

## HOW DOES PROXY ACQUISITION AND EXPERTISE INFLUENCE SURVEY REPORTS?

Monica L. Dashen, U.S. Bureau of Labor  
Statistics  
2 Massachusetts Ave. N.E., Suite 1950  
Washington D.C. 20212 USA<sup>1</sup>

### **Key words: self- and proxy-reports and response error**

The present work explores the effectiveness of cognitive tools associated with assessing the quality of self- and proxy reports of an event. Such reports can vary in quality for a variety of reasons. For one, the quality of proxy reports depends upon how proxies learned about the event. Those proxies who witnessed the event generally have a richer, fuller and more accurate account of the event than those who simply heard about it (e.g., Sudman, Bickart, Blair, & Menon, 1994; Thompson, Skowronski, Larsen, & Betz, 1996). Moreover, those proxies who already knew much about the event and its surrounding details may recollect it better than those who did not.

In contrast to proxy reports, self-reports usually do not vary much in quality because self-reporters always witness their own behavior. However, there is one exception: When self-reporters recount the details of an event to another, they usually remember the event better (e.g., Rundus & Atkinson, 1970; Suengas & Johnson, 1988). How much reporters already know about an event, how they learn about an event, and whether they tell another affects the quality of their reports. Therefore, such reports can vary widely in quality, and the survey designers need to distinguish good reports from bad ones. Cognitive tools can provide the means for sorting out these reports.

Cognitive tools have been implemented in survey situations in the past. One such class of tools, report improvement tools, involves implanting procedures during the interview to improve the respondent's memory of the event.

For example, bounding and landmark techniques typically improve the accuracy of event dating (e.g., Gaskell, Wright, & O'Muircheartaigh, 2001; Neter & Waksberg, 1964). Report improvement tools also may involve memory enhancement techniques, whereby the interviewer provides specific cues to improve the reporter's event memory (e.g., Dashen, 2000b; Geiselman, Fisher, Mac Kinnon, & Holland, 1985; Means & Loftus, 1991).

In contrast to report improvement tools, report assessment tools offer a way to evaluate the quality of a report, but do not provide any means for improving it. There are three such tools: (a) confidence ratings; (b) expertise; and (c) discussion. While all three tools focus on predictors of good reports, they differ in nature. The first tool, confidence ratings, relies on an individuals' ability to assess their own reports. In contrast, the remaining two tools focus on individual characteristics of the reporter as predictors. Similar inquiries have been conducted by Brian Kojetin and his colleagues (e.g., Kojetin & Miller, 1993). All proposed tools may provide ways to sort out good reports from bad ones.

With the first tool, confidence ratings, respondents are asked to assess their reports by evaluating the fit between their reported memory and the original event. (Usually, the higher the confidence rating, the better the fit between the reported memory and original event.) Individuals typically provide lower confidence ratings when they are relatively unsure about their report of an event.

There have been mixed reports on the utility of confidence ratings. While Thompson et al. (1996) found that confidence ratings were a good predictor of event dating accuracy, Sudman et al. (1994) found that high confidence ratings did not necessarily lead to high agreement between self- and proxy reports. One possibility that might explain these mixed reports, which this paper investigates, is that reporters' confidence ratings are shaped by their level of direct experience. That is to say, respondents' confidence levels may be derived from their ability to recollect the event and the details surrounding it. The inability to recollect many event details may lead respondents to assign a lower confidence rating value to their recollections than those who recollect many details. Those proxies who witnessed an event should assign a higher confidence rating to their recollections than individuals who only heard about it. For example, proxies who watched their spouses eat

---

<sup>1</sup> 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. The author would like to thank John Dixon, Bill Mockovak, and Christine Rho for their helpful comments.

a slice of pizza that came right out of the oven and smelled wonderful but burned their spouses' mouth when they tried to eat it too soon, may assign a higher confidence rating value to their recollection than might those proxies who were not privy to the details surrounding the purchase of the slice.

The second assessment tool, expertise, involves finding out how much the respondent knows about a particular topic. Knowing a lot about a particular topic often leads to a fuller and richer recollection of an event than knowing little about it. This is because experts typically have a more differentiated and broader knowledge base than do novices (e.g., Bickart, 1992). Therefore, to better evaluate report quality, interviewers should inquire after the interview about the respondents' expertise on topics related to the event. Or if the interviewers are in a position to do so, they may seek out the expert before the start of the interview in order to "stack the deck" in favor of a good report. For example, identifying the person in charge of household finances may produce high quality reports for a survey interested in household spending habits. Moreover, even if these financial experts do not witness other family members' purchases, the fact that these experts keep track of the families' expenses may compensate to some extent for the experts' absence at the time of the purchase.

