performance was intended to answer two questions: (1) Did people record fewer diary entries toward the end of the diary-keeping period? (2) Did people in the different groups report the same number of items in the diary-keeping phase? The answers to these questions provide a preliminary check of the data and may rule out possible alternative explanations for perceived differences in recall performance.

For an answer to the first question about diary compliance, the mean number of diary entries were submitted to a regression involving diary-keeping days as a predictor. When the average number of diary recordings is regressed by the diary-keeping days, it was found that the number of entries declined by -.115 each day ($t(599)=3.35$; $p<.01$). These results suggest that there was a decline in diary-keeping throughout the diary-keeping period. For an answer to the second question about uniformity amongst retention interval conditions, the mean number of diary entries were submitted to an one-way Analysis of Variance (ANOVA) involving Retention Intervals. The average number of recorded items did not differ as a function of retention intervals ($F(2,45) = .30$, $p = .74$, n.s.). The nearly equivalent numbers of diary entries for each recall group (one-, two- and four-weeks) indicate that any

items in the report and m be the number of matches between the diary and the report.

differences in recall between conditions can not be attributed to a different number of diary recordings.

B.4 Recall Test Performance. The analysis of the recall tests involved the examination of the unique match rates, duplicate match rates, and intrusion rates. The analyses for each dependent measure are discussed in the sections below.

Unique Match Rates. The analyses of the unique match rates were designed to answer the following three questions: (1) Do people consistently remember a wide variety of purchases over time (unique match rates)? (2) Do people remember recent items (those purchases recorded in the diary on days 11- 14) over time? (3) Do people remember remote items (those purchases recorded in the diary on days 1-4) over time?

For an answer to the first question about match rate consistency over time, the unique match rates were submitted to an one-way ANOVA involving Retention Intervals. Respondents remembered a fewer number of different types of items over time ($F(2,45) = 5.90$, $p < .01$). See Table 1.

Table 1. Unique Mean Match Rate Comparisons for Retention Intervals and Observed Outcomes

Retention Intervals (in weeks)	Observed Match Rates	Observed t-value	p-value
One vs. two	.46 > .32	2.22	$p < .05$
One vs. four	.46 > .24	3.38	$p < .01$
Two vs. four	.32 > .24	1.15	$p = .25$

Note: The d.f. for the LSD post-hoc t tests was 15.

For an answer to the second question about recent items (those purchases recorded in the diary on days 11- 14), the mean number of unique items were submitted to an one-way ANOVA involving Retention Intervals. Only at the shortest retention interval (one week), did respondents remember recent purchases, but the respondents' memories for recent purchases deteriorated rapidly from the one-week interval (.63) to the two-week interval (.27), ($F(2,29) = 5.83$ $p < .05$). No reliable difference in match rates was observed between weeks two and four.

For an answer to the third question about remote purchases (those purchases recorded in the diary on days 1-4), the mean number of unique remote items were submitted to an one-way ANOVA involving Retention Intervals. The data failed to

show any differences between the recollection of remote items by those people in the shorter retention intervals as compared to those in the longer retention intervals, ($F(2,33) = .942, p = .40$).

Duplicate Match Rates. The analysis of the duplicate match rates was designed to answer the following question: Do people consistently remember all of the purchases they have made over time? To answer this question, I submitted the mean duplicate match rate to an one-way ANOVA involving Retention Intervals. The mean duplicate match rate did decrease as a function of Retention Interval, but the decrease was not significant, $F(2,45) = 1.96, p = .153$. A closer look at the means reveal that there was a large drop in performance between the first (.56) and second (.39) week. However, this drop tapered off after the second week and the difference between two- weeks (.39) and four-weeks (.40) was nominal.

Intrusion Rates. The analysis of the intrusion rates was designed to answer the following question: Do people report more items falsely over time?⁵ To answer this question, I submitted the mean intrusion rates to an one-way ANOVA involving Retention Intervals. The data suggest that respondents tend to commit more intrusion errors over time ($F(2,45) = 6.76, p < .01$). See Table 2.

Table 2. Mean Intrusion Rate Comparisons for Retention Intervals and Observed Outcomes

Retention Intervals (in weeks)	Observed Intrusion rates	Observed t-value*	p-value
One vs. two	.55 < .69	2.25	p < .05
One vs. four	.55 < .78	3.64	p < .01
Two vs. four	.69 < .78	1.40	p = .17

* The d.f. for the LSD post-hoc t tests was 15.

The results of this analysis, support the conclusion that over-reporting of items increases over time. As indicated in Table 2, the intrusion rate is much lower for one-week than for the two- and four- weeks.

