

The Snodgrass picture fragment completion test: Alternate-form reliability

BEVERLY S. WYATT, FRANCES A. CONNERS, and MICHAEL D. CARR
University of Alabama, Tuscaloosa, Alabama

In two studies, the alternate-form reliability of the Snodgrass picture fragment completion test of implicit memory (Snodgrass, Smith, Feenan, & Corwin, 1987) was examined. In this test, identification thresholds are established for fragmented pictures. The same fragmented pictures are then shown again, intermixed with new fragmented pictures. Implicit memory is indicated by a decrease in identification threshold from the first to second presentation. Alternate-form reliability was low to moderate, depending on the measure used, regardless of the length of the test. A third study showed that explicit memory instructions did not increase the reliability. Recommendations for use of the test in correlational and experimental research are presented.

The picture fragment completion test used in this study was first designed by Gollin (1960) and later modified by Snodgrass, Smith, Feenan, and Corwin (1987). Snodgrass et al. (1987) developed a computerized version of the test. The Snodgrass test or a close version of it has been used in several studies to measure implicit memory (e.g., Bondi & Kaszniak, 1991; Parkin & Streete, 1988; Russo, Nichelli, Gibertoni, & Cornia, 1995; Snodgrass & Feenan, 1990; Snodgrass & Hirshman, 1991; Snodgrass & Surprenant, 1989; Takegata & Furutuka, 1993; Wyatt & Conners, in press). Implicit memory is demonstrated when, without intention, performance is affected by previous exposure to a particular stimulus or stimuli. Other picture fragment completion tests have also been used to measure implicit memory (e.g., DiGiulio, Seidenberg, O'Leary, & Raz, 1994; McDermott & Roediger, 1994; Srivinas, 1993; Warrington & Weiskrantz, 1968; Weldon, Roediger, Beitel, & Johnston, 1995), but the present study focused on the Snodgrass test. In the Snodgrass test, participants identify a set of fragmented pictures once and then a second time, intermixed with new pictures. Implicit memory is reflected in the decrease in identification difficulty for pictures that were previously presented. The present study investigated the alternate-form reliability of this test.

Two observations motivated the present study. One was that research on the development of implicit memory is increasing, and there is an increasing need for nonverbal tests that can be completed by young children. Of the many tests that have been used to measure implicit memory, most have used word stimuli. For example, word stem

completion tests and word fragment completion tests are among the most widely used (see Greene, 1992, for a discussion of implicit memory paradigms). An advantage to the picture fragment completion test is that it can be used with participants who have not learned to read or whose reading is poor. Several investigators have used a version of the test with children (e.g., Parkin & Streete, 1988; Russo et al., 1995; Wyatt & Conners, in press), and two have used a version with individuals with mental retardation (Takegata & Furutuka, 1993; Wyatt & Conners, in press). Although the present study did not include children or individuals with mental retardation as participants, it was an initial exploration of the reliability of a test that could be used for participants of many ages and learning levels.

The second observation that motivated the present study was that studies on individual differences in implicit memory have begun to emerge in which correlations are calculated among various measures of implicit and explicit memory (e.g., Perruchet & Baveux, 1989; Woltz, 1988; Woltz & Shute, 1993). In addition, basic studies of implicit memory report correlations from time to time (e.g., Parkin & Streete, 1988; Snodgrass, 1989a). Yet, there has been little discussion of reliability of implicit memory measures. In order to interpret correlations, it is important to have some knowledge of the reliability of the tests being used. The Snodgrass picture fragment completion test was a good candidate for a reliability study because it has been somewhat standardized and distributed on disk. Also, there are published "norms" for this test, corresponding to specific picture sets based on Snodgrass's work with college students (Snodgrass & Corwin, 1988; Snodgrass et al., 1987).

In the Snodgrass picture fragment completion test, participants view fragmented pictures and try to identify them. The pictures are fragmented by the random deletion of pixels. Each picture has eight versions, ranging in degree of fragmentation. A highly fragmented version of each picture is presented first (Version 1), and the participant guesses what it is. If the guess is wrong, a slightly less

Portions of this research were presented at the Southeastern Psychological Association meeting, April 1994, New Orleans. The authors thank Joan Gay Snodgrass for the use of her computerized word fragment completion program and for comments on an earlier version of the paper, and Sarah Pierce and Allison Sligh for assistance in data collection. Correspondence should be addressed to F. A. Conners, Department of Psychology, Box 870348, University of Alabama, Tuscaloosa, AL 35487-0348 (e-mail: fconners@gp.as.ua.edu).

fragmented picture is presented (Version 2), and so on, until the participant correctly identifies the picture. For each participant, an identification threshold is determined by calculating the mean version number at which the participant made a correct identification.

Participants complete a set of "training" pictures, do a filler task, and then complete a second set of pictures. The second set has twice as many pictures as the first. It contains the pictures presented in the first set as well as an equal number of new pictures. Implicit memory is reflected by the decrease in identification threshold for the "old" pictures relative to the "new" pictures.