The third assessment tool, discussion, involves identifying the self-reporters who often recount their activities to others. Recounting the event to someone else may affect one's own memory of the event because it serves as a form of rehearsal, which has been shown to improve memory (e.g., Rundus & Atkinson, 1970; Suengas & Johnson, 1988). As with expertise, after the interview, interviewers should inquire about how much the respondent talked about the event, thus evaluating the report quality. Or if the interviewers are in a position to do so, they may seek out the "talker" before the start of the interview in order to "stack the deck" in favor of a good report. Individuals who talk a lot about their purchases to others may be ideal candidates for a survey interested in household spending habits. Such a "talker" may relish recounting the thought process that went into choosing an item for purchase, for example, a birthday present for a child who is hard to please. Subsequently, this "talker" may have a richer and fuller recollection of the event than a "non-talker."

In summary, cognitive assessment tools are sorely needed in order to sort out good reports from bad ones because there is such a wide

variation in report quality. To fill this gap, I have suggested three tools as viable candidates: confidence ratings, expertise, and discussion. Although this list of candidates is hardly exhaustive, these tools offer discernable results and can be used in surveys that address family spending habits, such as the Consumer Expenditure survey.

Deciding how well these three tools can separate good quality reports from bad ones requires an in-depth view of respondents' recollection of an event. Simply asking people to recollect the name of the event does not provide enough material for evaluating these tools' effectiveness. Rather, a fuller picture of this event is needed, and this can be provided by focusing on the event's qualitative aspects. Qualitative aspects of the memory record, in general, refer to the sensory (e.g., sight, smell, taste, and touch), temporal (e.g., season, day, and hour) and emotional (e.g., thoughts and feelings) aspects of the event. After sanding a piece of wood, for example, one may recollect its blond color and smooth texture, and the amount of dust generated. (For a more detailed account of the qualitative aspects of a memory, see Suengas and Johnson [1988]).

Two points are worth mentioning in regards to the qualitative aspects of an event. First, the extent to which people may recollect an event varies in degree from a lot to a little. Rarely, is it an "all or nothing" situation, in which people remember all aspects of the event or none. Second, although people may remember more than one aspect, they may not remember each aspect to the same degree. For example, people may remember the frustration of buying a gift for a child who is hard to please more than the color of the gift's wrapping paper.

Not only does focusing on the qualitative aspects of an event provide insight into the three proposed assessment tools, it also lends some real world immediacy to survey results. Survey designers often ask respondents to consider qualitative aspects of an event when answering questions. For example, the National Health Interview Survey asks about the frequency one engages in light, moderate, and vigorous activities (as defined by general levels of sweating, breathing, and heart rate). To answer this question, respondents must consider their own levels of sweat and breathing, as well as other qualitative aspects of exercise. (See Conrad, Brown, and Dashen [2001] for a detailed account of such qualitative recollections.)

This work focuses on the qualitative aspects of the memory record as a dependent measure of interest. As such, this work extends beyond the previous self/proxy work that has relied on dependent measures, which have focused on sheer reporting volume (e.g., Silberstein, 1989), reporting accuracy (e.g., Dashen, 2000a), and reporting strategies (Beatty, Herrmann, Puskar, & Kerwin, 1998; Blair, Menon, & Bickart, 1991).

While the focus of this work is on assessment tools, the qualitative aspects of an event under study should be of interest to those individuals interested in improving reporters' memories for events. If, for example, a great many people remember based on sensory aspects (e.g., the look of packaging) but only under special circumstances (e.g., only for self-purchases), survey designers may be able to insert cues into those questions to help participants recall the memories that seem to work best. It is not only the questions that may be affected; in their instructions, survey designers may want to include tips for remembering the event (e.g., Dashen, 2000b).

**Aims of the Study**

The experiment reported in this work has three specific aims. The first is to find out whether proxies adjust their confidence ratings according to how they learned about an event. Because they rely on the details surrounding the event, proxies who witnessed the event ought to assign a higher confidence rating value than those who only hear about it. If confidence ratings are shaped by the proxy's amount of direct experience with the event, these ratings would be a useful tool for survey designers who are concerned about data quality.