In sum, the present findings indicate that over time, memory for purchases deteriorated, as signified by low match and high intrusion rates. Given the amount of expenditure under- and over-reporting observed in this study, it seems useful to

⁵ When we talk about intrusions in recollection, we must acknowledge that some of these errors may be due to poor diary-keeping.

devise a means of reducing these errors. One way would be to provide the respondent with cognitive techniques proven to enhance memory (e.g., recollection of the place of purchase) as a means to reduce under- and over-reporting. For this reason, Study 2 is aimed at finding ways to improve reporting performance.

III. STUDY II

Study 2 was designed to extend the findings of Study 1 by examining ways to enhance respondents' memory of purchases. The central idea of this study was to improve reporting performance by combining a series of techniques into one package, a memory enhancer, and compare it against an interview patterned after a CE Diary survey follow-up interview (no-memory enhancer).⁶ Combining cognitive techniques has been shown to be fairly effective for improving dietary intake (Fisher & Quigley, 1992). The combination of four cognitive techniques used in this study are as follows:

1. *Recollection of Specific Circumstances.* Asking the respondent to think about specific circumstances surrounding an item and its occurrence have been shown to be an effective technique for improving memory for re-occurring events.
2. *Temporal Time Line.* A time line has been shown to improve memory for events by reducing false positives due to misremembering the date of purchase (Means and Loftus, 1991).
3. *Focused Retrieval.* Focused retrieval or simply asking the respondent to concentrate during recollection has been shown to discourage guessing and thereby reduce the number of false-reports (Fisher & Quigley, 1992).
4. *Backward retrieval.* Backward retrieval, reporting the most recent purchase first and then tracing back to the most remote, has been shown to improve memory for events.

Participants were asked to participate in a two-phased study: (1) In the diary-keeping phase, participants kept diaries of their own purchases for one week, and (2) in the test phase, people were

⁶ In the CE follow-up interview, people are typically asked to report any thing they may not have mentioned in their Diaries.

asked to report their purchases under one of two conditions: (a) memory enhancer and (b) no-memory enhancer.

A. Methodology

A.1 Participants. Thirty-six paid volunteer participants were recruited by an advertisement in a local newspaper for this study. Each received \$25.00 for participation. Sixteen participants were randomly assigned to the memory enhancer and twenty to the non-memory enhancer. The average educational level was sixteen years of schooling (completion of college).

A.2 Procedure. The present study consisted of two experimental sessions as described below:

Session I: Daily Diary Recordings by Respondents. All subjects were instructed to keep a diary for one week. Recording categories and instructions were identical to those in Study 1.

Session II: Expenditure Reports by Respondents. Respondents were randomly assigned to either the memory enhancer or no-memory enhancer condition. The no-memory enhancer condition was identical to the one used in Study 1 which was patterned after the existing Consumer Expenditure Survey follow-up interview. All respondents in the memory enhancer condition were asked (a) to recreate the original context in which they made the purchase, (b) to concentrate on the task at hand and to be as accurate as possible, (c) to report purchases in backward order and, (d) to use a time line which should facilitate backward reporting. While all respondents engaged in each activity the order of activities varied across respondents. With the exception that instruction to use the time line always preceded the instruction to report backward.

B. Results/Discussion

B.1 Analytic Strategy. The data have been broken down to simplify explication of the analyses. Section B.2 describes the scoring of the diaries and recall tests. Section B.3 discusses the diary-keeping performance and section B.4 describes the recall performance.

B.2 Scoring Diaries and Recall Tests. Diary and test scoring was identical to Study 1. As in the previous study, two measures were calculated from the diaries and tests: (1) match rate (unique and duplicate), and (2) intrusion rate.

B.3. Diary-keeping Performance. The analysis of the diary-keeping performance was intended to answer two questions: (1) Did people record fewer diary entries toward the end of the diary-keeping

period? (2) Did people in the different groups record the same number of items in the diary-keeping phase?

To answer the first question about compliance, I submitted the mean number of unique diary entries to a regression involving diary-keeping days as a predictor. When the average number of diary recordings was regressed by days, it was found that the number of entries declined 2.19 each day through out the diary-keeping period ($t(210)=2.79$; $p<.01$).

For an answer to the second question about uniformity among retention intervals, the mean number of diary entries were submitted to a one-way ANOVA involving conditions. The average number of recorded items did not differ as a function of conditions ($F(1,34) = .368$, $p = .548$, ns.). The observed results demonstrate that participants recorded nearly equivalent numbers of diary entries irrespective of the conditions.

B.4 Recall Test Performance. The analysis of the recall tests involved the examination of the unique match rates, duplicate match rates and intrusion rates. These analyses are discussed in the sections below.

Unique Match Rates. The analyses of the unique match rates were designed to answer the following three questions: (1) Does the memory enhancer improve memory for a variety of distinct purchases? (2) Does the memory enhancer improve memory for recent items? (3) Does the memory enhancer improve memory for remote items?

