

# The generation effect: A meta-analytic review

SHARON BERTSCH

University of Pittsburgh, Johnstown, Pennsylvania

BRYAN J. PESTA

Cleveland State University, Cleveland, Ohio

RICHARD WISCOTT

Shippensburg University, Shippensburg, Pennsylvania

AND

MICHAEL A. MCDANIEL

Virginia Commonwealth University, Richmond, Virginia

---

The *generation effect* refers to the finding that subjects who generate information (e.g., produce synonyms) remember the information better than they do material that they simply read. Meta-analytic techniques were used to summarize 445 effect sizes over 86 studies, thereby assessing the magnitude and 11 potential moderators of the generation effect. The size of the generation effect across the 86 studies was .40—a benefit of almost half a standard deviation of generation over reading. The variability of the effect size due to moderator type was substantial, and we attempted to use this information to clarify several theories that have been proposed to explain the generation effect.

---

The *generation effect* is the experimental finding that when a subject is asked to generate all or part of a stimulus item, that item is almost always remembered better than material the subject only read (Jacoby, 1978; Slamecka & Graf, 1978). The proportion of the number of previously generated items to the previously read items that were remembered constitutes the size of the generation effect. Over the last 20-plus years, a substantial body of research has evolved around this seemingly simple cognitive task. Nevertheless, controversy still exists over many of the particulars of the generation effect, including its true magnitude (see, e.g., J. C. Brown, Niinikoski, & Duke, 1993; Toth & Hunt, 1990), the underlying cognitive processes that are responsible for it (e.g., Fiedler, Lachnit, Fay, & Krug, 1992; Gardiner, Gregg, & Hampton, 1988), the exact experimental conditions that are required to produce it (e.g., Nairne, Riegler, & Serra, 1991), the influences that moderate its size (e.g., Peynircioğlu & Mungan, 1993; Reardon, Durso, Foley, & McGahan, 1987), and even the conjecture as to whether it is real or merely an experimental design artifact (e.g., Slamecka & Katsaiti, 1987). Using the techniques of meta-analysis, the goal of the present article is to address two of these questions: the true population magnitude of the generation effect and the degree to which suggested moderator variables influence the size of the effect.

The basic generation effect paradigm involves the presentation of some type of paired-associates list to subjects. Nonsense word pairs, number and letter bigrams, word lists, and mathematical equations are some of the most common examples. Half the pairs are provided intact by the experimenter, and the subject is instructed to simply read the pair (e.g., COLD, HOT). For the remaining items, the subjects are presented with the first half of the pair intact (COLD, \_\_\_\_\_) and they are also provided with a rule that they must use to generate the second half of the pair (e.g., creation of synonyms, rhymes, or various category generation rules). Variations of this paradigm include presenting complete sentences whose last word is either read or generated (e.g., Peynircioğlu & Mungan, 1993); reading or completing multiplication (e.g., Pesta, Sanders, & Nemeč, 1996) or addition (e.g., McNamara & Healy, 2000); and providing anagrams whose solutions are intact or jumbled (e.g., Gardiner, Dawson, & Sutton, 1989).

Tests of recognition, cued recall, or free recall for the read and/or generated stimuli are conducted following the learning trials. The retention tests are scored as the proportion of generated and the proportion of read items correctly remembered out of the total tested. When the difference score is calculated (generate minus read), the resultant number indicates the memory benefits (in per-

---

S. Bertsch, bertsch@pitt.edu

---

centage terms) that self-generation of material had during the study phase.

Part of the difficulty in explaining the cognitive processes involved in the generation effect stems from the varied and sometimes conflicting results obtained in primary studies. For example, Fiedler et al. (1992, Experiment 1) found that the effect was larger when the subjects were required to generate more of the target word (i.e., completion blank vs. word fragment). Yet, a significant generation effect has also been obtained when simple generation rules, such as adding the letter "e" to the end of each word fragment (Donaldson & Bass, 1980) or switching letters (e.g., Nairne & Widner, 1987), have been used.

Even conclusions based on whether the read/generate condition was manipulated within or between subjects have been unclear. In many studies, a between-subjects design has been shown to reduce or even eliminate the generation effect altogether (e.g., Hertel, 1989; Schmidt, 1992). However, Kane and Anderson (1978), who used a between-subjects manipulation, found a generation effect comparable to that found by Gardiner (1989), who used a within-subjects design. In addition, Hara, Neumann, and Tajika (1989), who used a between-subjects experimental design in one experiment (Experiment 2), actually found a larger generation effect than that which they found in an earlier within-subjects experiment (Experiment 1).

Discrepancies can often be explained by the concomitant experimental manipulations used by researchers. Tyler, Hertel, McCallum, and Ellis (1979) found that the difficulty of word stem completion moderated the generation effect that was obtained: More difficult word stem completions were remembered better than less difficult ones. Other potential experimental design factors that are believed to moderate generation effect findings include level of encoding (e.g., Soraci et al., 1994), type of encoding (i.e., rhyme vs. sentence completion; McFarland, Frey, & Rhodes, 1980), and type of test (i.e., recognition vs. free or cued recall; e.g., Gardiner et al., 1989; Schmidt, 1992).

The goal of this review was to estimate the true population effect size of the generation effect as found by the cumulation of individual (or primary) studies. Using psychometric meta-analytic techniques to summarize past findings (Hunter & Schmidt, 1990), we assessed the magnitude and potential moderators of the generation effect.

## METHOD

Meta-analysis procedures are based on the assumption that much of the variability across studies is due to sampling error—a function of both the population effect size of the characteristic of interest and the sample size of the individual study. Estimates of the population effect size calculated from individual studies (i.e., correlation coefficients, standardized mean differences, and percentage differences) are distinguished from other statistics ( $t$  or  $F$ ) in that the magnitude of the effect size is not a function of the sample size of the study. Sampling error causes the results of some studies to over- or underestimate the population parameter of interest. This random variation in results across studies often causes readers of the literature to conclude falsely that the research results are mixed. Correcting for variance due to

sampling error across studies often allows researchers to show that results are much more consistent than previously believed. Removing sampling error from the distribution of effect sizes across studies also permits researchers to evaluate more accurately how much remaining variance can be attributed to substantive causes, such as moderators.

