

# Measuring strategy production during associative learning: The relative utility of concurrent versus retrospective reports

JOHN DUNLOSKY

*University of North Carolina, Greensboro, North Carolina*

and

CHRISTOPHER HERTZOG

*Georgia Institute of Technology, Atlanta, Georgia*

Strategy production during associative learning can be measured by self-reports made either concurrently with study or retrospectively. Both kinds of report presumably have strengths and weaknesses, yet a systematic comparison has not been conducted. Younger and older adults studied paired associates and reported strategy production using one or both kinds of report. Participants either received or did not receive descriptions of mediational strategies prior to study. Retrospective reports were not completely consistent with concurrent reports, suggesting that the validity of retrospective reports is somewhat diminished by forgetting. Making concurrent reports did not affect subsequent retrospective reports, but describing strategies affected reported strategy production for both age groups and paired-associate recall for older adults. A production deficiency constrained older adults' performance when they did not receive strategy descriptions prior to study. Discussion focuses on the relative utility of concurrent and retrospective reports of strategy production.

The importance of strategies to our understanding of individual differences in learning is evident from the extensive literature on the topic (Rogers, Hertzog, & Fisk, 2000). The strategies that people use to study paired associates have a substantial influence on memory performance, and in general, evidence from investigations that focus on differential strategy production has yielded many tests of theory-based hypotheses about memory. The success of such research partly depends on valid identification of the strategies that an individual produces. Given the number of difficulties that arise in measuring strategy use, how should one identify the strategies that an individual produces during study?

To address this question, we focus on subjective reports, a class of measures that has been widely used to assess strategy production during associative learning (for reviews, see Dunlosky & Hertzog, 1998; and Richardson, 1998). Subjective reports can be made either concurrently with study or retrospectively after study. For *concurrent reports*, an individual overtly reports the strategy used to study an item immediately after that item has been studied, so that the time between strategy production and each report is minimal. For *retrospective reports*, an individual studies the list of items and then, some time after study

(e.g., after the test), reports the strategies that have been used. Both concurrent reports and retrospective reports are made on an item-by-item basis, which allows fine-grained analyses of how strategies are distributed across items. Richardson has argued for the superiority of retrospective over concurrent reports, mainly because concurrent "reports obtained during presentation may significantly alter the subjects' cognitive activities" (pp. 608–609). By contrast, Dunlosky and Hertzog have argued for the superiority of concurrent reports because retrospective reports "will rely on retrieval from long-term memory, [and hence] these reports may yield inaccurate estimates of the relative production of strategies" (p. 598).

Which kind of report provides a more adequate measure of strategy production? Is one report better under all circumstances, or are concurrent reports superior in some circumstances, and retrospective reports superior in others? Our major goal involved answering these questions through a hypothesis-driven evaluation of the two kinds of subjective reports of strategy production.

Two concerns central to the validity of strategy reports are (1) the degree to which the reports themselves influence the strategies that an individual produces (see, e.g., Adams & Montague, 1967), and (2) the degree to which the strategies produced are forgotten prior to their being reported (e.g., Montague, Adams, & Kiess, 1966). The first concern pertains primarily to concurrent reports, which can result in somewhat inaccurate estimates of the strategies that would be produced given that strategy reports were not made. At different extremes, concurrent

---

This research was supported by a grant from the National Institute on Aging (5 R37 AG131148). We thank Ute Bayen for comments on an earlier draft of this manuscript. Send correspondence to J. Dunlosky, P.O. Box 26164, Department of Psychology, University of North Carolina, Greensboro, NC 27402-6164 (e-mail: dunlosky@uncg.edu).

reports may impair the discovery of strategies and hence constrain memory performance (Richardson, 1998), or they may prompt participants to use strategies that they otherwise would not. To better understand the latter possibility, consider that to collect strategy reports, researchers often use nominal self-report scales, with each point on the scale referring to one strategy (e.g., 1 = *imagery*, 2 = *rote repetition*, etc.). To provide a common interpretation of the scale, participants need a brief description of each strategy. When individuals make concurrent reports, the descriptions must be provided before study, which may inadvertently alter strategy production. The second concern above pertains to retrospective reports, which may result in inaccurate estimates of the strategies that have been produced during study. At one extreme, participants may forget all of the strategies produced during study, such as would be the case if they had a memory deficit or if the interval between study and reports was inordinately long.