Snodgrass and Corwin (1988) tested several sets of pictures and published mean identification thresholds for each set. In the present study, alternate sets of pictures, equivalent in identification thresholds based on Snodgrass' work, were used to test the alternate-form reliability of the test. In both Studies 1 and 2, college students completed the picture fragment completion test twice, using different picture sets. In Study 1, two sets of 15 pictures were used, as in the original test (Snodgrass et al., 1987). In Study 2, two sets of 30 pictures were used, as in some subsequent uses of the test (e.g., Snodgrass & Hirshman, 1991; Snodgrass & Surprenant, 1989). In Study 3, explicit retrieval instructions were substituted for the standard instructions to determine whether reliability estimates were specific to measurement of implicit memory.

STUDY 1

Method

Participants. The participants were 47 introductory psychology students, who participated for course credit. Two additional participants were tested but were eliminated from the sample because of errors in test administration.

Apparatus and Materials. A Macintosh SE/30 was used to present the tests. Snodgrass and Corwin's (1988) Picture Sets 6a and 6b and Sets 7a and 7b were used to produce alternate forms of the test. Snodgrass and Corwin found that all these sets have equivalent mean thresholds in the training phase and in the test phase. As a result, the sets are roughly equivalent in measures of implicit memory and skill learning. Furthermore, the sets are equated on 35th percentile identification threshold and have roughly the same number of exemplars in each of nine conceptual categories.

For half of the participants, Sets 6a and 6b were used for the first form of the test and Sets 7a and 7b were used for the second. For the other half of the participants, the order was reversed. Sets 6a and 7a were used for the training and old pictures, and Sets 6b and 7b were used for the new pictures.

Procedure. The procedure was the same for both forms of the test. The only difference between forms was the pictures used. The participants were tested individually on both forms within the same testing session. First, they were given task instructions verbally, and then they read the same instructions on the computer screen. Instructions are described in detail by Snodgrass et al. (1987) and Snodgrass and Corwin (1988). Essentially, the participants were told that the pictures were of common objects and animals, and that they would be shown at first in a very incomplete form, but that the completeness would increase with every version of the picture. The participants were told to press the return key to continue to the next version when they did not know what the picture was, and they were

told to type in the name of the picture when they did know what it was.

The participants were shown each picture in turn, starting with the most fragmented version of the picture and continuing with increasingly complete versions. When they correctly named a picture, the version number was recorded, and the complete picture was shown. They then went on to the next picture, starting with the most fragmented version. This procedure was repeated for all 15 training pictures.

After completing the 15 training pictures, the participants completed word fragments for a 5-min filler task. The participants then read the instructions for the second half of the test. The instructions included a statement that some of the pictures would be those that they had seen in the training phase. In the second half of the test, the participants completed the task with 30 pictures—15 old and 15 new pictures. After a short break, the participants performed the test again, following the same procedure, but this time with a different set of pictures. Testing lasted 50–60 min.

Results and Discussion

Measures. Four measures were used to reflect implicit memory. There were two *perceptual learning* measures (after Snodgrass, 1989a, 1989b, and Snodgrass et al., 1987) and two *savings* measures (after Parkin & Streete, 1988), which Snodgrass recognized (Snodgrass & Corwin, 1988) but did not use. The perceptual learning measures reflected the advantage in identification of old pictures over new pictures. The savings measures reflected improvement in identification of old pictures over training pictures. The savings measures have an advantage over the perceptual learning measures because they are based on the comparison of identification of the same set of pictures presented twice. If implicit memory is defined by increased sensitivity to stimuli *previously experienced*, then the savings measures tap the construct of implicit memory more directly. On the other hand, participants may do better on the second presentation simply because they have had practice, which has increased their skill in picture identification. Thus, the savings measures may overestimate implicit memory. The perceptual learning measures control for practice and skill learning because they are based on the comparison between performance on pictures that have been presented after practice. However, perceptual learning measures are only useful if the new and old picture sets produce the same identification thresholds when they are presented for the first time.

Absolute perceptual learning and *absolute* savings scores were derived merely by subtracting one mean threshold from another (absolute perceptual learning = threshold for new pictures – threshold for old pictures; absolute savings = threshold for training pictures – threshold for old pictures). As Snodgrass (1989a, 1989b) noted, the use of absolute scores implies that amount of learning is constant, regardless of how much is left to be learned. However, learning tends to take on a negatively accelerating function, whereby the amount learned decreases with the amount left to be learned. *Relative* perceptual learning and *relative* savings scores were derived to conform to this learning model. Following Snodgrass (1989a,

Table 1
Means and Standard Deviations for Thresholds and Derived Measures for Studies 1–3