In the psychometric meta-analytic approach (Hunter & Schmidt, 1990), it is recognized that artifacts other than sampling error (e.g., measurement error, range variation) can make the observed results biased estimates of the population. These artifacts are often the result of using small  $n$  sizes. Unfortunately, published research often lacks information regarding the reliability of the measures used, which makes it impossible to correct for the primary (nonsampling) error artifact—measurement error in the dependent variable. This error contributes to variance across studies (studies with reliable measures, on average, yield effect sizes of larger magnitude than those of studies with less reliable measures) and leads to the underestimation of the mean population effect. Our meta-analysis was only able to correct for sampling error; that is, our present estimates of the population variance are overestimates of the actual population variance, and our resulting estimates of the population mean are downwardly biased.

Analyzing multiple effect sizes from the same sample may also downwardly bias the sampling error variance estimate, which could also result in the underestimation of the population effect size. For example, including results from a sample that was used in a study with a within-subjects design, incidental learning, and a recognition test contributes three separate generation effect sizes to the meta-analysis on the basis of a single sample of subjects. We therefore calculated our estimation of the population parameter of the generation effect twice—once allowing every experiment to contribute one mean generation effect weighted by sample size (adjusted analysis results), and a second time using all the information contributed by all experiments, each individually weighted by sample size (overall analysis results). If the two differed by a substantial amount, the method that produced the most conservative estimate was used in the moderator analyses. If both methods produced similar estimations, then the method that included all generation effect information from all studies was used.

## Moderators

Moderators were selected on the basis of the popularity of the variable in the primary studies of the generation effect. Studies were coded for the type of memory test used (recognition, cued recall, or free recall); whether or not subjects were forewarned of the upcoming memory test (intentional or incidental learning); the manipulation of the read/generate condition (within or between subjects); the age group of the subjects (older adults, typically over 65 years old; or younger adults, generally college students); the type of list organization used in the presentation of the stimuli (blocked or randomized); the type of stimuli used (numbers, words, or nonwords); the type of generation rule used by the subjects (anagram, rhyme, associations, category membership, sentence completion, calculation, synonym, word fragment completion, antonym, or letter rearrangement); how much of the target was generated by the subject (part or whole); the total number of stimuli seen by the subject (<25, 25–50, or >50); the retention interval between learning and test (immediate, <60 sec, 60 sec to 1 day, or >1 day); and the difficulty of the generate task, as judged by the authors (easy, moderate, or hard). *Easy* tasks were those such as letter rearrangement, in which the subject had to switch only two underlined letters. *Moderate* tasks included simple math calculations and word fragment completions. *Hard* tasks were those such as mental multiplication. Questions about the difficulty status of particular tasks were resolved through discussion.

The literature search for individual studies included the PsycINFO database (1966–2005), reference lists of all relevant articles, and contact of primary researchers in the field for "file-drawer" studies. These searches resulted in usable data from 86 studies with a total of 445 measures of the generation effect size on the basis of 17,711 subjects,

some of which were derived from nonindependent samples within a single experiment. Primary studies were excluded for a number of reasons, including the use of memory tests other than free recall, cued recall, or recognition (e.g., J. C. Brown et al., 1993; Java, 1996); bilingual subjects (e.g., O'Neill, Roy, & Tremblay, 1993); subjects from clinical samples (e.g., Pring, 1988); or pictures as stimuli (e.g., Peynircioğlu, 1989). In addition, only the data from initial tests were used in studies that reported the results of multiple-trial testing (e.g., McFarland, Warren, & Crockard, 1985). Studies were excluded if they lacked the necessary statistical information (e.g., Mitchell, Hunt, & Schmitt, 1986), or if they reported results of read-only or generate-only conditions, without a comparison group in the other condition (e.g., A. S. Brown & Mitchell, 1991). Although some studies provided usable read and generate proportions, they lacked information with respect to one or more moderator variables. These studies were used only in moderator analyses for which information was provided.

A total of 13 meta-analyses were done: 1 that included all read and generate condition information (overall analysis), 1 that included an average for each condition from each experiment (adjusted analysis), and 11 that were grouped by moderator class. Each primary study provided two proportions: the proportion remembered correctly of the initially read items and the proportion correct of the initially generated items. Additionally, for each experiment the sample size ( $n$ ) and the total number of generation effects to which a particular sample contributed were obtained.

Very few meta-analyses have been conducted using proportionate data. We based our analyses on one such study—Viswesvaran and Schmidt (1992)—although our data set differed slightly from theirs. In their study, calculations were based on the cumulation of single proportions (quitting rates for smokers), whereas we calculated the generation effect as a difference score between two proportions (read and generate). Because difference scores are notoriously unreliable (they inflate the amount of variance that is due to error), effect sizes calculated on the basis of this variance without correction are inaccurate reflections of the population. Therefore, the formulas that Viswesvaran and Schmidt used may not correct for sampling error as precisely when they are used on the statistic with which generation effect researchers are most familiar.

We therefore performed the overall meta-analysis twice. In one, we used the Viswesvaran and Schmidt (1992) formulas to calculate sampling-error-corrected effect sizes on the basis of the cumulation of difference scores. In the second, we used these statistics to cumulate read and generate proportions separately. Observed and sampling-error variances were also calculated separately for read and generate conditions. These variances were often found to be substantial within the same meta-analytic group; for example, among studies in which numbers were used as stimuli, the variance for the read conditions was .039, whereas the variance for the generate condition was .024. Therefore, when calculating effect sizes for this type of analysis, we treated the read condition as a control group and used the sampling-error-corrected standard deviation of the entire collection of read proportions as an estimate of the standard deviation for the control group population (Glass, 1977). Our final estimates should then indicate the standardized size of the benefit of the generate condition over the control (read) condition.

Because the variance of difference scores is smaller than that of individual scores, the first type of analysis led to effect sizes that were approximately twice those of the analysis that cumulated read and generate conditions separately. The pattern of change in the effect sizes in moderator group comparisons, however, was largely consistent across both types of analyses.<sup>1</sup> We therefore chose to report the results of the second type of analysis, since it avoided the pitfalls associated with difference scores. For each meta-analysis, then, the mean sample-size-weighted proportion for read and generate conditions ( $\bar{P}$ ) across the included studies was computed as follows:

$$\frac{\sum nP}{\sum n}, \quad (1)$$

where  $P$  was the proportion correct in read or generate conditions in each experiment. The sample-size-weighted variance of the read and generate conditions in the group was given by

$$\frac{\sum n(P - \bar{P})^2}{\sum n}. \quad (2)$$

The formula for sample-size-weighted sampling error variance (of proportions) was computed as

$$\frac{\sum PQ}{\sum n}, \quad (3)$$

where  $Q = 1 - P$ . This error variance term was then subtracted from the total weighted variance of the group, and the result was an estimate of the amount of variance that could not be explained by sampling error plus other artifacts.

