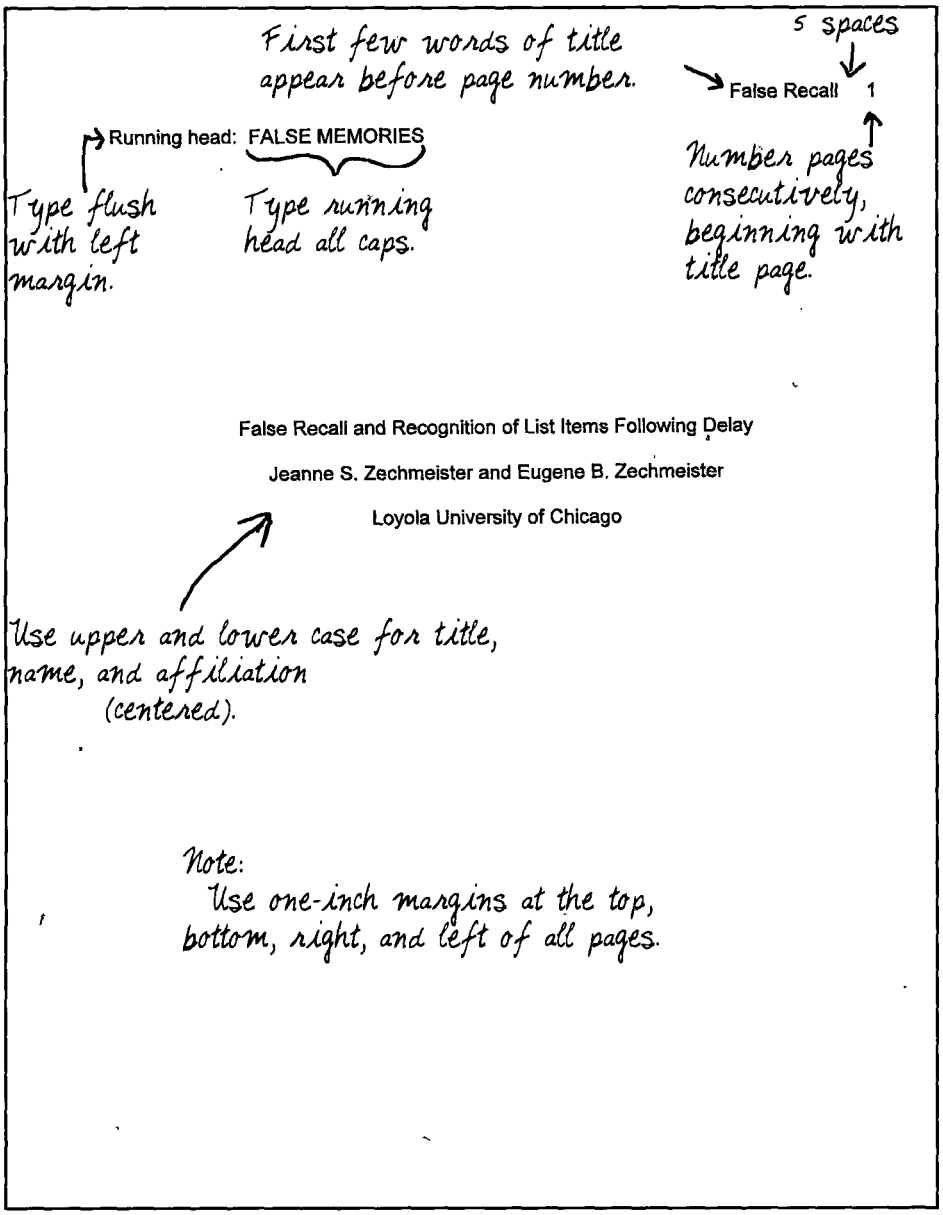
included in the appendix: actual lists with randomizations made, type of apparatus used for presentation, instructions to participants for all conditions, and randomizations of conditions.

Finally, a research proposal should include material to be submitted to an Institutional Review Board (IRB) or similar committee designed to review the ethics of the proposed research (see Chapter 3).