For an answer to the first question about distinct purchases, the unique match rates were submitted to an one-way ANOVA involving conditions. The results indicate that memory enhancer participants showed a clear improvement on purchase reporting (.65), compared to those in the non-enhancer memory group (.51), $F(1,34) = 4.60$, $p < .05$.

For an answer to the second question about recent purchases (those purchases recorded in the diary on days 6-7), the unique match rates were submitted to an one-way ANOVA involving conditions. Memory enhancer participants did not remember more recent purchases than did those in the non-memory enhancer group, $F(1,34) = .082$, $p=.77$.

For an answer to the third question about remote events (those purchases recorded in the diary on days 1-2), the unique match rates were submitted to an one-way ANOVA involving conditions. The memory enhancer participants remembered more

remote (.67) purchases than did those in the non-memory enhancer group (.41), $F(1,34) = 14.3$, $p < .01$.

Duplicate Match Rates. The analyses of the duplicate match rates were designed to answer the following question: Do memory enhancers improve recollection for a lot of purchases? To answer this question, I submitted the mean duplicate match rate to a one-way ANOVA involving conditions. Memory enhancer participants remembered more purchases (.71) than did those in the non-memory enhancer group (.53), $F(1,34) = 4.59$, $p < .05$.

Intrusion Rates. The analysis of the intrusion rates was designed to answer the following question: Do memory enhancers reduce intrusions? To answer this question, I submitted the mean intrusion rates to a one-way ANOVA involving conditions. Memory enhancer participants reported fewer intrusions (.37) than did those in the non-memory enhancer group (.51), $F(1,34) = 4.102$, $p < .05$.

IV. Concluding Remarks

The current studies underscore the complexity of purchase recollection in retrospective surveys. Study 1 clearly points to memory deterioration over time and documents how recency improves memory for events. Study 2 provides an example of how the contribution of a cognitive psychology perspective can make towards reducing sources of response error in retrospective reports.

This work is of practical interest. For the designer of retrospective surveys, the most obvious message is that respondents will forget and that this memory lapse will affect the interpretation of any results. Whenever possible, the time between the completion of the diary and the follow-up interview should be minimized since forgetting occurs over time.

Do we know what would be the optimal retention interval or delay between the diary completion and interview? These findings suggest that the delay should be ideally one day because a large fraction of the information is lost within the first week. In contrast, the effects of a two-week delay on reporting performance are relatively small because a smaller fraction of the information was lost after two weeks than after one week.

Earlier studies examining reports of events over time (Smith, Jobe & Mingay, 1991) used unique match rates as a measure of reporting performance. This work suggests that models should also include the extent to which the item is reported multiple items rather than consolidate them into discrete events. In doing so, this type of measure will better reflect respondents' recollection over time.

V. REFERENCES

- Crowder, R. G. (1976). Principles of learning and memory. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Dashen, M.L. (2000). Effects of Retention Intervals on Self- and Proxy-Reports. Memory, 8(3), 129-143
- Ebbinghaus, H. (1964). Memory: a contribution to experimental psychology. (H. A. Ruyser and C. E., Bussenias, trans.). New York: Dover (original work published 1885).
- Fisher, R.P., & Quigley, K.L. (1992). "Applying cognitive theory in public health investigations: enhancing food recall with cognitive interviews." In J. Tanur (Ed.) Questions about questions: inquiry into the cognitive bases for surveys. New York: Russell Sage Foundation.
- Loftus, E.F. & Marburger, W. (1983). Since the eruption of Mt. St. Helens, has anyone beaten you up? Improving the accuracy of retrospective reports with landmark events. Memory and Cognition, 11, 114-120.
- Means, B. & Loftus, E. F. (1991). When a personal memory repeats itself: Decomposing memories for reoccurring events. Applied Cognitive Psychology, 5, 297-318.
- Schwarz, N., & Wellens, T. (1997). Cognitive dynamics of proxy responding: The diverging perspectives of actors and observers. Journal of Official Statistics, 13(2) 159-179.
- Shiffrin, R.M., (1970). Forgetting: Trace erosion or retrieval failure? Science 168 (3939) 1601-1603.
- Smith, A. F., Jobe, J. B., & Mingay, J. D. (1991). Retrieval from memory of dietary information. Applied Cognitive Psychology, 5, 269-296.
- Sudman, S., & Bradburn, N. M. (1973). Effects of time and memory factors on response in surveys. Journal of American Statistical Association, 68: 805-815.
- Smith, A., Jobe, J. & Mingay, D. (1991). Retrieval from memory of dietary information. Applied Cognitive Psychology, 5 269-296.
- U.S. Bureau of Labor Statistics (1987). "Consumer Expenditure Survey, 1987," BLS Bulletin, No. 2354. Washington, DC: U.S. Government Printing Office.