We then estimated the population effect size ( $d$ ) of the generation effect in each collection of studies by taking the difference between the mean proportions for the read and generate conditions in that group and dividing it by the corrected standard deviation of all read proportions representing the control group (Glass, 1977). Our final estimates indicate the standardized size of the benefit of generating in different conditions over the control (read) condition, remembering that even our estimate of the variance in the control group will be larger than the population (true) variance due to remaining uncorrectable artifacts. Thus, our final effect sizes (which we achieved through use of this estimate) still underestimate the true effect size of the population generation effect.

Mean differences in effect sizes, grouped by moderator type, were potentially influenced by second-order sampling-error variance because of the sampling of primary studies included in the analyses. Two methods are available to assess the likelihood that a particular moderator variable had a genuine effect in the population: (1) testing of the statistical significance of the mean effect sizes and (2) comparison of confidence intervals built around those means. We evaluated the probability that moderator variables had an effect by using confidence intervals. Nonoverlapping 95% confidence intervals for group means (split by moderator type) were taken as evidence that the moderator does influence population effect size. Confidence intervals for each effect size were calculated as follows:

$$d \pm 1.96(S), \quad (4)$$

where  $d$  was an observed mean effect size and  $S$  was the square root of the corrected observed variance of the control group divided by the number of studies included (Hunter & Schmidt, 1990, p. 437).

## RESULTS AND DISCUSSION

The results of our analyses are presented in Table 1, which lists the results of the overall and adjusted analyses, plus those of the 11 moderators. The number of generation effects contributing to the calculation of the effect size for this statistic within each group is listed, along with the total  $n$  size, the  $n$ -size-weighted means for the read and generate proportions, the  $n$ -size-weighted total variance, the  $n$ -size-weighted sampling-error variance, the percentage of observed variance accounted for by sampling error, the  $n$ -size-weighted generation effect for that group, the effect size, and the 95% confidence interval. These analyses indicate that the generation effect is a robust and consistent finding. Calculations made with 17,711 subjects revealed that there was almost a one-half standard deviation advantage (.40) in memory performance when material was generated versus when it was just read (remembering that this is actually an underestimation of the population

**Table 1**  
**Read and Generate Effect Size Data**

Group	No. of Generation Effects	<i>n</i> Size	<i>n</i> -Size-Weighted $\bar{P}$	<i>n</i> -Size-Weighted Total Variance	<i>n</i> -Size-Weighted Sampling-Error Variance	Percentage of Total Variance Accounted For	<i>n</i> -Size-Weighted Generation Effect	Effect Size	95% Confidence Interval
Overall	445	17,711							
Read			.471	.054	.006	11.1	.088	.40	.38-.42
Generate			.559	.054	.006	11.1			
Adjusted	276	11,043							
Read			.447	.052	.006	11.5	.091	.41	.39-.43
Generate			.556	.050	.006	12.0			
Type of Test									
Recognition	139	4,980							
Read			.66	.056	.007	12.5	.10	.46	.44-.48
Generate			.76	.058	.007	12.1			
Cued Recall	91	3,640							
Read			.56	.057	.006	10.5	.12	.55	.53-.57
Generate			.68	.058	.006	10.3			
Free Recall	215	9,091							
Read			.33	.049	.006	12.2	.07	.32	.30-.34
Generate			.40	.051	.006	11.8			
Type of Learning									
Intentional	275	11,430							
Read			.49	.055	.006	10.9	.07	.32	.30-.34
Generate			.56	.056	.006	10.7			
Incidental	149	5,481							
Read			.44	.050	.007	14.0	.14	.65	.63-.67
Generate			.58	.049	.007	14.3			
Design									
Within Subjects	306	9,170							
Read			.45	.058	.008	13.8	.11	.50	.48-.52
Generate			.56	.056	.008	14.3			
Between Subjects	138	8,517							
Read			.50	.049	.004	8.2	.06	.28	.26-.30
Generate			.56	.053	.004	7.5			
Subject Population									
Older Adults	18	454							
Read			.34	.077	.01	13.0	.11	.50	.48-.52
Generate			.45	.086	.01	11.6			
Younger Adults	427	17,257							
Read			.47	.053	.006	11.3	.09	.41	.39-.43
Generate			.56	.054	.006	11.1			
List Presentation									
Blocked	202	10,371							
Read			.49	.048	.005	10.4	.07	.32	.30-.34
Generate			.56	.050	.005	10.0			
Random	223	6,736							
Read			.47	.060	.008	13.3	.12	.55	.53-.57
Generate			.59	.058	.008	13.8			
Stimulus Type									
Numbers	36	1,240							
Read			.34	.039	.007	17.9	.19	.87	.85-.89
Generate			.53	.024	.007	29.2			
Words	362	14,710							
Read			.48	.054	.006	11.1	.09	.41	.39-.43
Generate			.57	.054	.006	11.1			
Nonwords	47	1,761							
Read			.47	.062	.007	11.3	.01	.05	.03-.07
Generate			.48	.078	.007	9.0			
Generate Rule									
Anagram	18	1,083							
Read			.47	.031	.004	12.9	-.01	-.05	-.07 to -.03
Generate			.46	.027	.004	14.8			
Rhyme	51	1,693							
Read			.45	.053	.008	15.1	.10	.46	.44-.48
Generate			.53	.063	.008	12.7			
Association	67	3,214							
Read			.46	.061	.005	8.2	.07	.32	.30-.34
Generate			.53	.056	.005	8.9			

Table 1 (Continued)