The second aim is discover whether experts have a richer and fuller recollection of an event than novices do. If indeed this is the case, then a further aim is to find out whether expertise partially compensates for the fact that the proxy only hears about the event. Moreover, those expert proxies who hear about the event ought to have a fuller and richer recollection of the event than do those novice proxies who hear about it.

The third aim is to find out whether recounting an event to another strengthens a person's memory of it. If the recounting strengthens memory, this finding suggests that not all self-reports are equivalent and other mediating factors besides direct involvement in an event can improve memory.

**Study**

**Method**

*Participants.*

Twenty-seven men and twenty-seven women participated in this study. The participants' mean age was 52.24 years, and their average educational level was 16 years of schooling (four years of college). The couples had been married for an average of 22.55 years.

*Design*

The experiment employed a 2 (expert versus novice) x 3 (engage-witness, engage-hear, versus engage-neither witness nor hear type of self/proxy acquisition) x 4 (sanding, unscrewing, dusting, and cleaning activities) x 2 (self- versus proxy reporter) mixed design, with the first two factors varying between groups, and the fourth factor varying within participants. The third factor is nested within couple. The expertise factor has two levels, expert and novice, which refer to how much the individual knows about a particular topic (e.g., housekeeping or home-repair). Table 1 describes the self/proxy acquisition factor levels.

Table 1. Description of each type of self/ proxy acquisition.

Level of Self/Proxy Acquisition	Explanation
Engage-Witness	The proxy reporter <i>observes</i> the self-reporter engage in an activity. *
Engage-Hear (Perceptual)	The proxy reporter <i>hears about the perceptual aspects</i> of the event from the self-reporter.
Engage-Hear (Appreciative)	The proxy reporter <i>hears about the appreciative aspects</i> of the event from the self-reporter.
Engage-Neither Witness or Hear	The proxy reporter is <i>unaware</i> of the self-reporter's involvement in the activity.

(\* ) Neither reporter discussed the event at the time of its occurrence.

Columns 1 and 2 of Table 1 describe the three types of self/proxy acquisition: engage-witness, engage-hear (perceptual/appreciative), and engage-neither witness nor hear. (Participants were randomly assigned to each type of self/proxy acquisition.)

With regard to the first type of self/proxy acquisition, “engaged-witness,” individuals both observed their spouses engage in two separate activities and engaged in another pair of activities themselves (in the presence of their spouse). With regard to the second type of self/proxy acquisition, “engaged-hear,” participants discussed certain aspects—perceptual and appreciative—of the activity. (While perceptual aspects refer to the physical or sensory details of the activity, appreciative aspects refer to the emotional or thought-related details of the activity.)

Those “perceptual-hear” proxies heard about three of the following six aspects of an event: (1) position and spatial arrangement of people, (2) shape and spatial arrangement of objects, (3) anything that the individual was touching, (4) colors, (5) noises and voices that you [as the participant in this study] were hearing, and (6) anything the individual was looking at. For example, proxies heard about the texture and color of the sandpaper and wood from the self-reporter. (The order in which each aspect was discussed varied by person.)

Those “appreciative-hear” proxies heard about three of the following six appreciative aspects: (1) any negative feelings that the individual had, (2) any positive feelings that the individual had, (3) any comments the individual could have made but did not make, (4) any difficulties experienced, (5) any ideas that came to mind, and (6) any other time in which the person thought or felt the same way. For example, the proxy reporters heard about how much sand was generated and how messy it was to sand from the self-reporter. (The order in which each aspect was discussed varied by person.)

Individuals in the “engaged-neither witness nor hear knowledge” type of self/proxy acquisition were only told the name of the activity (e.g., sanding a piece of wood). For example, proxies heard that their spouses engaged in a sanding activity but did not know any other details about the event. This condition allowed me to see what happens when the proxy has little knowledge of the actual event but nonetheless is asked about it.

#### *Procedure.*

The present study consisted of three experimental phases as described below:

*Phase I: Completion of the Knowledge Tests.* To assess the expertise in either home-repair or housekeeping, each participant completed a ten-question knowledge test. Of the ten questions, five pertained to housekeeping and the remaining to home-repair. The five questions per topic varied in level of difficulty.

*Phase II: Engagement of the Activities.* Participants engaged in one home-repair (sanding or unscrewing) and one housekeeping activity (dusting or cleaning).