### **An Empirical Comparison of Concurrent and Retrospective Reports**

We evaluated the relative utility of the two kinds of report by testing several hypotheses. We will first outline our design and then describe the hypotheses along with the corresponding planned comparisons. All participants read instructions for paired-associate learning (which did not describe any strategies), studied paired associates (e.g., *dog-spoon*), and made retrospective reports after the test. Three groups differed with respect to whether they made concurrent reports and/or received strategy descriptions. The *concurrent group* read descriptions of three strategies (imagery, sentence generation, and rote repetition), which did not mention the effectiveness of the strategies. Immediately after studying each item, these participants made a concurrent report. The *informed retrospective* group also received the strategy descriptions prior to study, but they did not make concurrent reports. The *uninformed retrospective* group did not receive the strategy descriptions.

The *reactive effects* hypothesis is that concurrent reports have reactive effects on the kinds of strategy reported and on paired-associate recall. As suggested above, reactive effects may result from two sources: making concurrent reports or receiving strategy descriptions. Given the present design, sources that contribute to any reactive effects can be inferred from comparisons among all three groups. If making concurrent reports has a reactive effect on strategy production, the retrospective reports will differ for the concurrent group as compared with the informed retrospective group.<sup>1</sup> If receiving descriptions about the strategies has a reactive effect on strategy production, the retrospective reports will differ for the informed retrospective group as compared with the uninformed retrospective group. Analogous comparisons were also conducted to evaluate the source of any reactive effects on paired-associate recall.

The *inconsistent reports* hypothesis is that retrospective reports differ from the concurrent reports made during study. Such inconsistency presumably results from forgetting that occurs in the filled interval between studying and the subsequent reports. Evidence from Montague et al. (1966) confirmed this hypothesis. Namely, the majority of mediators reported during study either were not reported or were different when retrospective reports were collected 24 h after study. Adams and McIntyre (1967) found that about 19% of the mediators originally reported were not reported or were different during retrospective reports made 1 week after study. Such inconsistency suggests that retrospective reports are not entirely valid when relatively long-term retention of mediators is required. Accordingly, researchers often use a minimal interval between study and retrospective reports. What is presently unknown, however, is whether forgetting can influence retrospective reports even when they are made soon after study. We evaluated this possibility by having the concurrent group also make retrospective reports immediately after the test. To the degree that the validity of retrospective reports is constrained by factors such as forgetting, retrospective reports made after the test will diverge from the concurrent reports made during study.

### **Do Production Deficiencies Contribute to Age Deficits in Associative Learning?**

We also evaluated the production deficiency account of age-related differences in associative memory in which older adults' production of inferior strategies (or no strategies) at study contributes to age deficits in memory performance. In agreement with some of the previous literature (see Light, 1991), Dunlosky and Hertzog (1998) have concluded that production deficiencies do not describe age differences in associative learning. However, this conclusion was based on age equivalence in concurrent reports, which necessitated providing strategy descriptions prior to study. Perhaps when older adults receive no information about strategies, they less often produce effective strategies. We conducted planned comparisons to test two predictions: first, that the profile of reported strategies for older adults would differ for the informed retrospective group versus the uninformed retrospective group, with the latter more often reporting ineffective or no strategies; and second, that paired-associate recall would be greater for the two informed groups than for the uninformed group. Obtaining both outcomes would indicate that production deficiencies partly constrain older adults' paired-associate recall.

## **METHOD**

### **Design and Participants**

The design was a 3 (group: concurrent, informed retrospective, uninformed retrospective)  $\times$  2 (age: young vs. older adults) factorial. Participants were randomly assigned to groups by order of appearance, which included 98 younger adults and 94 older adults.