A Sample Research Report



Center, do not italicize.

Abstract

No paragraph indentation for abstract.

Whether false memories persist over a delay was investigated using H. L. Roediger and K. B. McDermott's (1995) false memory procedure. College students (N = 19) studied 12 15-item lists and either recalled items immediately or following a 90-s delay. A first recognition test immediately followed recall of lists; a second test occurred 2 days later. The proportion of correctly recalled list items decreased with the 90-s delay, whereas the proportion of falsely recalled critical items increased. The overall proportion of items recognized as "old" was greater for false critical items than for correct list items, but both false recognition and correct recognition decreased over the 2-day period. These findings are compared to other studies that examined false memories following delays.

Abstract should be a single paragraph not exceeding 120 words.

Note:
Double-space between all lines of the manuscript. There should be no more than 27 lines of text on each page.
When using a word processor, use left justification only (not full). Use italics for statistical symbols and infrequently otherwise (underline if using a typewriter), and use one font size throughout the manuscript.

Leave one space at end of all sentences, and after all commas, colons, and semicolons.

Title appears centered on page 3, with first letter of major words in caps.

False Recall and Recognition of List Items Following Delay

Roediger and McDermott (1995) captured the attention of memory researchers by reminding us of the importance of examining participants' errors in recall and that these errors can sometimes be quite predictable. Moreover, these "false memories" are relevant to ongoing debates about false memories that may occur in more natural settings (e.g., clinics, courtrooms). Roediger and McDermott, and many others, have demonstrated that false memories are easily created and that participants' phenomenological experience of these memories mimics their memories for actual events (e.g., Mather, Henkel, & Johnson, 1997; Norman & Schacter, 1997; Payne, Elie, Blackwell, & Neuschatz, 1996). The purpose of the present study was to replicate Roediger and McDermott's (1995) findings and to extend their findings by examining the persistence of false memories following delays. If false memories persist following delays, important implications exist regarding courtroom testimony and recovered memories in therapy. For example, memory retrieval in these cases follows varying periods of delay and the accuracy of these memories may thus be questioned.

Identify all authors (surnames only) the first time you cite a source (five or fewer authors).

List multiple sources in alphabetical order, separated by a semicolon.

Indent every paragraph using the tab key (set at five to seven spaces or 1/2 inch).

In this false memory paradigm word lists are constructed such that each word is highly associated with a critical, not-presented word. For example, participants may hear the words *bed*, *rest*, *awake*, *tired*, *dream*, etc., but never the critical word, *sleep*. False recall and recognition for the word *sleep* is quite high. Underwood (1965) proposed that an implicit associative response (IAR) occurs when participants study an associated word. For example, when participants hear "bed" they may think of "sleep" (the IAR). With the next word, "rest," they again may experience "sleep" as an IAR, and so on. When their memory is tested participants may have difficulty discriminating the

Abbreviated terms are first spelled out, followed by the abbreviation in parentheses.

Use Latin abbreviations (such as e.g. and i.e.) only in parentheses.

Use numbers to express amounts 10 and above, and when numbers are grouped for comparison with numbers 10 and above.

An "s" (with no period) is the abbreviation for second.

Use number for units of time or measure.

Note: Use past tense to describe the study.

Flush with left margin, italicized, only first letter in caps.

Always use words to express numbers that begin a sentence.

IAR from the presented items, resulting in false recall or recognition. An alternative account relies on the extent to which the lists of associated words activate a schema for the words (e.g., a "sleep" schema). According to this account, recall after a delay may be based on memory for the schema rather than the actual list presentation.

In the present study participants listened to 12 lists comprised of 15 words that were associated to a critical, not-presented word. Participants were randomly assigned to recall 4 of the 12 lists in either an immediate-recall condition or following a 90-s delay in a single trial of free recall. When not recalling words (and during the 90-s delays) participants completed distractor tasks consisting of math problems. Additionally, two recognition tests were administered. The first recognition test followed the presentation and recall of the 15 lists; the second was 2 days later.