*Phase III: Completion of a Memory Characteristic Questionnaires (MCQ).* All participants rated their own and their spouses activities on the MCQ. The MCQ, which contains Lickert-type scales, offers an effective way to assess differences in the qualitative aspects of an event. The MCQ, which was developed by Marcia Johnson (of Princeton University), has been proven to be a useful tool for distinguishing between real and imaginary events (Johnson, Foley, Suengas, & Raye, 1988; Sueguas & Johnson, 1988;). However, the MCQ has not been used to distinguish between actual memories of an event and second-hand reports. The present work will be the first to report the MCQ’s effectiveness at assessing the differences between such reports.

#### **Results/Conclusions**

The present analyses have been broken down into three sections. Section 1 describes the selection of the dependent variables to be used in these analyses. Section 2 describes a close-up examination of each of the confidence rating, expertise, and discussion tools.

#### *Selection of the Dependent Variables.*

The MCQ used to rate participant memories included questions designed to assess a range of memory qualities. I evaluated the effects of self/proxy acquisition and expertise on self- and proxy differences by averaging across groups of items that were expected to draw on common memory characteristics. These common characteristics were patterned after Suengas & Johnson’s (1988) four composite factors: (1) clarity (2) context, (3) thoughts and feelings and, (4) intensity of feelings factor.

*A close-up view of the proposed assessment tools.*

The analyses of the five factors were designed to answer three questions intended to give a general idea of the effectiveness of the proposed assessment tools: (1) Does the act of telling another (proxy) about an event shape the self-reporter's memory of the event for the better? (2) When making confidence ratings do proxy reporters adjust their ratings according to how they learned about it? (3) Do experts have a fuller and richer recollection of the event than do novices? If so, does expertise compensate to some extent for the fact that proxies neither engaged in the event nor were present at the time the event occurred? To address these questions the four mean factor ratings and confidence ratings per participant were submitted to the above mentioned mixed design. Relevant LSD post-hoc comparisons among various factors were made.

The goal of the first question was to find out whether recounting an event to another (proxy) produces a stronger memory of the clarity, sensory, thoughts, feelings and context aspects of the event. The significant reporter ( $F[1,122]=49.23$ ,  $Mse=33.7$ ,  $p<.001$ ) and self/proxy acquisition ( $F[2,72]=22.73$ ,  $Mse=15.6$ ,  $p<.001$ ) main effects and interaction ( $F[2,122]=15.00$ ,  $Mse=10.15$ ,  $p<.001$ ) for clarity support this conjecture. Also, the significant reporter ( $F[1,122]=23.07$ ,  $Mse=33.08$ ,  $p<.001$ ) and self/proxy acquisition ( $F[2,50]=12.85$ ,  $Mse=25.17$ ,  $p<.001$ ) main effects and interaction ( $F[2,122]=6.17$ ,  $Mse=8.50$ ,  $p<.01$ ) for feelings support this conjecture. Moreover, self-reporters who discussed the event with another had reliably higher mean clarity ratings (6.67) than did those self-reporters who did not discuss it (5.84),  $LSD \ t(1)=3.45$   $p<.01$ . In addition, self-reporters who discussed the event with another individual had reliably greater mean feeling of intensity ratings (5.96) than did those self-reporters who did not discuss it (4.74),  $LSD \ t(1)=3.48$ ,  $p<.01$ . No other significant comparisons were observed.

The goal of the second question was to find out whether confidence ratings were sensitive to how the proxy learned about the event. The significant reporter ( $F[1,122]=32.8$ ,  $Mse=37.29$ ,  $p<.001$ ) and self/proxy acquisition ( $F[2,65]=3.11$ ,  $Mse=4.86$ ,  $p<.05$ ) main effects and interaction ( $F[2,121]=6.44$ ,  $Mse=7.32$ ,  $p<.01$ ) for confidence ratings support this conjecture.

Proxies who witnessed their spouse engage in an activity had a reliably higher mean confidence rating (6.13) than did those who only heard about it (5.06)  $LSD, \ t(1)=3.54$ ,  $p<.001$ . In contrast, proxies who hear about the event did not have a reliably higher mean confidence rating (5.06) than did those who only heard the activity name (5.04)  $LSD, \ t(1)=.082$ ,  $p=.94$ . The latter finding suggests that proxies rely on the details of the event to evaluate the fit between the original event and their memory of the event.