Picture Set	Training				Threshold				Old				Perceptual Learning				Savings				Skill Learning			
	M		SD		M		SD		M		SD		M		SD		M		SD		M		SD	
	M	SD	M	SD	M	SD	M	SD	Absolute	Relative	Absolute	Relative	Absolute	Relative	Absolute	Relative	Absolute	Relative	Absolute	Relative	Absolute	Relative		
6	4.37	0.68	4.09	0.54	2.75	0.59	1.34	0.52	.43	.16	1.62	0.46	.49	.12	0.29	0.49	.07	.13	0.29	0.49	.07	.13		
	4.61	0.62	4.46	0.51	3.06	0.71	1.40	0.48	.41	.16	1.55	0.48	.44	.14	0.16	0.50	.03	.15	0.16	0.50	.03	.15		
10	4.31	0.51	4.22	0.47	2.76	0.54	1.46	0.34	.46	.11	1.55	0.36	.47	.11	0.10	0.34	.02	.10	0.10	0.34	.02	.10		
	4.58	0.57	4.43	0.51	3.15	0.63	1.28	0.38	.38	.12	1.44	0.33	.41	.11	0.15	0.29	.04	.08	0.15	0.29	.04	.08		
6	4.47	0.52	4.14	0.50	2.55	0.54	1.59	0.39	.51	.13	1.92	0.45	.56	.12	0.33	0.51	.09	.14	0.33	0.51	.09	.14		
	4.54	0.61	4.44	0.46	2.80	0.50	1.64	0.29	.48	.10	1.75	0.50	.49	.11	0.11	0.45	.02	.13	0.11	0.45	.02	.13		

1989b), they were derived by dividing the threshold difference by a relevant baseline threshold. For relative perceptual learning, the relevant baseline is the threshold for new pictures [relative perceptual learning = absolute perceptual learning/(threshold for new pictures - 1)]. For relative savings, the relevant baseline is the threshold for training pictures [relative savings = absolute savings/(threshold for training pictures - 1)].¹ Snodgrass (1989b) recommended using relative scores because they fit actual learning curves better than do absolute scores.

Two additional measures used by Snodgrass (1989a, 1989b)—*absolute skill learning* (threshold for training pictures - threshold for new pictures) and *relative skill learning* [absolute skill learning/(threshold for training pictures - 1)]—were also included. These are not measures of implicit memory but of the within-session increase in task-related skill experienced by participants from the first to second presentation of fragmented pictures. The skill learning measures were included because they have been considered part of the picture fragment completion test, and because they help interpret the savings scores. If there is no evidence of skill learning, then savings scores should be closely indicative of implicit memory; if there is a great deal of skill learning, then savings scores should be interpreted with caution.

Implicit memory and skill learning. Table 1 shows threshold means, implicit memory means (mean perceptual learning and mean savings), and skill learning means for each of the two forms of the test. To determine whether there was significant evidence of perceptual learning, mean thresholds for old pictures were compared with those for new pictures, using an alpha level of .05, as for all analyses in the study. There was significant perceptual learning for both Set 6 [$t(46) = 17.48, SE = .076, p < .001$] and Set 7 [$t(46) = 20.16, SE = .069, p < .001$]. Also, mean relative perceptual learning was significantly different from zero for both Set 6 [$t(46) = 18.22, SE = .024, p < .001$] and Set 7 [$t(46) = 17.37, SE = .024, p < .001$]. These analyses suggest that there was evidence of implicit memory, aside from any influence from skill learning.

To determine whether there was significant evidence of savings, mean thresholds for old pictures were compared with those for training pictures. There was significant savings for both Set 6 [$t(46) = 24.23, SE = .067, p < .001$] and Set 7 [$t(46) = 22.19, SE = .070, p < .001$]. Also, mean relative savings was significantly different from zero for both Set 6 [$t(46) = 27.68, SE = .018, p < .001$] and Set 7 [$t(46) = 21.36, SE = .021, p < .001$]. These analyses also suggest significant levels of implicit memory; however, these levels of savings could be inflated by skill learning.

Finally, analyses were conducted that would reveal evidence of skill learning. Mean thresholds for training pictures were compared with those for new pictures. There was significant skill learning for both Set 6 [$t(46) = 4.04, SE = .071, p < .001$] and Set 7 [$t(46) = 2.13, SE = .073, p < .05$]. Mean relative skill learning was significantly above zero for Set 6 [$t(46) = 3.65, SE = .019, p < .001$], but not for

Table 2
Alternate-Form and Split-Half Reliabilities for Studies 1–3

Reliability	Threshold			Perceptual Learning		Savings		Skill Learning	
	Training	New	Old	Absolute	Relative	Absolute	Relative	Absolute	Relative
Study 1 (15 pictures)									
Alternate Form	.506	.610	.817	.323	.570	.036	.546	-.066	-.039
Split-Half									
Set 6	.909	.782	.832	.518	.631	.566	.606	.475	.317
Set 7	.821	.697	.792	-.027	.323	.424	.218	.296	.327
Study 2 (30 pictures)									
Alternate Form	.704	.770	.831	.275	.544	.026	.501	-.189	-.151
Split-Half									
Set 10	.832	.807	.881	.323	.595	.532	.689	.130	-.071
Set 11	.895	.801	.905	.461	.719	.374	.646	.466	.336
Study 3 (15 pictures, explicit memory instructions)									
Alternate Form	.143	.749	.676	.427	.524	-.133	.375	-.008	-.053
Split-Half									
Set 6	.764	.650	.934	-.074	.295	.069	.359	.526	.452
Set 7	.831	.700	.660	-.768	-.133	.516	.406	.259	.086

Set 7 [$t(46) = 1.36$, $SE = .022$, $p > .05$]. Thus, overall, there was evidence of a practice effect, which indicates that savings measures in this study should not be considered good measures of implicit memory.