For the younger adults, 32, 32, and 34 participated in the concurrent, informed retrospective, and uninformed retrospective groups, respectively. For the older adults, 33, 30, and 31 participated in the concurrent, informed retrospective, and uninformed retrospective groups, respectively. The younger adults ( $M$  age = 19.9,  $SD$  = 1.4) received course credit for participating and were students at the Georgia Institute of Technology. The older adults ( $M$  age = 69.8,  $SD$  = 5.9) received a nominal fee and were normal, community-dwelling adults recruited from the Atlanta area. All participants were tested in a campus laboratory, and participants were tested individually.

A demographic questionnaire and a standard vocabulary test (Ekstrom, French, Harman, & Dermen, 1976) were administered after the completion of the experiment. For each question on the vocabulary test, the participants had to select the closest synonym for a word from five alternatives. The participants had 4 min to answer as many of the 36 questions as possible, and points were not deducted for errors. Performance on the vocabulary test was greater for the older ( $M$  = 22,  $SD$  = 7.7) than for the younger adults ( $M$  = 17,  $SD$  = 2.7) [ $t(124) = 5.13$ ]. The older adults were well educated, as indexed by years of education ( $M$  = 15.4,  $SD$  = 2.7). The participants also rated their health, with 1 indicating *excellent health* and 4 indicating *poor health*. Both the younger adults ( $M$  = 1.5,  $SD$  = .56) and the older adults ( $M$  = 1.8,  $SD$  = .72) reported being in relatively good health.

### Materials and Apparatus

Items were 46 paired associates. The two words of a given pair were unrelated, concrete nouns. Six of the items (the same for each participant) were used for practice only. Macintosh computers displayed all the instructions and items and recorded all responses during the computer-controlled experiment. The items were presented in 24-point Geneva font.

### Procedure

The procedure included detailed instructions on the task, a practice study trial, and a study–test trial. The participants were told that their task was to learn each pair so that later they would be able to recall the second word of a given pair when prompted with the first word of that pair. For the groups receiving information about the strategies (concurrent group and the informed retrospective group), three strategies commonly used for paired-associate learning were described: sentence generation, rote repetition, and interactive imagery. For each strategy, the description was relatively minimal, including only two examples (with the same two items used in the examples for each strategy) and a one-sentence description of each strategy (e.g., “when you use *sentence generation*, you try to link the two words together by completing a sentence that includes both words”). The instructions did not mention the normative effectiveness of the strategies. The participants were also told that many other kinds of strategy could be used, and that a given person might (or might not) use multiple strategies. They were instructed to do whatever worked best for them. For the uninformed retrospective groups, strategies were not mentioned during the pre-study instructions.

Next, the concurrent group received instructions on how to make ratings. Immediately after the offset of the presentation of a given item for study, the query for a strategy rating appeared: “How did you study the last pair? Type one of the following numbers: 1, if you used interactive imagery; 2, if you used rote repetition; 3, if you used sentence generation; 4, if you used some other strategy; 5, if you used no strategy; and 6, if you tried to use a strategy but ran out of time.”

To ensure that the participants understood their task, we conducted a practice trial. The practice trial included studying six prac-

tice items, and for the concurrent group, it also involved making a rating for each item. Practice did not include paired-associate recall. The participants were encouraged to ask questions if they did not understand any aspect of the procedure.

After the practice trial, the participants began the study–test trial. During study, each item was presented for study for 8 sec (order of presentation was randomized anew for each participant). All participants studied the same pairs during the study–test trial. The concurrent group also made a strategy rating. The ratings were self-paced, and the participants were instructed to make each rating as quickly as possible. Even so, concurrent reports allowed further presentation of an item for study, which in turn could boost paired-associate recall. However, given that the extra time involved massed presentation of the item and that making other judgments similar to those used here does not typically influence memory performance (e.g., Nelson & Dunlosky, 1992), any extra time used to make ratings was expected to have a negligible influence on paired-associate recall.

After the study trial, each of the 40 critical items was presented individually for paired-associate recall in a newly randomized order. For each item, the stimulus was presented, and each participant was instructed to type his or her response on the computer keyboard. The participants pressed the “Return” key to advance to the next item. Although the participants were not forced to provide a response for each stimulus, they were encouraged to guess if they had one. To reduce the role of incorrect spelling, we scored a response as correct if the first three letters were correct.