We hypothesized that the proportions of correct recall and false recall would not differ in the immediate-recall condition because of participants' inability to differentiate their memory for the presented words from the IARs generated for the critical items. In the delayed-recall condition, we expected the proportion of false items recalled to be greater than the proportion of correct items recalled because of participants' activation of a schema for the list to aid recall. This activation was assumed to lead to recall of the highly associated, critical item. Furthermore, we anticipated that correct recall would decrease over time but false recall would increase with the delay. Similar predictions were made for participants' recognition of correct list items and false critical items.

Method

Center, no italics, only the first letter in caps.

Participants

Nineteen students (16 women, 3 men) from a research methods in psychology

Note: Method, Results, and Discussion sections do not begin on new page (unless coincidentally).

Note: "Data" is plural.

Indented, and italicized paragraph heading, followed by a period. Only first letter of first word in caps.

Use words to express numbers less than 10.

The word "participants" is preferred over "subjects."

class at Loyola University of Chicago participated as part of a class project. Data for one student were omitted from all analyses because the student attempted recall for the wrong lists; recognition data for two students were omitted from analyses because they missed the second class session in which the experiment was conducted.

Materials

Word lists. We used 12 15-item word lists from Roediger and McDermott's (1995) study. They constructed 24 lists by selecting 15 words that were highly associated with a critical word that was not presented. Words were ordered such that the strongest associates were presented first, followed by more weakly associated words. For example, the list for the critical word, *needle*, was *thread*, *pin*, *eye*, *sewing*, *sharp*, *point*, *prick*, *thimble*, *haystack*, *thorn*, *hurt*, *injection*, *syringe*, *cloth*, and *knitting*. Six lists in the present study were identified by Roediger and McDermott as having a high likelihood of creating false memories (identified by the critical item): chair, mountain, needle, rough, sleep, sweet. The other six lists used in the present study for recall were randomly selected from their remaining lists (identified by the critical item): bread, cold, doctor, foot, music, slow.

Participants recalled words from 4 of the 12 lists in a single trial of free recall. In order to rule out the possibility that proportions of correct and false recall would be determined by the particular lists used for the recall task, the four recall lists were counterbalanced. The 12 lists were randomly divided into three sets of four lists, with the provision that two lists in each set have high likelihoods for creating false memories. Participants were randomly assigned to recall words for four lists. Seven participants recalled the chair, bread, slow, and sleep lists; seven participants recalled the needle,

foot, doctor, and rough lists, and four participants recalled the sweet, music, cold, and mountain lists.

Distractor task. Participants completed a distractor task consisting of math problems when not recalling words and during the 90-s delay. We generated 20 pages of math problems. Each page had six complex math problems involving addition, subtraction, multiplication, and division.

Recall packets. We created six different packets to manipulate the recall condition (immediate, delayed) and counterbalance the lists used for recall. Four pages presented recall instructions based on those used by Roediger and McDermott (1995): "Please write down as many words from the list that you remember. A good strategy is to write down the last few items first, then the rest of the words in any order." Twenty pages had math problems. Between each page was a colored sheet of paper, which masked the participants' next task (recall or math problems). The order of the 24 pages depended on whether participants were in the immediate- or delayed-recall condition and which list set participants were assigned to recall. For example, the "music" list was presented first. Following the presentation of the list, participants assigned to recall the music list in the immediate condition opened their packet to a page with recall instructions; all other participants opened their packet to a page of math problems. After 90 s, a tone signaled participants to turn to the next white page (following the colored sheet). If assigned to recall the music list in the delayed-recall condition, participants' next sheet had recall instructions; all other participants received math problems. This sequence was followed for the remaining 11 word lists.