Group	No. of Generation Effects	<i>n</i> Size	<i>n</i> -Size-Weighted $\bar{P}$	<i>n</i> -Size-Weighted Total Variance	<i>n</i> -Size-Weighted Sampling-Error Variance	Percentage of Total Variance Accounted For	<i>n</i> -Size-Weighted Generation Effect	Effect Size	95% Confidence Interval
Generate Rule (Continued)									
Category	31	1,419							
Read			.56	.044	.005	11.4	.09	.41	.39-.43
Generate			.65	.065	.005	7.7			
Sentence Completion	46	1,961							
Read			.52	.043	.006	14.0	.13	.60	.58-.62
Generate			.65	.043	.006	14.0			
Calculation	32	1,136							
Read			.35	.040	.007	17.5	.20	.92	.90-.94
Generate			.55	.022	.007	31.8			
Synonym	15	346							
Read			.44	.078	.011	14.1	.09	.41	.39-.43
Generate			.53	.073	.011	15.1			
Word Fragment	92	3,889							
Read			.48	.059	.006	10.2	.08	.37	.35-.39
Generate			.56	.056	.006	10.7			
Antonym	18	600							
Read			.37	.060	.008	13.3	.06	.28	.26-.30
Generate			.43	.061	.008	13.1			
Letter Rearrangement	75	2,370							
Read			.48	.061	.008	13.3	.08	.37	.49-.53
Generate			.56	.070	.008	11.4			
Stimuli Generated									
Part	283	11,645							
Read			.47	.057	.006	10.5	.08	.32	.30-.34
Generate			.54	.057	.006	10.5			
Whole	162	6,066							
Read			.48	.047	.007	14.9	.12	.55	.53-.57
Generate			.60	.050	.007	14.0			
Number of Stimuli									
25 or Fewer	138	4,855							
Read			.44	.058	.007	12.1	.13	.60	.58-.62
Generate			.57	.058	.007	12.1			
26 to 50	204	8,471							
Read			.44	.056	.006	10.7	.09	.41	.39-.43
Generate			.53	.061	.006	9.8			
More Than 50	101	4,241							
Read			.58	.045	.006	13.3	.02	.09	.07-.11
Generate			.60	.038	.006	15.8			
Retention Interval									
Immediate	176	7,275							
Read			.50	.058	.006	10.3	.09	.41	.39-.43
Generate			.59	.059	.006	10.2			
Up to 1 min	111	4,740							
Read			.46	.047	.006	12.8	.07	.32	.30-.34
Generate			.53	.051	.006	11.8			
1 min to 1 day	115	4,404							
Read			.44	.058	.006	10.3	.09	.41	.39-.43
Generate			.53	.055	.006	10.9			
More Than 1 day	30	971							
Read			.48	.032	.008	25.0	.14	.64	.62-.66
Generate			.62	.031	.008	25.8			
Generation Difficulty									
Easy	257	9,638							
Read			.49	.055	.007	12.7	.09	.41	.39-.43
Generate			.58	.053	.007	13.2			
Moderate	171	7,388							
Read			.44	.050	.006	12.0	.08	.37	.35-.39
Generate			.52	.054	.006	11.1			
Hard	17	685							
Read			.49	.059	.006	10.2	.10	.46	.44-.48
Generate			.59	.071	.006	8.5			

Note— $\bar{P}$ , mean percent correct.

effect). In comparison, S. M. Smith and Vela (2001) found that the influence of contextual reinstatement on memory had an effect size of .28, and Christensen-Szalanski and Willham (1991) calculated the effect of hindsight bias as  $r = .17$  (which converts to  $d = .34$ ).

The differences between the overall and adjusted analyses are extremely small. Both have similar average generation effects (.088 vs. .091), similar effect sizes (.40 vs. .41), and overlapping confidence intervals. Both leave approximately the same amount of variance to be explained by moderators after corrections are made for sampling error (88%–89%). Because these differences are small, we chose to use the data set that would allow all the read and generate proportions to contribute to the moderator analyses.

In our overall analysis, sampling error only accounted for a small amount of variance in effect size (11%). Even across analyses grouped by moderator (some of which included a relatively small number of effects), the influence of sampling error was smaller than we expected (7.5%–31.8%). The small amount of variance accounted for in these analyses may represent moderators used in individual designs that we did not code for, or may be the result of the “bare bones” nature of this work (Hunter & Schmidt, 1990, p. 293).

The moderators we included showed varying amounts of influence on the effect size of the generation effect. The type of memory test showed variation from .46 for recognition testing, .55 for cued recall, and .32 for free recall. Incidental learning conditions resulted in an effect size (.65) that was more than twice as large as the one obtained under intentional-learning conditions (.32). Within-subjects designs resulted in an effect size (.50) approximately twice the effect size for between-subjects designs (.28). Analysis of generation effect experiments with subjects grouped by age showed small differences; the comparison was made difficult because of the large difference in contributing effect sizes (18 studies used older adults, 427 used younger ones). The data presently indicate that the older adults had an edge in effect size over the younger adults (.50 and .41, respectively).

The type of stimulus presentation also influenced the effect size: Random (or mixed) list format (.55) outperformed blocked (or mixed) presentation (.32). The type of stimuli also influenced effect size; numbers showed the largest read–generate differences (.87), followed by words (.41) and nonwords (.05). All the generation effect tasks that we examined resulted in strong positive effects (.28–.92), with the exception of anagrams. Across 18 studies totaling over 1,000 subjects, the use of anagrams resulted in a negative generation effect (i.e., read conditions were more effective than generate conditions). Although it is a small effect (–.05), the confidence interval does not include zero and is therefore likely to represent real population differences. Even simple rules, such as the rearrangement of the order of two underlined letters, produced a strong effect (.37). Effect sizes were larger when subjects were asked to generate an entire target (.55) versus only part of the target (.32), and effect sizes were larger when smaller stimulus sets were used. The influence of interval length before testing indicates that

there is a general trend toward larger effect sizes with longer retention intervals: Immediate testing was at .41, <1 min at .32, 1 min to 1 day at .41, and >1 day at .64.

The effect sizes associated with the difficulty of the generation task were all very similar. The tasks designated by the coders as easy (e.g., letter switching) had an effect size of .41, moderate tasks (e.g., rhyming) had an effect of .37, and hard tasks (e.g., mental multiplication) had an effect of .46. There are several possible explanations for these counterintuitive results: The easy condition included more than 13 times as many subjects as did the hard condition (9,638 vs. 685); the level of difficulty may also overlap with the type of processing that is involved (many easy tasks require only shallow types of encoding); and difficulty may also be confounded with the type of generation task (many difficult studies used number calculation).

### Theoretical Implications

To examine several theories that attempt to explain the generation effect, we performed several additional analyses. One of the most intuitively appealing theories is that of mental effort. It is believed that read conditions require a lesser amount of cognitive work than do generate conditions, a situation that may result in less accurate test performance (e.g., McFarland et al., 1980). According to this theory, the larger the effort required to process stimuli, the larger the generation effect size. Our results seem to lend partial support to this theory: Subjects' generation of only part of the target stimuli—presumably an easier task—yielded a smaller generation effect size (8%,  $d = .32$ ) than did their generation of the whole target (12%,  $d = .55$ ).