As in Richardson (1998), retrospective reports were collected immediately upon completion of the test trial. All participants first received the brief descriptions of the strategies that were originally provided to the informed groups. Next, each stimulus–response item was presented individually (with the order of items randomized anew), and the participants indicated the strategy that they had used during study. Even if a participant did not recall a response on the previous paired-associate recall trial, he or she had to make a strategy rating for that item. The prompt for these ratings was identical to the prompt for the concurrent reports.

## RESULTS

The data relevant to evaluating the two main hypotheses involve how often individuals reported using each strategy. Accordingly, we present self-reported strategy production first. Next, we present paired-associate recall performance to demonstrate the validity of the strategy reports and to explore whether production deficiencies contribute to age-related differences in memory. All differences declared as reliable had  $p < .05$ .

### Self-Reported Strategy Production

For each participant, we calculated the proportion of times that he or she reported producing a given strategy. As in Dunlosky and Hertzog (1998), reports of producing no strategy or running out of time were collapsed into one category, called “no strategy.” Means across participants’ proportions for each of the categories are presented in Table 1. Several planned comparisons were conducted to evaluate the inconsistent-reports and reactive-effects hypotheses.

For the inconsistent-reports hypothesis, we contrasted the concurrent reports and retrospective reports from the concurrent group. A 2 (age)  $\times$  2 (kind of report: concu-

**Table 1**  
**Proportion of Items that Participants**  
**Reported Studying by a Given Strategy**

Group	Strategy Reported									
	Imagery		Repetition		Sentence		Other		No Strategy	
	Prop.	SEM	Prop.	SEM	Prop.	SEM	Prop.	SEM	Prop.	SEM
Young adults										
Concurrent report	.38	.05	.20	.04	.29	.04	.07	.01	.06	.01
Retrospective report										
Concurrent group	.44	.06	.19	.05	.26	.05	.07	.02	.04	.01
Informed	.37	.06	.20	.05	.33	.05	.03	.01	.08	.02
Uninformed	.51	.06	.14	.04	.19	.04	.05	.02	.12	.03
Older adults										
Concurrent report	.34	.06	.15	.03	.38	.06	.03	.02	.10	.03
Retrospective report										
Concurrent group	.34	.06	.15	.03	.31	.05	.04	.02	.17	.04
Informed	.31	.06	.15	.04	.29	.06	.03	.01	.23	.05
Uninformed	.28	.05	.29	.07	.12	.03	.05	.02	.28	.06

Note—Values under “Other” include reported production of a strategy other than imagery, repetition, or sentence generation. “No strategy” indicates values collapsed across reports of producing no strategy and reports of running out of time.

rent vs. retrospective)  $\times$  4 (strategy: report of imagery, repetition, sentence, or no strategy) analysis of variance (ANOVA) was conducted. Because the proportion of reported strategies summed to 1.0 across the five categories, we dropped the response category of “using some other strategy”—which was least relevant to our hypotheses—to break the statistical dependence between cells.

The main effects for age and for kind of report, and the interaction between these factors, were not reliable [ $F(1,63) < 2.80$ ,  $MS_e s < .005$ ]. The main effect of strategy was reliable [ $F(3,189) = 13.5$ ,  $MS_e = .16$ ], but the strategy  $\times$  age interaction was not reliable [ $F(3,189) = 1.36$ ,  $MS_e = .16$ ]. Most important, the kind of report  $\times$  strategy interaction [ $F(3,189) = 3.49$ ] and the three-way interaction [ $F(3,189) = 2.93$ ,  $MS_e = .01$ ] were reliable. Follow-up analyses to the three-way interaction were conducted to explore whether strategy reports differed for concurrent and retrospective reports. For younger adults, imagery was reported less often during concurrent as opposed to retrospective reports [ $t(31) = 3.25$ ], whereas using no strategy was reported more often during concurrent reports [ $t(31) = 2.95$ ]. For older adults, sentence generation was reported less often during retrospective as opposed to concurrent reports [ $t(32) = 2.83$ ], whereas using no strategy was reported more often during retrospective reports [ $t(32) = 2.30$ ]. This outcome is consistent with the a priori hypothesis that older adults forget some mediators (in this case, sentence generation), and hence report using no strategy when one had purportedly been used. One interpretation of the age-related differences is that in contrast to younger adults, older adults are more conservative in making responses, and hence, when they forget a mediator, they are less likely to guess which mediator had been used during study. In contrast to this interpretation, the extant literature suggests that older adults are not more conservative (and at times are even more liberal) in making responses during tests of memory (for discus-