Recognition tests. Participants completed two different recognition tests, the first

Place periods or commas within quotation marks.

followed the presentation and recall of lists (immediate), the second was 2 days later (delayed). Because three list sets were used for recall (to counterbalance recall lists across participants), we created three versions of the first recognition test (session 1) and three versions of the second recognition test (session 2). Following Roediger and McDermott's terminology, recognition-test items that were presented during the recall portion of the study are referred to as "studied," and new distractor items are referred to as "not studied." Each 72-item recognition test consisted of six different types of items: 10 Studied/Recalled items (5 each from two of the four lists participants were asked to recall); 2 corresponding critical items for these lists; 20 Studied/Not Recalled items (5 each from four of the eight lists participants did not recall); 4 corresponding critical items for these lists; 30 Not Studied Distractor items (5 each from six lists that were not presented); and 6 corresponding critical items for the distractor lists (these lists also were from Roediger & McDermott, 1995). The five items selected from the word lists for each recognition test were the first, third, fifth, seventh, and ninth words. The 72 words on each list were randomly ordered. Participants made recognition judgments using a 4-point scale (1 = *sure new*, 2 = *probably new*, 3 = *probably old*, and 4 = *sure old*).

Procedure

A 2 (condition) X 2 (memory type) mixed factorial design was used for the recall portion of the experiment. Students were randomly assigned to either the immediate-recall condition or the delayed-recall condition; thus, the condition variable was manipulated as a random groups design. Memory type was a repeated measures variable with two levels: correct and false memory. The recognition portion of the experiment was a repeated measures independent variable. Participants completed the

A comma separates authors' surnames and year of publication when citation is within parentheses.

Use ampersand to indicate "and" for references within parentheses.

Italicize the anchors of a scale.

Numbers are used to indicate the levels of a design. Include one space to left and right of multiplication sign.

"min" (without a period) is the abbreviation for minute.

first recognition test immediately following the presentation and recall of word lists and completed the delayed recognition test 2 days later. The dependent variables were proportion of correct and false recall and proportion of correct and false recognition.

The experiment was conducted during two scheduled class periods. In the first session, lasting 50 min, the experimenter gave students a consent form and a packet that corresponded to their condition when they entered the class (students were randomly assigned to conditions prior to their arrival). A participant number appeared on top of the packet. Instructions requested students not to open their packet. After students read and signed their consent form, they listened to tape-recorded instructions for the study. Following this, the experimenter stopped the tape recorder and asked if there were any questions. Tape-recorded word lists were then presented. The random order of the lists was as follows (identified by the critical item): music, needle, sleep, doctor, slow, mountain, rough, bread, sweet, cold, chair, and foot. Words were presented at the rate of approximately one word every 1½ s. Following the list presentation, participants heard a tone informing them to turn to the next white page. They then either recalled list items or completed math problems for 90 s. A tone then signaled participants to turn to the next white page. This page instructed them to recall the list items or complete math problems. This procedure was followed for all 12 lists.

After the study and recall phase of the experiment, the experimenter distributed the first recognition test and asked participants to record their participant number on the top of the recognition test. After completing the recognition test, participants were asked to write their participant number in their notebook; they were then dismissed. At the beginning of the next class session (2 days later), participants reported their number

Note: Words are not hyphenated at the end of a line.

Note: Always double-space. Do not insert extra spaces between sections.

Center, no italics.

and they were given the corresponding second recognition test. Each recognition test took approximately 5 min to complete. The instructor then fully debriefed participants regarding the purpose of the experiment.

Results

Results are presented first for participants' recall and then for their recognition of list items. The dependent variables were proportion of correct recall/recognition and proportion of false recall/recognition. Proportion of correct recall was determined by counting the number of words participants correctly recalled and dividing by 60 (the total number of list items they were asked to recall). Proportion of false recall was determined by counting the number of critical items falsely recalled and dividing by 4 (the total number of critical items for the four lists participants recalled).

Similar proportions were computed for recognition data. For ease of comparisons, participants' recognition responses were collapsed such that ratings of 1 or 2 indicated judgments that the word was "new" (i.e., not presented during the recall portion of the study), and ratings of 3 or 4 indicated the word was "old" (i.e., presented during the recall portion). The number of items judged "old" for each type of word (e.g., studied, recalled; not studied distractor) was counted and divided by the total number of words of that type presented on the recognition test.