There is a greater amount of evidence, however, that “effort” by itself is an insufficient explanation for the generation effect. First, our classification of generation rules into groups of easy, moderate, or hard levels of difficulty resulted in effect sizes that were very similar to each other and, for the easy and moderate groups, overlapping confidence intervals. Second, the presentation of the task via incidental (effortless) learning yielded an effect (14%,  $d = .65$ ) that was twice the size of the one for intentional (effortful) learning (7%,  $d = .32$ ). Third, although there were no primary studies with both older adults and tasks we coded as hard, older adults had a larger effect size than did younger subjects on tasks that were rated as moderate (10%,  $d = .46$  vs. 7%,  $d = .32$ ). The difference in  $d$  was smaller on tasks rated as easy (13%,  $d = .55$  vs. 10%,  $d = .46$ ), an unexpected pattern, given the substantial amount of literature on the negative relationship between task difficulty and age (see Zacks, Hasher, & Li, 2000, for a review). It is likely that “effort” as a construct may be a term that is too broad to be an effective explanation for the generation effect.

The selective rehearsal displacement hypothesis is a second theory that is used to account for the benefit that generating has over reading (Slamecka & Katsaiti, 1987). This theory suggests that when read and generate items are presented in a mixed (random) list under free-recall testing conditions, subjects sacrifice rehearsal of the read items

and favor the generate items. Evidence that seems to lend support to this theory was taken from the finding that generate performance was better under mixed-presentation conditions than it was under blocked presentation; the read conditions showed the opposite pattern. Our analyses across studies with a total of 9,091 subjects that took free-recall tests did not support this pattern: Subjects in both the read and generate conditions actually scored significantly higher when presentation of stimuli was blocked ( $M_R = .38, M_G = .42$ ) than they did when presentation of stimuli was in a mixed list ( $M_R = .23, M_G = .35$ ; both  $ps < .05$ ). In addition, according to this theory, there should be no generation effect in (free-recall) experiments with between-subjects designs, since selective rehearsal of generate items would not take time from rehearsal of read items. We found that there was a small average generation effect of 4% ( $d = .17$ ) in the experiments that had between-subjects designs.

Our third examination actually pertains to several theories, all of which emphasize the processes used in generation versus reading, especially as to how those processes overlap the ones used at test (Crutcher & Healy, 1989; deWinstanley & Bjork, 1997; Soraci et al., 1994). In all these theories, the contention is that the more the processes used at study overlap those used at test, the better the test performance becomes (transfer-appropriate processing; Morris, Bransford, & Franks, 1977). We examined this idea in several ways. First, the processing done by subjects asked to generate only part of the target words at study would appear to overlap more with that needed for a cued-recall test, whereas those asked to generate the complete target word would match more with free-recall test processing. The generation effect was 5% ( $d = .24$ ) when subjects generated part of the target and then took a free-recall test. Those who generated the whole target under free-recall test conditions had a generation effect of 13% ( $d = .57$ ). The reverse was not true, however; the performance of subjects who took a cued-recall test after generating whole targets was approximately the same as that of those subjects who had generated partial targets (12% vs. 11%).

To further examine the influence of processing match between study and test, we looked at the generation rules that subjects used when they produced targets. First, the 11 rules for which we coded were reorganized into two classes: (1) cue-based rules, in which subjects' generation of the correct target depended on the target's relation to a provided cue (e.g., rhyming rules, math calculation), and (2) target-based rules, in which subjects needed no cues to correctly generate the target word (i.e., word fragment completion, in which the target word could only be completed using specific letters, or letter rearrangement). Next, we compared the size of the generation effects in these two classes of rules on the basis of whether a free-recall or cued-recall test was given. If transfer-appropriate processing were a valid explanation for the generation effect, we would expect that cue-based rules would be more successfully tested using cued recall, whereas successful testing of target-based rules would best be achieved through free recall. Additionally, since letter rearrangement is the only generation rule in which the entire target

is present, we examined the generation effect for this rule in terms of recognition versus recall testing.

Under free-recall test conditions, experiments with cue-based rules yielded larger effects (8%,  $d = .35$ ) than did those with target-based rules (3%,  $d = .13$ ). Under cued-recall conditions, the opposite was true: Target-only rules yielded larger generation effects (16%,  $d = .73$ ) than did cue-based rules (12%,  $d = .55$ ). One caution we should note is that there are large differences in the number of studies within each of these cells. Our final comparison was of subjects' performance in recognition versus recall tests that used the letter rearrangement rule. The generation effect that was based on this rule was indeed smaller for recall tests (7%,  $d = .33$ ) than it was for recognition tests (10%,  $d = .47$ )—as the transfer-appropriate processing theory would predict. On the basis of this set of analyses, we concluded that the usefulness of transfer-appropriate processing as a general explanation for the generation effect remains unclear.<sup>2</sup>

The recognition that measurement issues are overlooked despite their obvious importance is not new (see Cone & Foster, 1991). However, these issues become more critical when they restrict the ability of meta-analytic techniques to correct for sources of error. To the extent that primary studies tend to omit reporting basic measurement-related information or critical statistical data (e.g.,  $n$  sizes), the full potential this type of meta-analysis has to correct for artifactual variance is limited. In future studies, it is critical that researchers (and editors) emphasize the reporting of all relevant data, including the reliabilities of dependent measures, even if nonsignificant differences are found. As meta-analyses become more common, this information will likely be needed to deepen our understanding of research results.

In conclusion, our findings represent estimates of the magnitude of and moderating influences on the generation effect. We recognize that some of our conclusions are based on moderator classes with relatively low  $n$  sizes, so we encourage others to replicate and extend these findings, as more studies become available. Additional studies (and the cumulation of those studies) can only help to clarify existing theories or to assist in the creation of new ones. Regardless of what the underlying cognitive mechanisms may be, the generation effect appears to be a real phenomenon that deserves further empirical study.

#### AUTHOR NOTE

The authors thank Richard A. Block and Alan S. Brown for their suggestions and advice in the preparation of the manuscript. Please send all correspondence to S. Bertsch, University of Pittsburgh at Johnstown, Johnstown, PA 15904 (e-mail: bertsch@pitt.edu).