sion, see Kausler, 1991, pp. 380–381). Moreover, this interpretation of the age-related differences does not qualify the apparent forgetting of mediators by older adults.

In summary, results from both age groups confirm the inconsistent-reports hypothesis, although the general consistency in reported strategy production appeared to be relatively high. One caveat to this conclusion pertains to averaging the proportion of strategy reports across individuals. Such averaging may obscure intraindividual shifts in strategy reports, which would otherwise provide stronger evidence against the consistency of retrospective reports. Accordingly, we also evaluated the inconsistent-reports hypothesis by computing the consistency of reports within individuals in the concurrent group. For each participant, we computed the percentage of items that were matched with the same strategy during the concurrent reports and retrospective reports. The mean across participants was 64% ( $SEM = 3.5$ ) for younger adults and 61% ( $SEM = 3.3$ ) for older adults, showing moderate intraindividual inconsistency across the reports. These results indicate that retrospective reports do not completely reflect the kinds of strategy that an individual purportedly has used during study.

To evaluate the reactive-effects hypothesis, we compared the retrospective reports from the three groups by conducting separate analyses to evaluate two planned comparisons. First, to discover whether making concurrent reports had a reactive effect on subsequent reports, we compared the retrospective reports from the concurrent group and the informed retrospective groups. In a 2 (group)  $\times$  2 (age)  $\times$  4 (strategy) ANOVA, the main effect of strategy [ $F(3,369) = 16.0$ ] and the age  $\times$  strategy interaction [ $F(3,369) = 3.13$ ,  $MS_e = .095$ ] were reliable, with the interaction indicating that older adults (vs. younger adults) more often reported producing no strategies [ $t(125) = 3.96$ ]. Most important, the main effect of group and all interactions involving group were not statistically reliable

**Table 2**  
**Proportion of Correct Paired-Associate Recall Performance as**  
**a Function of the Strategies Participants Reported Producing During Study**

Group	Strategy Reported									
	Imagery		Repetition		Sentence		Other		No Strategy	
	Prop.	SEM	Prop.	SEM	Prop.	SEM	Prop.	SEM	Prop.	SEM
Young adults										
Concurrent report	.71	.04	.32	.05	.63	.05	.65	.07	.42	.08
Retrospective report										
Concurrent group	.76	.04	.29	.07	.61	.06	.43	.08	.25	.09
Informed	.81	.03	.38	.09	.81	.04	.81	.12	.16	.07
Uninformed	.78	.03	.19	.06	.65	.07	.59	.12	.12	.07
Older adults										
Concurrent report	.37	.06	.09	.03	.33	.04	.23	.13	.18	.07
Retrospective report										
Concurrent group	.39	.06	.08	.04	.38	.05	.19	.12	.06	.03
Informed	.45	.06	.08	.05	.43	.07	.42	.14	.03	.02
Uninformed	.42	.06	.05	.03	.26	.08	.36	.14	.01	.01

Note—Values under “Other” include reported production of a strategy other than imagery, repetition, or sentence generation. “No strategy” indicates values collapsed across reports of producing no strategy and reports of running out of time.

[ $F_s < 2.75$ ,  $MS_{e,s} < .10$ ]. Thus, making concurrent reports had a negligible impact on retrospective reports.