The goal of the third and final question was to find out whether experts have a stronger recollection of the event than do novices, and if indeed this were the case, then does expertise compensate for a proxy's absence? The insignificant results of the Expertise main effect and the Reporter x Self/Proxy acquisition x Expertise interaction for all factors suggest that experts do not have a stronger memory of the event than do novices. (The range of  $F_s(2,122)$  and  $MS_e$  for insignificant interactions ( $p>.05$ ) are as follows: ,  $F_s$  between .02 and 1.78,  $MS_e$ s between .01 and 3.5.) These findings also suggest that self/proxy acquisition overrides any advantage an expert brings to the test.

To summarize, three principal findings have been demonstrated. First, recounting the details of an event shapes the self-reporter's memory. This finding suggests that not all self-reports are created equal and that there are factors mediating good quality self-reports other than simply engaging in an activity, such as the level of discussion. Second, proxy reporters adjusted their confidence ratings according to how they learned about it. Third, the Memory Characteristic Question (MCQ) is a viable instrument for detecting self-reporter differences.

## References

- Beatty, P., Hermann, D., Puskar, C., & Kewin, J. (1998). "Don't know" responses in surveys: Is what I know what you want to know and do I want you to know it?" *Memory*, *6*, 407-426.
- Bickart, B. (1992). Question-Order Effects and Brand Evaluations: Moderating Role of Consumer Knowledge. In Schwarz, N. & Sudman, S. (Eds.) (pp. 63-80) *Context Effects in Social and Psychological Research*. New York: Springer-Verlag.

Blair, J., Menon, G. & Bickart, B. (1991). Measurement effects in self- vs. proxy responses: An information-processing perspective. In P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz & S. Sudman (Eds.) Measurement Errors in Surveys (pp. 145-166). New York: Wiley.

Conrad, F. Brown, N. and Dashen, M. (2001). Estimating the frequency of events from unnatural categories. (Under review.)

Dashen, M.L. (2000a). The effects of retention intervals on self- and proxy reports of purchases. Memory. 8(3), 129-143.

Dashen, M.L. (2000b). Improving Purchase Recall. Proceedings of the 160<sup>th</sup> Annual American Statistical Association (Survey Methods Section). (In press.)

Dobbins, I., Kroll, N., & Linn, Q. (1998). Confidence-accuracy inversions in scene recognition: A remember-know analysis. Journal of Experimental Psychology: Learning, Memory, and Cognition. 24(5). 1306-1315.

Gaskell, G., Wright, D. & O'Muircheartaigh, C., (2001). Telescoping of Landmark Events: Implications for Survey Research. Public Opinion Quarterly, 64 77-89.

Johnson, M., Foley, M. Suengas, A. & Raye, C. (1988). Phenomenal characteristics of memories of perceived and imagined autobiographical events. Journal of Experimental Psychology: General. 117(4) 371-376.

Kojetin, B. & Miller, L. (1993). The intrahousehold communication study: Estimating the accuracy of proxy responses at the dyadic level. Proceedings of the 153<sup>th</sup> Annual American Statistical Association (Survey Methods Section), 2 1095-1110.

Menon, G., Bickart, B., Sudman, S., & Blair J. (1995). How well do you know your partner? Strategies for formulating proxy reports and their convergence with self-reports. Journal of Marketing Research, 32 75-84.

Neter, J. & Waksberg, J. (1964). A study of response errors in expenditure data from household interview. Journal of the American Statistical Association, 59 18-55.

Rundus, D. & R. Atkinson (1970). Rehearsal Processes in Free Recall: A Procedure

for Direct Observation. Journal of Verbal Learning and Verbal Behavior 9, 99-105.

Silberstein, A. (1989). Recall effects in the U.S. Consumer Expenditure Interview Survey. Journal of Official Statistics. 5(2). 125-142.

Sudman, S., Bickart, B., Blair, J., & Menon, G. (1994). "The effect of participation level on reports of behavior and attitudes by proxy reporters." in N. Schwarz & S. Sudman (Eds.), Autobiographical memory and the validity of retrospective reports. New York: Springer-Verlag (pp. 251-265).

Suengas, A.G. & Johnson, M.K. (1988). Qualitative effects of rehearsal on memories for perceived and imagined complex events. Journal of Experimental Psychology: General 117, 377-389.

Thompson, C., Skowronski, J., Larsen, S. & Betz, A. (1996). Autobiographical Memory: Remembering What and Remembering When. Mahwah, New Jersey: Lawrence Erlbaum Associates.

U.S. Bureau of Labor Statistics (1987). "Consumer Expenditure Survey, 1987," BLS Bulletin, No. 2354. Washington, DC: U.S. Government Printing Office.