Reliability. To examine alternate-form reliability, the two perceptual learning measures, the two savings measures, and the two skill measures were correlated across forms (see Table 2). Three generalizations can be made from Study 1 about the alternate-form reliability of the picture fragment completion test: (1) implicit memory measures were low to moderate in reliability, (2) relative measures of implicit memory were more reliable than absolute measures, and (3) skill learning measures were not reliable. Scatterplots related to each reliability estimate showed no unusual shape and no influential outliers.

To assist in interpreting this set of reliabilities and to add breadth to the results, additional reliability estimates were calculated. Spearman-Brown split-half reliability of each perceptual learning, savings, and skill measure was figured separately for Set 6 and Set 7, by deriving and then correlating scores that were based on only odd-numbered pictures and only even-numbered pictures. There were 8 odd-numbered pictures and 7 even-numbered pictures. Also, both alternate-form and split-half reliabilities were calculated for training, old, and new threshold measures. All of these additional reliability estimates are included in Table 2.

The usual interpretation of low-to-moderate alternate-form reliabilities is that there is a substantial amount of measurement error in the test. This could be caused by poor internal consistency—a state in which the test items do not all work in a unified way to produce a total score. Possibly, poor internal consistency of either Set 6 or Set 7 or both was responsible for the low alternate-form reliabilities. For skill learning measures, split-half reliabilities were higher than were alternate-form reliabilities. For implicit memory measures, though split-half reliabilities for Set 6 were in line with the highest alternate-form reliabilities

from Study 1, split-half reliabilities from Set 7 were, in fact, lower. It is possible that poor internal consistency for Set 7 weakened the alternate-form reliabilities.

The threshold measures themselves, which are used to derive the measures of implicit memory and of skill learning, had generally higher alternate-form reliability than did the derived measures. Still, with a range of .506–.817, they were not optimal. Because derived measures can be only as reliable as the measures from which they are derived, the reliability of the threshold measures most certainly affected the reliability of the implicit memory and skill learning measures. On the basis of threshold reliabilities alone, it would be expected that skill learning measures would show the lowest reliability, followed by savings, followed by perceptual learning. This is, in fact, the pattern observed.

The split-half reliabilities of the threshold measures were the highest reliabilities in Study 1, ranging from .697 to .909. Those for Set 7 were lower than those for Set 6. However, the difference was not large enough to explain the large difference in split-half reliability of implicit memory and skill learning measures between Sets 6 and 7. It must be noted that the odd-number and even-number measures used to compute split-half reliabilities were based on only a few pictures, and, therefore, they would be expected to be somewhat volatile.

STUDY 2

In Study 2, the number of pictures in the picture fragment completion test was doubled. It was reasoned that if the weak reliabilities of Study 1 were due to high measurement error, additional pictures would help stabilize the threshold scores as well as the overall scores somewhat, and the result would be an increase in reliability.

Method

Participants. Forty-nine introductory psychology students participated in Study 2 to fulfill a course requirement. An additional 6

participants were tested, but their data were lost due to a computer-related error. Also, one participant did not complete the tests, and one did not follow instructions.

Apparatus and Materials. The same computer was used as that in Study 1. The pictures were drawn from the same pool described by Snodgrass and Corwin (1988). Snodgrass and Corwin's Sets 6a and 6b and Sets 9a and 9b were combined into what were called Sets 10a and 10b in the present study. Their Sets 7a and 7b and Sets 8a and 8b were combined into what were called Sets 11a and 11b in the present study. Snodgrass and Corwin showed that Sets 6–9 have equivalent mean thresholds when used as training, old, and new pictures, similar 35th percentile threshold, and similar distribution of exemplars across nine categories. Because the mean thresholds are equivalent, mean perceptual learning, mean savings, and mean skill learning presumably would also be equivalent.

As in Study 1, order of presentation of Sets 10 and 11 was counterbalanced. Sets 10a and 11a were used for training and old pictures, whereas Sets 10b and 11b were used for new pictures.

Procedure. The procedure was the same as that for Study 1, except that (1) the number of pictures viewed was doubled, (2) the testing took place in two separate 1-h sessions, and (3) number search puzzles were used for the filler task.

Results

Measures. The same six measures were calculated as those in Study 1: absolute perceptual learning, relative perceptual learning, absolute savings, relative savings, absolute skill learning, and relative skill learning.