Recall

Mean proportions for correct and false recall were entered into a 2 (condition) X 2 (memory type) mixed-design analysis of variance (ANOVA). Across the two recall conditions the proportion of falsely recalled critical items (M = .60, SD = .23) tended to be greater than the proportion of correctly recalled list items (M = .48, SD = .12), F(1,

M is the abbreviation for arithmetic mean.

SD is the abbreviation for standard deviation.

Is the abbreviation for Mean Square Error.

Italicize all statistical terms.

Do not capitalize names of conditions.

Footnotes appear on a separate page near end of manuscript (not at bottom of the page).

Do not abbreviate "Figure." Figure caption and figure appear on separate pages at end of manuscript.

Degrees of freedom for statistical tests are reported in parentheses (with no space after statistical term).

16) = 4.38, p = .053 (*MSE* = .03, effect size r = .47).[1] The observed power to detect this large effect with the small sample size (n = 9 per condition), at alpha = .05, was estimated to be only .50. The main effect of condition was not significant, $F(1, 16) = .10$, $p > .05$ (*MSE* = .02, r = .10, power = .06). As predicted, however, recall condition and memory type interacted to influence participants' recall, $F(1, 16) = 9.50$, $p < .01$ (*MSE* = .03, r = .61). Simple main effects analyses revealed that recall differed in the immediate- and delayed-recall conditions in the predicted direction (see Figure 1). Participants in the immediate-recall condition did not differ in their proportions of correct recall and false recall, $F(1, 8) = .41$, $p > .05$ (*MSE* = .036, r = .22, power = .09); however, participants in the delayed-recall condition were more likely to recall a greater proportion of false critical items than correct list items, $F(1, 8) = 16.57$, $p < .005$ (*MSE* = .024, r = .82).[2]

A second set of simple main effects analyses examined the effect of delay for each type of recall. As expected, the proportion of correctly recalled list items decreased following delay and the proportion of false recall increased following delay (see Figure 1). For correct list items, participants in the delayed-recall condition recalled fewer items than participants who recalled list items immediately, $F(1, 16) = 15.42$, $p < .005$ (*MSE* = .008). The effect size (r) was .70, indicating a very large effect of immediate- vs. delayed-recall on participants' correct recall (see Cohen, 1988). In contrast, the proportion of falsely recalled critical items increased following the 90-s delay relative to the immediate-recall condition, $F(1, 16) = 3.77$, p = .07. The effect size (r) was .44; however, the estimated power to detect this large effect was only .45 (alpha = .05).

Table appears on separate page near end of manuscript.

Note use of spaces and punctuation when typing statistics.

Variable names in caps when they appear with a multiplication sign.

Capitalize nouns that are followed by a number.

Recognition

Across two testing sessions separated by 2 days, we examined the proportion of items recognized as "old" as a function of whether items were correct list items or false critical items and whether the items were recalled (vs. not recalled) during the recall portion of the experiment (see Table 1).[3] Distractor items were excluded from the analyses because the proportion of these items judged "old" was generally low across item type and sessions, $M = .18$ ($SD = .12$). Main effects were observed for the time, memory type, and recalled/not recalled variables. As expected, memory for the items decreased over the 2-day period, as the proportion of all items recognized as "old" decreased at Time 2 ($M = .71$, $SD = .14$) relative to Time 1 ($M = .87$, $SD = .11$), $F(1, 14) = 16.40$, $p < .005$ ($MSE = .04$, $r = .73$). Additionally, the main effect of memory type indicated that participants rated a higher proportion of false critical items "old" ($M = .85$, $SD = .11$) than correct list items ($M = .73$, $SD = .13$), $F(1, 14) = 14.72$, $p < .005$ ($MSE = .026$, $r = .71$). Thus, in terms of proportions, false recognition was greater than correct recognition at Time 1 and Time 2. Finally, the proportion of items judged "old" was greater for lists participants were asked to recall ($M = .87$, $SD = .09$) relative to lists they did not recall ($M = .71$, $SD = .14$), $F(1, 14) = 23.57$, $p < .001$ ($MSE = .032$, $r = .79$). This is the "testing effect" identified by Roediger and McDermott (1995), in which rehearsal during the recall portion of the experiment increases subsequent recognition.