Second, to discover whether providing information about strategies had a reactive effect on subsequent reports, we compared the retrospective reports from the informed retrospective group and the uninformed retrospective group. The 2 (group)  $\times$  2 (age)  $\times$  4 (strategy) ANOVA revealed a reliable main effect of strategy, a strategy  $\times$  age interaction, and a strategy  $\times$  group interaction [ $F_s(3,369) > 3.30$ ,  $MS_{e,s} = .10$ ]. (All other effects were not reliable,  $F_s < 1.90$ ,  $MS_{e,s} < .11$ .) Follow-up tests uncovered several straightforward, interpretable findings. For younger adults, the trends were to report more imagery [ $t(64) = 1.69$ ,  $p = .10$ ] and less sentence generation [ $t(64) = 2.17$ ] after receiving no strategy information than after receiving strategy information. By contrast, older adults reported more repetition [ $t(59) = 1.79$ ,  $p = .08$ ] and less sentence generation [ $t(59) = 2.62$ ] after receiving no strategy information than after receiving strategy information. Thus, at least for older adults, providing descriptions of the strategies increased the likelihood that they would report using a normatively effective strategy during study.

### Proportion of Correct Performance for Paired-Associate Recall

Analyses of paired-associate recall were aimed at accomplishing several interrelated goals. We first describe whether making concurrent reports (or receiving strategy descriptions) affected younger adults' performance, which pertains to the reactive effects of the reports. Next, we evaluate a production-deficiency hypothesis by conducting a planned comparison on overall performance for older adults. In the final section, we explore further the validity of the strategy reports.

**Performance collapsed across strategy reports.** For each participant, we computed the proportion of items

correctly recalled. For younger adults, correct paired-associate recall did not differ among the three groups [ $F(2,95) = 0.23$ ,  $MS_e = .059$ ], with means across individuals' proportions being .57 ( $SEM = .04$ ) for the concurrent group, .61 ( $SEM = .04$ ) for the informed retrospective group, and .59 ( $SEM = .05$ ) for the uninformed retrospective group.

For the older adults, a planned comparison was conducted between the uninformed group and the two informed groups. As expected, performance was less for the uninformed retrospective group (.17) than for the concurrent group (.29) and the informed retrospective group (.27) [planned  $F(1,91) = 7.66$ ,  $MS_e = .032$ , all  $SEMs = .03$ ]. These data indicate that a production deficiency constrains performance when older adults do not receive strategy descriptions prior to study. Just as important, making concurrent reports did not have a reactive effect on subsequent performance for either age group, which appears inconsistent with Richardson's (1998) concern that making concurrent reports would impair the discovery of strategies and hence constrain memory performance.

**Relation between reported strategy production and paired-associate recall.** For each participant, we computed the proportion of correct recall performance as a function of reported strategy. Means across individual participants' values are reported in Table 2.

Separate analyses were conducted for concurrent reports and for retrospective reports. For the former, a 2 (age)  $\times$  3 (concurrent strategy: imagery, repetition, and sentence generation) ANOVA was conducted.<sup>2</sup> The main effects of age [ $F(1,38) = 28.7$ ,  $MS_e = .082$ ] and strategy [ $F(2,76) = 20.8$ ,  $MS_e = .041$ ] were reliable, whereas the interaction was not reliable [ $F(2,76) = 1.66$ ,  $MS_e = .041$ ]. As was expected from previous research, paired-associate recall was greater for items reportedly studied using imagery and sentence generation as opposed to repetition [smaller of two  $t_s(46) = 5.80$ ], but it did not differ for items

reportedly studied using imagery and sentence generation [ $t(45) = 1.29$ ].

For paired-associate recall as a function of retrospective reports, a 2 (age)  $\times$  3 (retrospective strategy: imagery, repetition, and sentence generation)  $\times$  3 (group) ANOVA revealed main effects of age [ $F(1,73) = 39.0$ ,  $MS_e = .129$ ], group [ $F(2,73) = 3.25$ ,  $MS_e = .129$ ], and strategy [ $F(3,146) = 63.1$ ,  $MS_e = .053$ ]. All interactions were not reliable ( $F_s < 2.75$ ). Although performance also tended to be greater for items reportedly studied by imagery than by sentence generation [ $t(119) = 2.88$ ], this trend was not evident across all groups. Most important, performance was greater for items reportedly studied with imagery and sentence generation than with repetition [smaller of two  $t_s(87) = 10.6$ ]. Thus, both concurrent and retrospective reports demonstrated construct validity.