*Note*—This article was accepted by the previous editorial team, when Colin M. MacLeod was Editor.

#### REFERENCES

- (\* indicates work included in meta-analysis)  
 \*BEGG, I., VINSKI, E., FRANKOVICH, L., & HOLGATE, B. (1991). Generating makes words memorable, but so does effective reading. *Memory & Cognition*, *19*, 487-497.  
 BROWN, A. S., & MITCHELL, D. B. (1991). Age differences in retrieval consistency and response dominance. *Journal of Gerontology*, *46*, P332-P339.

- BROWN, J. C., NIINIKOSKI, J., & DUKE, L. W. (1993). Generation effect and frequency judgment in young and elderly adults. *Experimental Aging Research*, **19**, 147-164.
- \*BURNS, D. J. (1992). The consequences of generation. *Journal of Memory & Language*, **31**, 615-633.
- \*BURNS, D. J. (1996). The item-order distinction and the generation effect: The importance of order information in long-term memory. *American Journal of Psychology*, **109**, 567-580.
- \*BURNS, D. J., CURTI, E. T., & LAVIN, J. C. (1993). The effects of generation on item and order retention in immediate and delayed recall. *Memory & Cognition*, **21**, 846-852.
- \*BUYER, L. S., & DOMINOWSKI, R. L. (1989). Retention of solutions: It is better to give than to receive. *American Journal of Psychology*, **102**, 353-363.
- \*CARROLL, M., & NELSON, T. O. (1993). Failure to obtain a generation effect during naturalistic learning. *Memory & Cognition*, **21**, 361-366.
- \*CHECHILE, R. A., & SORACI, S. A., JR. (1999). Evidence for a multiple-process account of the generation effect. *Memory*, **7**, 483-508.
- CHRISTENSEN-SZALANSKI, J. J., & WILLHAM, C. F. (1991). The hindsight bias: A meta-analysis. *Organizational Behavior & Human Decision Processes*, **48**, 147-168.
- \*CLARK, S. E. (1995). The generation effect and the modeling of associations in memory. *Memory & Cognition*, **23**, 442-455.
- CONE, J. D., & FOSTER, S. L. (1991). Training in measurement: Always the bridesmaid. *American Psychologist*, **46**, 653-654.
- \*CRUTCHER, R. J., & HEALY, A. F. (1989). Cognitive operations and the generation effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 669-675.
- \*DEWINSTANLEY, P. A., & BJORK, E. L. (1997). Processing instructions and the generation effect: A test of the multifactor transfer-appropriate processing theory. *Memory*, **5**, 401-421.
- \*DEWINSTANLEY, P. A., & BJORK, E. L. (2004). Processing strategies and the generation effect: Implications for making a better reader. *Memory & Cognition*, **32**, 945-955.
- \*DICK, M. B., KEAN, M.-L., & SANDS, D. (1989). Memory for internally generated words in Alzheimer-type dementia: Breakdown in encoding and semantic memory. *Brain & Cognition*, **9**, 88-108.
- \*DONALDSON, W., & BASS, M. (1980). Relational information and memory for problem solutions. *Journal of Verbal Learning & Verbal Behavior*, **19**, 26-35.
- \*FIEDLER, K., LACHNIT, H., FAY, D., & KRUG, C. (1992). Mobilization of cognitive resources and the generation effect. *Quarterly Journal of Experimental Psychology*, **45A**, 149-171.
- \*FLORY, P., & PRING, L. (1995). The effects of data-driven and conceptually driven generation of study items on direct and indirect measures of memory. *Quarterly Journal of Experimental Psychology*, **48A**, 153-165.
- \*GARDINER, J. M. (1988). Generation and priming effects in word-fragment completion. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 495-501.
- \*GARDINER, J. M. (1989). A generation effect in memory without awareness. *British Journal of Psychology*, **80**, 163-168.
- \*GARDINER, J. M., & ARTHURS, F. S. (1982). Encoding context and the generating effect in multitrail free-recall learning. *Canadian Journal of Psychology*, **36**, 527-531.
- \*GARDINER, J. M., DAWSON, A. J., & SUTTON, E. A. (1989). Specificity and generality of enhanced priming effects for self-generated study items. *American Journal of Psychology*, **102**, 295-305.
- \*GARDINER, J. M., GREGG, V. H., & HAMPTON, J. A. (1988). Word frequency and generation effects. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 687-693.
- \*GARDINER, J. M., & HAMPTON, J. A. (1985). Semantic memory and the generation effect: Some tests of the lexical activation hypothesis. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 732-741.
- \*GARDINER, J. M., & HAMPTON, J. A. (1988). Item-specific processing and the generation effect: Support for a distinctiveness account. *American Journal of Psychology*, **101**, 495-504.
- \*GARDINER, J. M., & ROWLEY, J. M. C. (1984). A generation effect with numbers rather than words. *Memory & Cognition*, **12**, 443-445.
- \*GHATALA, E. S. (1983). When does internal generation facilitate memory for sentences? *American Journal of Psychology*, **96**, 75-83.
- GLASS, G. V. (1977). Integrating findings: The meta-analysis of research. *Review of Research in Education*, **5**, 351-379.
- \*GLISKY, E. L., & RABINOWITZ, J. C. (1985). Enhancing the generation effect through repetition of operations. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 193-205.
- \*GRAF, P. (1981). Reading and generating normal and transformed sentences. *Canadian Journal of Psychology*, **35**, 293-308.
- \*GREENWALD, A. G., & JOHNSON, M. M. (1989). The generation effect extended: Memory enhancement for generation cues. *Memory & Cognition*, **17**, 673-681.
- \*GROSOFSKY, A., PAYNE, D. G., & CAMPBELL, K. D. (1994). Does the generation effect depend upon selective displaced rehearsal? *American Journal of Psychology*, **107**, 53-68.
- \*HARA, K., NEUMANN, E., & TAJIKA, H. (1989). Effects of word versus nonword rehearsal frequency on the generation effect. *Psychologia*, **32**, 230-235.
- \*HERTEL, P. T. (1989). The generation effect: A reflection of cognitive effort? *Bulletin of the Psychonomic Society*, **27**, 541-544.
- \*HIRSHMAN, E., & BJORK, R. A. (1988). The generation effect: Support for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 484-494.
- HUNTER, J. E., & SCHMIDT, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.
- \*JACOBY, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning & Verbal Behavior*, **17**, 649-667.
- \*JACOBY, L. L. (1983). Remembering the data: Analyzing interactive processes in reading. *Journal of Verbal Learning & Verbal Behavior*, **22**, 485-508.
- \*JAVA, R. I. (1994). States of awareness following word stem completion. *European Journal of Cognitive Psychology*, **6**, 77-92.
- \*JAVA, R. I. (1996). Effects of age on state of awareness following implicit and explicit word-association tasks. *Psychology & Aging*, **11**, 108-111.
- \*JOHNS, E. E., & SWANSON, L. G. (1988). The generation effect with nonwords. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 180-190.
- \*JOHNSON, M. K., RAYE, C. L., FOLEY, H. J., & FOLEY, M. A. (1981). Cognitive operations and decision bias in reality monitoring. *American Journal of Psychology*, **94**, 37-64.
- \*JOHNSON, M. M., SCHMITT, F. A., & PIETRUKOWICZ, M. (1989). The memory advantages of the generation effect: Age and process differences. *Journal of Gerontology*, **44**, P91-P94.
- \*KANE, J. H., & ANDERSON, R. C. (1978). Depth of processing and interference effects in the learning and remembering of sentences. *Journal of Educational Psychology*, **70**, 626-635.
- \*KINOSHITA, S. (1989). Generation enhances semantic processing? The role of distinctiveness in the generation effect. *Memory & Cognition*, **17**, 563-571.
- \*LIU, I.-M., & LEE, Y.-S. (1990). Memorial consequences of generating words and non-words. *Quarterly Journal of Experimental Psychology*, **42A**, 255-278.
- \*LUTZ, J., BRIGGS, A., & CAIN, K. (2003). An examination of the value of the generation effect for learning new material. *Journal of General Psychology*, **130**, 171-188.
- \*MACLEOD, C. M., & DANIELS, K. A. (2000). Direct versus indirect tests of memory: Directed forgetting meets the generation effect. *Psychonomic Bulletin & Review*, **7**, 354-359.
- \*MCCLELLAND, A. G., & PRING, L. (1991). An investigation of cross-modality effects in implicit and explicit memory. *Quarterly Journal of Experimental Psychology*, **43A**, 19-33.
- \*MCDANIEL, M. A., RIEGLER, G. L., & WADDILL, P. J. (1990). Generation effects in free recall: Further support for a three-factor theory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 789-798.
- \*MCDANIEL, M. A., WADDILL, P. J., & EINSTEIN, G. O. (1988). A contextual account of the generation effect: A three-factor theory. *Journal of Memory & Language*, **27**, 521-536.
- \*MCELROY, L. A. (1987). The generation effect with homographs: Evidence for postgeneration processing. *Memory & Cognition*, **15**, 148-153.
- \*MCELROY, L. A., & SLAMECKA, N. J. (1982). Memorial consequences of generating nonwords: Implications for semantic-memory interpre-