There was also a Time X Recalled/Not Recalled interaction, $F(1, 14) = 7.49$, $p < .05$ ($MSE = .024$, $r = .59$). This interaction indicated that recognition of words for recalled lists was slightly higher at Time 1 ($M = .90$, $SD = .07$) relative to Time 2 ($M = .83$, $SD = .16$), $F(1, 15) = 3.22$, $p < .10$ ($MSE = .013$, $r = .42$, power = .39). In contrast,

recognition of words from not-recalled lists dropped significantly at Time 2 (M = .60, SD = .18) relative to Time 1 (M = .83, SD = .18), $F(1, 15) = 18.63$, $p < .005$ (MSE = .023, r = .74). Thus, rehearsal of the lists (through recall) increased the proportion of correct items and false critical items recognized as "old" even after a 2-day period, relative to lists that were not rehearsed.

As noted above, we observed differential effects of delay on the proportion recall of correct list items and false critical items. However, a similar effect was not observed for recognition data, as the Time X Memory Type interaction was not statistically significant, $F(1, 14) = .028$, $p > .05$ (MSE = .014). The effect size, r, for this interaction was quite small (.04, power = .05). Thus, the 2-day delay between the recognition tests resulted in similar decreases in the proportions of correct list items and false critical items that were recognized as "old."

Note: Make sure you use the "greater than" or "less than" sign correctly.

Center, no italics.

Discussion

Participants recalled list items in a single trial of free recall immediately after list presentation or following a 90-s filled delay. The proportion of correctly recalled list items decreased as a function of delay, but the proportion of false critical items recalled increased following the 90-s delay. Participants' recognition of correct list items and false critical items was tested following a 2-day delay. In terms of proportions, false recognition of critical items was greater than correct recognition of list items at Time 1 and Time 2, but recognition for both types of items decreased over the 2-day period. A testing effect occurred such that correct and false recognition were greater for lists participants had attempted recall relative to lists not recalled. The effects of rehearsal persisted over the 2-day period. In general, delay produced large effect sizes for both

recall and recognition.

McDermott (1996) used a 30-s filled delay to examine the effects of delay on participants' recall of correct list items and false critical items. Although she observed a decrease in correct recall following delay, her participants' false recall of critical items was not affected by delay. In contrast, we observed an increase in the proportion of false critical items recalled following a 90-s delay. One reason for this discrepancy may be derived from our differing findings for correct list recall. In McDermott's study the proportion of correct list items recalled following a 30-s delay was .50; in our study, the proportion of correct recall following a 90-s delay was .39. To the extent that a longer delay increased participants' forgetting of correct list items in our study, they may have had to rely more on schema-based recall, which we hypothesized to increase the occurrence of false memories. Alternatively, discrimination between IARs and memory for list items may become more difficult with longer delays. As a result, greater proportions of false recall would be expected (as was observed). This explanation is corroborated by McDermott's findings that after a 2-day delay, the proportion of falsely recalled critical items was greater than the proportion of correctly recalled items (similar to our findings for recognition). Thus, in both studies, participants demonstrated high levels of false recall that persisted over time.

Payne et al. (1996) examined participants' recognition memory for correct list items and false critical items in immediate (2 min) and delayed (24 hr) conditions. Similar to our recognition data, Payne et al. observed that false critical items were recognized with higher likelihood at both testing sessions than correct list items. However, they also observed an interaction between item type (correct, false) and

Year of publication in parentheses after author's surname.

Do not use contractions such as "wasn't."

Year of publication is not repeated within a paragraph if it is clear that the same study is being described.

Year of publication appears in parentheses even when reference has appeared previously in manuscript.