Implicit memory and skill learning. Means and standard deviations for all implicit memory measures and skill learning measures, calculated separately for Sets 10 and 11, are shown in Table 1.

The *t* tests comparing old pictures with new pictures showed significant perceptual learning for both Set 10 [$t(48) = 30.44$, $SE = .048$, $p < .001$] and Set 11 [$t(48) = 23.95$, $SE = .054$, $p < .001$]. Mean relative perceptual learning was also significantly different from zero for both Set 10 [$t(48) = 28.93$, $SE = .016$, $p < .001$] and Set 11 [$t(48) = 21.97$, $SE = .017$, $p < .001$]. Thus, there was significant evidence of implicit memory aside from skill learning.

A comparison of old picture thresholds with training picture thresholds showed significant savings for both Set 10 [$t(48) = 30.26$, $SE = .051$, $p < .001$] and Set 11 [$t(48) = 30.41$, $SE = .047$, $p < .001$]. Mean relative savings was also significantly different from zero for both Set 10 [$t(48) = 29.56$, $SE = .016$, $p < .001$] and Set 11 [$t(48) = 25.79$, $SE = .016$, $p < .001$].

The *t* tests comparing training and new pictures showed marginally significant skill learning for Set 10 [$t(48) = 1.99$, $SE = .048$, $p = .053$] and significant skill learning for Set 11 [$t(48) = 3.70$, $SE = .041$, $p = .001$]. Relative skill measures were not significantly above zero for Set 10 [$t(48) = 1.39$, $SE = .014$, $p > .05$], but were significantly above zero for Set 11 [$t(48) = 3.48$, $SE = .012$, $p < .01$]. Though skill learning seemed weak, particularly in Set 10, there is enough evidence of skill learning to warrant caution in using savings measures of implicit memory.

Reliability. Alternate-form reliability estimates and Spearman-Brown split-half reliability estimates for the

implicit memory measures, the skill learning measures, and the threshold measures are listed in Table 2. For the Spearman-Brown estimates, there were 15 pictures in each of the odd and even halves for both Set 10 and Set 11.

Study 2 replicated the pattern of results of Study 1. Reliabilities were low to moderate for implicit memory, higher for relative measures of implicit memory than for absolute measures, and very low for skill learning. Doubling the number of pictures did not increase the reliability of the test. Scatterplots showed no unusual patterns or influential outliers.

Split-half reliabilities of the implicit memory measures increased slightly from Study 1. However, the increase did not result in an increase in alternate-form reliability. Split-half reliabilities for the skill learning measures did not increase from Study 1. For implicit memory measures and skill learning measures, split-half reliabilities were slightly higher than the corresponding alternate-form reliabilities. It seems that internal consistency was not the major factor resulting in low-to-moderate alternate-form reliability.

Alternate-form and split-half reliabilities of threshold measures were also slightly better in Study 2 than in Study 1 and were fairly good. However, the increase in threshold reliability did not result in an increase in reliability of implicit memory or skill learning measures.

STUDY 3

The low-to-moderate reliability estimates generated in Studies 1 and 2 may mean that the picture fragment completion test contains a large amount of error variance. An alternate interpretation, however, is that individuals simply do not greatly differ from one another systematically in implicit memory. In other words, there may not be strong individual differences in this aspect of memory. Although individual differences are fairly stable for many measures of memory, measures that are meant to reflect automatic memory processes may not show individual differences (e.g., Hasher & Zacks, 1979, 1984; Reber, 1991). In Study 3, the same pictures were used in the same format as in Study 1, but instructions were altered so as to prompt explicit memory rather than implicit memory. If reliability estimates did not increase, it would suggest that the manner of testing and/or the characteristics of the picture sets allow substantial measurement error. On the other hand, if reliability estimates did increase, it would suggest that reliability depends on the characteristic that the test is used to measure.

Method

Participants. Forty-three introductory psychology students participated in Study 3 to fulfill a course requirement. An additional participant was tested, but her data were not used because her first language was not English.

Apparatus and Materials. Apparatus and materials were identical to those used in Study 1.

Procedure. The procedure was the same as that for Study 1, except for the instructions. The participants heard and read the same

instructions used in Study 1 and described by Snodgrass et al. (1987) and Snodgrass and Corwin (1988). However, before beginning the training phase, the participants were told they should try to remember the fragments and pictures they would see because they would have to identify them again later. Before the test phase, the participants were told to think back on the fragments and pictures that they had previously identified and to use that information to identify the fragmented pictures they would soon see. They were told that if they could do this, they would be able to identify old pictures with fewer fragments, relative to new pictures.