- tations of the generation effect. *Journal of Verbal Learning & Verbal Behavior*, **21**, 249-259.
- \*MCFARLAND, C. E., JR., FREY, T. J., & RHODES, D. D. (1980). Retrieval of internally versus externally generated words in episodic memory. *Journal of Verbal Learning & Verbal Behavior*, **19**, 210-225.
- \*MCFARLAND, C. E., JR., WARREN, L. R., & CROCKARD, J. (1985). Memory for self-generated stimuli in young and old adults. *Journal of Gerontology*, **40**, 205-207.
- \*MCNAMARA, D. S., & HEALY, A. F. (1995). A procedural explanation of the generation effect: The use of an operand retrieval strategy for multiplication and addition problems. *Journal of Memory & Language*, **34**, 399-416.
- \*MCNAMARA, D. S., & HEALY, A. F. (2000). A procedural explanation of the generation effect for simple and difficult multiplication problems and answers. *Journal of Memory & Language*, **43**, 652-679.
- MITCHELL, D. B., HUNT, R. R., & SCHMITT, F. A. (1986). The generation effect and reality monitoring: Evidence from dementia and normal aging. *Journal of Gerontology*, **41**, 79-84.
- MORRIS, C. D., BRANSFORD, J. D., & FRANKS, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning & Verbal Behavior*, **16**, 519-533.
- \*MULLIGAN, N. W. (2001). Generation and hypermnesia. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 436-450.
- \*MULLIGAN, N. W. (2002a). The emergent generation effect and hypermnesia: Influences of semantic and nonsemantic generation tasks. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 541-554.
- \*MULLIGAN, N. W. (2002b). The generation effect: Dissociating enhanced item memory and disrupted order memory. *Memory & Cognition*, **30**, 850-861.
- \*MULLIGAN, N. W., & DUKE, M. D. (2002). Positive and negative generation effects, hypermnesia, and total recall time. *Memory & Cognition*, **30**, 1044-1053.
- \*NAIRNE, J. S., PUSEN, C., & WIDNER, R. L., JR. (1985). Representation in the mental lexicon: Implications for theories of the generation effect. *Memory & Cognition*, **13**, 183-191.
- \*NAIRNE, J. S., RIEGLER, G. L., & SERRA, M. (1991). Dissociative effects of generation on item and order retention. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 702-709.
- \*NAIRNE, J. S., & WIDNER, R. L., JR. (1987). Generation effects with nonwords: The role of test appropriateness. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13**, 164-171.
- \*NAIRNE, J. S., & WIDNER, R. L., JR. (1988). Familiarity and lexicality as determinants of the generation effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 694-699.
- \*NICOLAS, S., & TARDIEU, H. (1996). The generation effect in a word-stem completion task: The influence of conceptual processes. *European Journal of Cognitive Psychology*, **8**, 405-424.
- \*OLOFSSON, U., & NILSSON, L.-G. (1992). The generation effect in primed word-fragment completion reexamined. *Psychological Research*, **54**, 103-109.
- O'NEILL, W., ROY, L., & TREMBLAY, R. (1993). A translation-based generation effect in bilingual recall and recognition. *Memory & Cognition*, **21**, 488-495.
- \*PAYNE, D. G., NEELY, J. H., & BURNS, D. J. (1986). The generation effect: Further tests of the lexical activation hypothesis. *Memory & Cognition*, **14**, 246-252.
- \*PESTA, B. J., SANDERS, R. E., & MURPHY, M. D. (1999). A beautiful day in the neighborhood: What factors determine the generation effect for simple multiplication problems? *Memory & Cognition*, **27**, 106-115.
- \*PESTA, B. J., SANDERS, R. E., & NEMEC, R. J. (1996). Older adults' strategic superiority with mental multiplication: A generation effect assessment. *Experimental Aging Research*, **22**, 155-169.
- PEYNIRCIOĞLU, Z. F. (1989). The generation effect with pictures and nonsense figures. *Acta Psychologica*, **70**, 153-160.
- \*PEYNIRCIOĞLU, Z. F., & MUNGAN, E. (1993). Familiarity, relative distinctiveness, and the generation effect. *Memory & Cognition*, **21**, 367-374.
- PRING, L. (1988). The "reverse-generation" effect: A comparison of memory performance between blind and sighted children. *British Journal of Psychology*, **79**, 387-400.
- \*RABINOWITZ, J. C. (1989). Judgments of origin and generation effects: Comparisons between young and elderly adults. *Psychology & Aging*, **4**, 259-268.
- \*RABINOWITZ, J. C., & CRAIK, F. I. M. (1986). Specific enhancement effects associated with word generation. *Journal of Memory & Language*, **25**, 226-237.
- \*REARDON, R., DURSO, F. T., FOLEY, M. A., & MCGAHAN, J. R. (1987). Expertise and the generation effect. *Social Cognition*, **5**, 336-348.
- \*SCHMIDT, S. R. (1990). A test of resource-allocation explanations of the generation effect. *Bulletin of the Psychonomic Society*, **28**, 93-96.
- \*SCHMIDT, S. R. (1992). Evaluating the role of distinctiveness in the generation effect. *Quarterly Journal of Experimental Psychology*, **44A**, 237-260.