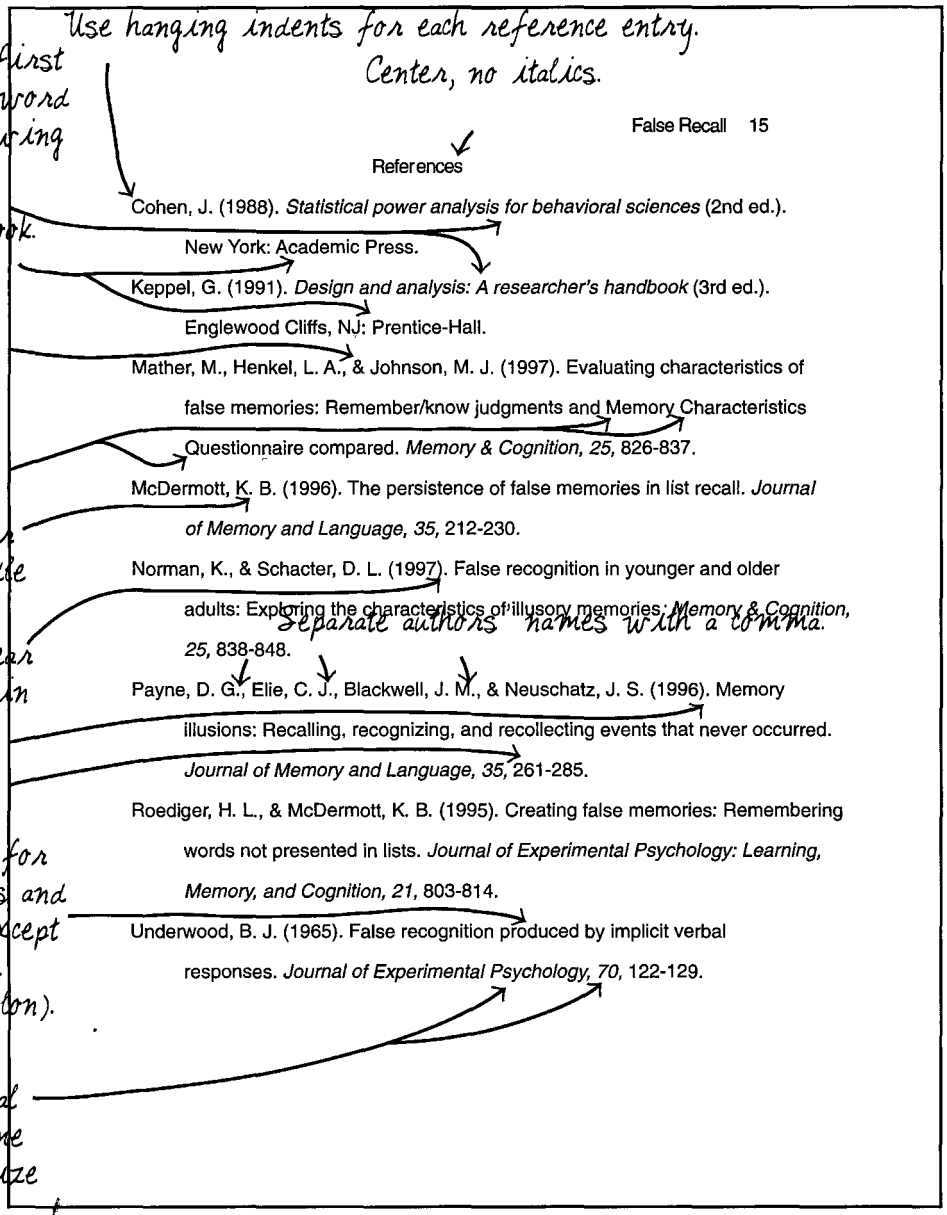
Use "et al." following surname of first author for work with three or more authors.
Note: First citation to this work lists all authors (see p. 3).