Results

Measures. The same measures were calculated as those in Studies 1 and 2. However, because of the instructions to use explicit retrieval, the measures have different meanings. The training threshold, for example, reflects the mean level of fragment needed to identify a picture *while effortfully trying to encode the stimuli for later retrieval*. This is slightly different from the meaning of the training threshold in Studies 1 and 2. The old and new thresholds reflect the mean level of fragment needed to identify an old or new picture *while effortfully trying to retrieve information from the prior session*. The perceptual learning measures (absolute and relative) reflect the gain in identification resulting from explicit retrieval. The savings measures (absolute and relative) reflect the gain in identification resulting from explicit retrieval and from skill learning. Perceptual learning and savings measures are no doubt influenced to some degree by implicit memory. The skill learning measures (absolute and relative) reflect skill learning as before, but for a slightly different task from that used in Studies 1 and 2. Though the measures have different meanings in Study 3, for simplicity their names have not been changed.

Explicit memory and skill learning. Means and standard deviations for all measures, calculated separately for Sets 6 and 7, are shown in Table 1. Although training and new thresholds were comparable to those of Study 1, old thresholds were lower, and absolute and relative perceptual learning as well as relative savings were higher in Study 3. With explicit memory instructions, the participants benefited more from prior exposure. There was no systematic difference in skill learning between Studies 1 and 3.

The t tests comparing old picture thresholds with new picture thresholds showed significant perceptual learning for both Set 6 [$t(42) = 26.61, SE = .060, p < .001$] and Set 7 [$t(42) = 36.82, SE = .045, p < .001$]. Mean relative perceptual learning was also significantly different from zero for both Set 6 [$t(42) = 25.37, SE = .020, p < .001$] and Set 7 [$t(42) = 31.17, SE = .015, p < .001$]. Thus, there was significant evidence of memory aside from skill learning.

A comparison of old picture thresholds with training picture thresholds showed significant savings for both Set 6 [$t(42) = 28.25, SE = .068, p < .001$] and Set 7 [$t(42) = 23.14, SE = .076, p < .001$]. Mean relative savings were also significantly different from zero for both Set 6 [$t(42) = 30.27, SE = .019, p < .001$] and Set 7 [$t(42) = 28.87, SE = .017, p < .001$].

The t tests comparing training and new pictures showed significant skill learning for Set 6 [$t(42) = 4.24, SE = .078, p < .001$], but not for Set 7 [$t(42) = 1.55, SE = .069, p > .10$]. Relative skill measures were significantly above zero for Set 6 [$t(42) = 4.17, SE = .022, p < .001$], but not for Set 7 [$t(42) = 1.0, SE = .020, p > .20$]. Because skill learning was present for Set 6, savings measures should be used with caution.

Reliability. Alternate-form reliability estimates and Spearman-Brown split-half reliability estimates for the explicit memory measures, the skill learning measures, and threshold measures are listed in Table 2. To compute the Spearman-Brown estimates, the 8 odd-numbered pictures were used for one half and the 7 even-numbered pictures were used for the other half. This was done separately for Sets 6 and 7. Scatterplots showed no unusual shapes or influential outliers.

As in Studies 1 and 2, reliabilities were low to moderate for perceptual learning and savings, higher for relative measures of implicit memory than for absolute measures, and very low for skill learning. However, there were some specific differences in Study 3.

Compared with Studies 1 and 2, there was some increase in the alternate-form reliability for absolute perceptual learning, though there was little difference for relative perceptual learning. The increase for absolute perceptual learning is explained neither by differences in internal consistency nor by differences in threshold reliability. In Study 3, Set 6 had poor internal consistency (split-half reliability), and Set 7 had moderate internal consistency, whereas the reverse was true for Study 1. Overall, however, there was little difference in internal consistency. Although there was some fluctuation in reliability estimates for old and new thresholds from Study 1 to Study 3, there was, overall, no difference. The increase in alternate-form reliability for absolute perceptual learning could be attributable to the fact that the test deliberately elicited explicit memory. However, the reliability was still in the moderate-to-low range.

For absolute savings, alternate-form reliability remained around zero, though it decreased for relative savings. The decrease for relative savings may be explained by internal consistency or by threshold reliability. Internal consistency of savings measures was lower in Study 3 than in Study 1. Reliability for training and old thresholds was also lower in Study 3, particularly for training pictures.

For skill learning, alternate-form reliabilities remained at about zero, as in Studies 1 and 2. Split-half reliabilities were slightly higher for Set 6, but slightly lower for Set 7. The lower alternate-form reliability for the training threshold was offset somewhat by the higher alternate-form reliability for the new threshold.

DISCUSSION

The results suggest that alternate-form reliability of the Snodgrass picture fragment completion task is low to

moderate, depending on which measure is used. When an absolute difference measure of implicit memory was used (absolute perceptual learning or absolute savings), reliability was low, ranging from .026 to .323. When a relative difference measure was used, reliability was moderate, ranging from .501 to .570. Skill learning measures showed the lowest reliabilities, ranging from $-.189$ to $-.039$. Further research would be needed to know whether these results would be any different if participants were children rather than college students.