- \*SCHMIDT, S. R., & CHERRY, K. (1989). The negative generation effect: Delineation of a phenomenon. *Memory & Cognition*, **17**, 359-369.
- \*SCHWEICKERT, R., MCDANIEL, M. A., & RIEGLER, G. (1994). Effects of generation on immediate memory span and delayed unexpected free recall. *Quarterly Journal of Experimental Psychology*, **47A**, 781-804.
- \*SERRA, M., & NAIRNE, J. S. (1993). Design controversies and the generation effect: Support for an item-order hypothesis. *Memory & Cognition*, **21**, 34-40.
- \*SLAMECKA, N. J., & FEVREISKI, J. (1983). The generation effect when generation fails. *Journal of Verbal Learning & Verbal Behavior*, **22**, 153-163.
- \*SLAMECKA, N. J., & GRAF, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory*, **4**, 592-604.
- \*SLAMECKA, N. J., & KATSAITI, L. T. (1987). The generation effect as an artifact of selective displaced rehearsal. *Journal of Memory & Language*, **26**, 589-607.
- \*SMITH, R. W., & HEALY, A. F. (1998). The time-course of the generation effect. *Memory & Cognition*, **26**, 135-142.
- SMITH, S. M., & VELA, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review*, **8**, 203-220.
- \*SOLOWAY, R. M. (1986). No generation effect without semantic activation. *Bulletin of the Psychonomic Society*, **24**, 261-262.
- \*SORACI, S. A., JR., CARLIN, M. T., CHECHILE, R. A., FRANKS, J. J., WILLS, T., & WATANABE, T. (1999). Encoding variability and cuing in generative processing. *Journal of Memory & Language*, **41**, 541-559.
- \*SORACI, S. A., JR., FRANKS, J. J., BRANSFORD, J. D., CHECHILE, R. A., BELLI, R. F., CARR, M., & CARLIN, M. (1994). Incongruous item generation effects: A multiple-cue perspective. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 67-78.
- STEFFENS, M. C., & ERDFELDER, E. (1998). Determinants of positive and negative generation effects in free recall. *Quarterly Journal of Experimental Psychology*, **51A**, 705-733.
- \*TACONNAT, L., & ISINGRINI, M. (2004). Cognitive operations in the generation effect on a recall test: Role of aging and divided attention. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **30**, 827-837.
- \*THOMPSON, C. P., & BARNETT, C. (1981). Memory for product names: The generation effect. *Bulletin of the Psychonomic Society*, **18**, 241-243.
- \*TOTH, J. P., & HUNT, R. R. (1990). Effect of generation on a word-identification task. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 993-1003.
- TYLER, S. W., HERTEL, P. T., MCCALLUM, M. C., & ELLIS, H. C. (1979). Cognitive effort and memory. *Journal of Experimental Psychology: Human Learning & Memory*, **5**, 607-617.
- VISWESVARAN, C., & SCHMIDT, F. L. (1992). A meta-analytic comparison of the effectiveness of smoking cessation methods. *Journal of Applied Psychology*, **77**, 554-561.
- \*WATKINS, M. J., & SECHLER, E. S. (1988). Generation effect with an incidental memorization procedure. *Journal of Memory & Language*, **27**, 537-544.
- \*WIDNER, R. L., JR. (1995). Associative spread as a mediating variable in the generation effect. *Memory*, **3**, 1-19.
- ZACKS, R. T., HASHER, L., & LI, K. Z. H. (2000). Human memory. In F. I. M. Craik & T. A. Salthouse (Eds.), *Handbook of aging and cognition* (pp. 293-357). Mahwah, NJ: Erlbaum.

**NOTES**

1. Out of all the group comparisons, these two types of analyses indicated different patterns of effect sizes for only two classes of moderators: test type and generation rule. Using the difference score as the basis for calculation led to a steady decrease in effect size from recognition through cued- to free-recall testing conditions. When we used the read and generate proportions individually as the basis of calculation, we found that the strongest effect size was for the cued-recall condition, followed, in order, by recognition and free recall. Within the generation rule analysis, the ranking of effect sizes by the rule used in generation changed for 2 of the 10 rules examined: Both synonym and category rules had a higher ranking when difference scores were used as the basis of the effect size calculation.

2. There is another class of generation effect theories—a class that is based on the contributions of the different sources of information that are used in the generate process (e.g., Hirshman & Bjork, 1988; McDaniel, Waddill, & Einstein, 1988; Steffens & Erdfelder, 1998). Some of the data we have presented (i.e., stronger effects for cue–target relational processing over target alone) may also be useful to these theorists, but since our data were not coded for one component of these theories (whole-list or intertarget processing), we were unable to discuss them completely.

(Manuscript received November 18, 2002;  
revision accepted for publication October 23, 2005.)