## DISCUSSION

Evidence from the present research provides insight into the strengths and weaknesses of both concurrent reports and retrospective reports. First, consider concurrent reports. In contrast to intuition, concurrent reports had negligible reactive effects on strategy production and on paired-associate recall. These outcomes somewhat obscure one difficulty, however, because the strategy descriptions that are arguably necessary to collect concurrent reports had reactive effects. For younger adults, these effects were minimal: Those who did not receive strategy descriptions tended to report using more imagery than did those who received them, but recall did not differ between these groups. By contrast, the reactive effects of informing older adults were more substantial. Those who did not receive strategy descriptions reported fewer effective strategies and had lower performance than did those who received them.

Concerning retrospective reports, we examined the degree to which retrospective and concurrent reports were consistent. Given the hypothesis that forgetting constrains the validity of retrospective reports, the two kinds of report were expected to diverge.<sup>3</sup> When the reported strategies are computed across individuals (Table 1), it can be seen that substantial consistency occurred, suggesting that retrospective reports largely portray those strategies purportedly used during study. Even so, when inconsistencies in reports within individual participants are examined, on average retrospective reports agreed with concurrent reports for only 64% of the items for younger adults and 61% of the items for older adults. Also, although age-related differences were not substantial, older adults' reports provided further evidence that the validity of retrospective reports was constrained by forgetting. Namely, when making retrospective reports, older adults more often reported using no strategy and less often reported using sentence generation. A general conclusion here is that estimates of strategy production should be interpreted with caution when they are derived from retrospective reports from a sample of participants who are

prone to episodic memory deficits, such as older adults, children, Alzheimer's patients, amnesiacs, and so on.

What is the relative utility of the two kinds of strategy reports? An answer from the present research is that neither kind of report is uniformly superior under all circumstances. Also important is that our evidence provides recommendations concerning when to use one report versus the other. Because concurrent reports have minimal reactive effects, when research questions are not undermined by providing descriptions of strategies, concurrent reports should be used. Certainly, when the focal subject population is healthy younger adults, concurrent reports of strategy production appear to be most appropriate. In many cases, however, providing strategy descriptions would be inappropriate, as in investigating spontaneous generation of strategies during incidental learning, evaluating some production-deficiency hypotheses, and so on. In these cases, retrospective reports would be more appropriate, especially if the delay between study and the subsequent reports was minimized. If memory performance is less central to the research questions, this delay could be reduced even further by collecting retrospective reports immediately after all items have been studied. Of course, when this delay must be long, the tradeoff between the reactive effects of instructions and forgetting of strategies will need to be considered. An alternative approach would be to employ a design in which both kinds of report are used, as in the present experiment, so that reactive effects and forgetting can be assessed.

Used judiciously, item-level strategy reports provide data valuable for testing hypotheses of human memory and learning. One such hypothesis concerns explaining age-related differences in associative learning. The production-deficiency hypothesis is that in comparison with younger adults, older adults are less likely to produce effective strategies, which in turn constrains their memory performance. Several outcomes converge on the conclusion that a production deficiency contributes to age-related differences in memory performance. Paired-associate recall was lower for older adults who did not receive strategy descriptions than for those who did receive them. Also, those who were not given strategy descriptions reported using less sentence generation and more often reported using rote repetition or no strategy. Why might a production deficiency exist? If some deficit in central resources limits older adults' production of strategies (see, e.g., Salthouse, 1991), one would not expect that the production deficiency would be overcome with simple strategy descriptions and without instructions to use a specific strategy. However, the data presented here and in our previous research (Dunlosky & Hertzog, 1998) demonstrate that merely producing several strategies to older adults minimizes production deficiencies. Such findings may implicate metacognitive explanations of the deficiency (e.g., less prior knowledge of strategies).