In Study 2, the length of the test was doubled in order to determine whether reliability was underestimated in Study 1 due to the small number of pictures. However, there was no effect of doubling the length of the test on reliability estimates of implicit memory and skill learning. It is possible that any increase in reliability afforded by doubling the number of pictures was offset by the split-session format used in Study 2. Because situational influences may change from one testing session to the next, more error variance could have been introduced into the Study 2 measures. However, alternate-form reliability of the threshold measures increased in spite of these conditions.

Or, it may be that the increase in test length from Study 1 to Study 2 was not sufficient to affect the stability of the measures. Perhaps, only if the test had been lengthened by four or five times would there have been an impact on reliability. Though there was an increase in reliability of the threshold measures as a result of doubling the number of pictures, it may take a larger increase to affect reliability of derived measures. Increasing the test by four or five times, however, has two drawbacks: the test would take well over an hour to complete, and the priming effect could diminish due to the amount of information to be processed. If it were the small number of pictures that caused the low reliability, then the results of the study apply to any similarly structured picture fragment completion test that uses up to 30 items.

It is also possible that, though individuals may differ somewhat in implicit memory, the differences simply are not strong—the low-to-moderate reliabilities are more reflective of lack of individual differences in an automatic function than of the level of measurement error in the test (see Hasher & Zacks, 1979, 1984, and Reber, 1991, for discussions of reduced individual differences in automatic processes). In one correlational study of implicit and explicit memory, Perruchet and Baveux (1989) reported low-to-moderate correlations among four different measures of implicit memory (ranging from $-.026$ to $.379$) and a higher correlation between two measures of explicit memory (.501). This pattern is consistent with the notion that there may not be great individual differences in implicit memory. On the other hand, Woltz and Shute (1993) found split-half reliabilities of .75 and .80 for implicit memory measures and an intercorrelation of .46. Furthermore, in the present Study 3, the Snodgrass picture fragment completion test was used to measure explicit memory. The task layout and pictures were identical to those in

Study 1, and only the instructions were changed. The picture fragment completion test was no more reliable as a test of explicit memory than as a test of implicit memory.

Other than the small number of pictures, there are three other aspects of the Snodgrass picture fragment completion test that may be related to the test's reliability. One is the constrained range of scores obtainable, which is a function of the number of fragment levels per picture. With eight fragment levels per picture, the possible range for threshold scores is 1–8, but in our study the actual range was about 3–6 (training and new) and 2–5 (old). Given these actual ranges, the possible ranges for absolute measure scores are 1–4 (perceptual learning and savings) and -3 – -3 (skill learning). The possible ranges for relative measure scores are 0–1.25 (perceptual learning and savings) and -3 – -0.67 . Although Snodgrass's test incorporated more fragment levels per picture than did Gollin's (1960), which had only five, it may need even more than eight levels to achieve good reliability. If a constrained range of possible scores affected the reliability of the test, then similar picture fragment completion tests using eight fragments or fewer would be also expected to suffer in reliability.

Another aspect of the Snodgrass picture fragment completion test that may be related to its reliability is the fact that the scores produced by the test are derived, or computed, from more basic scores. Error variance in basic scores is compounded in derived scores. In the present study, in almost every case, the basic scores were more reliable than the derived scores. This aspect of the test is shared with many tests of implicit memory, including those that do not use pictures. Thus, if the use of derived scores was what compromised the reliability of Snodgrass's test mediocre, then the problem applies more broadly in the implicit memory field.

Finally, it may be that the alternate-forms of the test—equivalent in Snodgrass's samples (Snodgrass & Corwin, 1988; Snodgrass et al., 1987)—were not equivalent for the samples in the present study. Although Snodgrass's normative study included 100 participants, only 10 were used for each picture set. Sampling error may have resulted in construction of picture sets that were not optimal. Our data suggest that this is a real possibility. For example, in Study 1, when used as training pictures, Set 6a produced a significantly lower mean threshold than did Set 7a [$t(46) = 2.53, p < .02$]. When used as old pictures, Set 6a also produced a significantly lower mean threshold than did Set 7a [$t(46) = 5.13, p < .001$]. When used as new pictures, Set 6b produced a significantly lower mean threshold than did Set 7b [$t(46) = 5.45, p < .001$]. There was no difference in absolute or relative perceptual learning, absolute savings, or absolute and relative skill learning ($p > .10$). However, there was significantly more savings for Set 6 than for Set 7 [$t(46) = 2.62, p < .02$]. A similar pattern was present for Study 2 in the comparison of Sets 10 and 11, with Set 10 producing lower thresholds and more absolute savings. However, Set 10 also produced

more absolute perceptual learning [$t(48) = 2.81, p < .01$] and relative perceptual learning [$t(48) = 4.84, p < .001$] than did Set 11. The alternate-form reliabilities in the present study may have been higher had the alternate forms been more equivalent. If alternate-form nonequivalence led to low reliability, then the results do not necessarily apply to picture fragment completion tests or implicit memory tests other than Snodgrass's.