"hr" (with no period) is the abbreviation for hour.

retention (immediate, delayed) for their recognition data that we did not. They observed more forgetting for correct list items than false critical items over the 24-hr period, whereas over our 2-day period we observed similar decreases for correct list items and false critical items. Differences in the retention interval (24 vs. 48 hr), the number of word lists (16 vs. 12), or other procedural differences may account for this discrepancy.

The findings of this study and those of other recent studies indicate that false memories are frequent and persist following delays. Moreover, there is accumulating evidence to suggest that false memories for highly associated information may exceed correct memory over time. These data are limited to individuals' recall for highly associated list items. Important areas for future research include false memories for natural categories such as people and events. Also, it is not clear to what extent individuals' emotions may affect correct and false recall. This question is particularly relevant to the issue of false memory in the emotionally charged situations of eyewitness testimony and therapy.

Note: References section begins on a new page immediately after the Discussion section.



Note: List references in alphabetical order using first author's surname. Double-space references, no extra space between references.

Center, no italics

Author Note

Jeanne S. Zechmeister, Department of Psychology; Eugene B. Zechmeister, Department of Psychology.

Authors' department affiliations in first paragraph.

Portions of this research were presented at the Annual Meeting of the Psychonomic Society, November 1996. The authors thank the students in the first author's Research Methods in Psychology course at Loyola University of Chicago who participated in and discussed this study, and John J. Shaughnessy for his suggestions for this manuscript.

Acknowledge people who assisted with the research and manuscript preparation.

Send correspondence concerning this manuscript to Jeanne S. Zechmeister, Department of Psychology, Loyola University of Chicago, 6525 N. Sheridan Road, Chicago, IL 60626; e-mail: jzechme@luc.edu.

Last paragraph in Author Note specifies author's address for correspondence. Use this form exactly with your name and address.

Footnotes

[1] The effect size, *r*, was derived from ANOVA output that presented η^2 as a measure of effect size. η^2 represents the proportion of variance in the dependent variable that is accounted for by the independent variable. η^2 is analogous to r^2 in analyses involving one *df* for the effect. The square root of η^2 represents *r*, the magnitude of linear relationship between two variables. Cohen (1988) offers the following guidelines for interpreting the size of an effect using *r*: small effect, $r = .10$; medium effect, $r = .25$; large effect, $r = .40$.

[2] Simple main effect analyses were calculated by treating the comparison as a one-way ANOVA, as recommended by Keppel (1991) for mixed designs.

[3] We also examined whether immediate- vs. delayed-recall conditions influenced recognition in the ANOVA. Because this variable did not produce any statistically significant effects in the recognition analyses, it will not be discussed further.

Note: Use footnotes sparingly in manuscript.

Type flush with left margin.

Table title, italicized, with first letters of major words in caps.

Line separates title from body of table.

Center column headings over appropriate columns.

Line separates headings from table data.

Line at end of table data.

Table "notes" are typed flush with left margin.

Table 1

Mean Proportions for Recognizing Items as "Old" at Time 1 and Time 2 (48-hr Delay)

	Studied and recalled	Studied and not recalled	Distractor
Correct list items			
Time 1	.84	.79	.19
	(.12)	(.17)	(.12)
Time 2	.76	.56	.18
	(.22)	(.16)	(.16)
False critical items			
Time 1	.97	.88	.28
	(.12)	(.20)	(.13)
Time 2	.91	.64	.17
	(.20)	(.24)	(.10)

Note. Standard deviations are in parentheses. "Recalled" and "Not Recalled" refer to whether participants were asked to recall list items at Time 1, not whether these items were produced by participants during recall.

Note: Use double-space in table. Place only one table on a page.

Center, no italics.

↓
Figure Caption

Flush with left margin. Do not abbreviate "Figure."

Figure 1. Mean proportions of correct and false recall in the immediate- and delayed-recall conditions.

Note: More than one figure caption may appear on a "Figure Captions" page.
The figure itself appears on a separate page at end of manuscript. If there is more than one figure, each appears on a separate page.

Note: Short title and page number appear on the back of the figure page. Include figure number.