In conclusion, the informed use of subjective reports has great potential benefit for evaluating critical aspects

of human learning. And although researchers have known for decades about the overall utility of item-by-item strategy reports, a systematic comparison of the most widely used reports—concurrent and retrospective—has not previously been conducted. Our research not only provides guidance concerning when to use each report but highlights how strategy reports can be central to evaluating hypotheses about human memory.

#### REFERENCES

- ADAMS, J. A., & MCINTYRE, J. S. (1967). Natural language mediation and all-or-none learning. *Canadian Journal of Psychology*, **21**, 436-449.
- ADAMS, J. A., & MONTAGUE, W. E. (1967). Retroactive inhibition and natural language mediation. *Journal of Verbal Learning & Verbal Behavior*, **6**, 528-535.
- DUNLOSKY, J., & HERTZOG, C. (1998). Aging and deficits in associative memory: What is the role of strategy production? *Psychology & Aging*, **13**, 597-607.
- EKSTROM, R. B., FRENCH, J. W., HARMAN, H. H., & DERMEN, D. (1976). *Manual for kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.
- ERICSSON, K. A., & SIMON, H. A. (1980). Verbal reports as data. *Psychological Review*, **87**, 215-251.
- KAUSLER, D. H. (1991). *Experimental psychology, cognition, and human aging* (2nd ed.). New York: Springer-Verlag.
- LIGHT, L. L. (1991). Memory and aging: Four hypotheses in search of data. *Annual Review of Psychology*, **42**, 333-377.
- MONTAGUE, W. E., ADAMS, J. A., & KIESS, H. O. (1966). Forgetting and natural language mediation. *Journal of Experimental Psychology*, **72**, 829-833.
- NELSON, T. O., & DUNLOSKY, J. (1992). How shall we explain the delayed-judgment-of-learning effect? *Psychological Science*, **3**, 317-318.
- RICHARDSON, J. T. E. (1998). The availability and effectiveness of reported mediators in associative learning: A historical review and an experimental investigation. *Psychonomic Bulletin & Review*, **5**, 597-614.
- ROGERS, W. A., HERTZOG, C., & FISK, A. D. (2000). An individual differences analysis of ability and strategy influences: Age-related differences in associative learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 359-394.
- SALTHOUSE, T. A. (1991). *Theoretical perspective on cognitive aging*. Hillsdale, NJ: Erlbaum.

#### NOTES

1. These and other planned comparisons involving reported strategy production focus on interactions between conditions (e.g., between concurrent and retrospective reports) and the strategies produced. To understand why, consider that the proportions of strategies reported across all items sum to 1.0. Thus, for two conditions to differ in strategy reports, if a given condition leads to more reports of a strategy (e.g., imagery) than does another condition, then the former will likely lead to reporting less of another strategy (e.g., rote repetition), which would be manifested in a condition  $\times$  strategy interaction.

2. As in Dunlosky and Hertzog (1998), the ANOVA for these analyses included three strategies (imagery, repetition, and sentence generation), which have been the focus of past research on self-reported strategy production (Richardson, 1998). Because strategy reported is a within-subject variable, a missing value from any cell for a given participant results in dropping that participant from the repeated measures analyses. Most individuals had a missing value for either the "other" category or the "no-strategy" category, and hence including these categories left few participants in the analyses. Even when the analysis was reduced to three strategies, participants were still excluded. Most important, however, is that the outcomes from the statistical analyses are consistent with the trends evident from inspection of Table 2.

3. Even if making concurrent reports has reactive effects on strategy reports, concurrent reports likely provide the most valid measure of the kind of mediators that participants had produced during study because the mediator produced for an item would still be heeded in working memory when the concurrent report was made (Ericsson & Simon, 1980). Thus, unless one makes the (unreasonable) assumption that retrospective reports from the concurrent group reflect the kinds of strategies that participants would have produced if they had not made concurrent reports, the consistency between concurrent and retrospective reports provides one indicator of the validity of retrospective reports.

(Manuscript received September 16, 1999;  
revision accepted for publication July 9, 2000.)