The results of the present study provide a perspective for researchers who may be considering using the Snodgrass picture fragment completion task in correlational studies. First, it is important to distinguish between perceptual learning and savings measures. Savings measures may seem to reflect the construct of implicit memory better than perceptual learning measures, because they directly assess the impact of having had previous experience with specific stimuli. However, skill learning is inherent in savings measures. In the present study, there was ample evidence of significant skill learning to warrant preference for the perceptual learning measures over the savings measures. Still, perceptual learning measures should be used only if it is clear that the sets of pictures used for training and new pictures are, in fact, equivalent. Second, relative measures tend to be more reliable than absolute measures. An analysis by Snodgrass (1989b) showed that college students' performance on the picture fragment completion task clearly fit a relative learning curve better than an absolute learning curve. Thus, it is likely that the relative measures come closer to measuring the construct of implicit memory (as measured by this test) than do the respective absolute measures. The results of the present study suggest that the more valid measures are also more reliable. Thus, correlational researchers would do best to use relative measures.

Finally, correlations involving measures from the Snodgrass picture fragment completion task should be interpreted with reliability in mind. In the present study, alternate-form reliability peaked at .570, and split-half reliability peaked at .719 (both for relative perceptual learning). Therefore, correlations involving these measures that are in the moderate range should be considered actually quite high.

The results are also pertinent to researchers who wish to use the Snodgrass picture fragment completion task in experimental studies. Picture sets that are equivalent in one sample may not be equivalent in other samples. In the present study, we were able to compare Sets 6a and 7a, Sets 6b and 7b, Sets 10a and 11a, and Sets 10b and 11b. Though Snodgrass and Corwin (1988) found these sets to be equivalent, they were not optimally equivalent in the present study. Caution should be used when choosing picture sets for training/old and new components of picture fragment completion tests. Interpretation of perceptual learning measures derived from this test (new threshold – old threshold) should also be done carefully, because these measures assume equivalence of picture sets. Finally, the practice of complete counterbalancing of items

should not be relaxed because picture sets have been deemed equivalent.

REFERENCES

- BONDI, M. W., & KASZNAK, A. W. (1991). Implicit and explicit memory in Alzheimer's disease and Parkinson's disease. *Journal of Clinical & Experimental Neuropsychology*, **13**, 339-358.
- GIULIO, D. V., SEIDENBERG, M., O'LEARY, D. S., & RAZ, N. (1994). Procedural and declarative memory: A developmental study. *Brain & Cognition*, **25**, 79-91.
- GOLLIN, E. S. (1960). Developmental studies of visual recognition of incomplete objects. *Perceptual & Motor Skills*, **11**, 289-294.
- GREENE, R. L. (1992). *Human memory: Paradigms and paradoxes*. Hillsdale, NJ: Erlbaum.
- HASHER, L., & ZACKS, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General*, **108**, 356-388.
- HASHER, L., & ZACKS, R. T. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist*, **39**, 1372-1388.
- MCDERMOTT, K. B., & ROEDIGER, H. L., III (1994). Effects of imagery on perceptual implicit memory tests. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 1379-1390.
- PARKIN, A. J., & STREETE, S. (1988). Implicit and explicit memory in young children and adults. *British Journal of Psychology*, **79**, 361-369.
- PERRUCHET, P., & BAVEUX, P. (1989). Correlational analyses of explicit and implicit memory performance. *Memory & Cognition*, **17**, 77-86.
- REBER, A. S. (1991). Implicit and explicit learning: Individual differences and IQ. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 888-896.
- RUSSO, R., NICHELLI, P., GIBERTONI, M., & CORNIA, C. (1995). Developmental trends in implicit and explicit memory: A picture completion study. *Journal of Experimental Child Psychology*, **59**, 566-578.
- SNODGRASS, J. G. (1989a). How many memory systems are there really?: Some evidence from the picture fragment completion task. In C. Izawa (Ed.), *Current issues in cognitive processes: The Tulane Flowerree Symposium on Cognition* (pp. 135-173). Hillsdale, NJ: Erlbaum.
- SNODGRASS, J. G. (1989b). Sources of learning in the picture fragment completion task. In S. Lewandowsky, J. C. Dunn, & K. Kirsner (Eds.), *Implicit memory: Theoretical issues* (pp. 259-282). Hillsdale, NJ: Erlbaum.
- SNODGRASS, J. G., & CORWIN, J. (1988). Perceptual identification thresholds for 150 fragmented pictures from the Snodgrass and Vanderwart picture set. *Perceptual & Motor Skills*, **67**, 3-36.
- SNODGRASS, J. G., & FEENAN, K. (1990). Priming effects in picture fragment completion: Support for the perceptual closure hypothesis. *Journal of Experimental Psychology: General*, **119**, 276-296.
- SNODGRASS, J. G., & HIRSHMAN, E. (1991). Theoretical explanations of the Bruner-Potter (1964) interference effect. *Journal of Memory & Language*, **30**, 273-293.
- SNODGRASS, J. G., SMITH, B., FEENAN, K., & CORWIN, J. (1987). Fragmenting pictures on the Apple Macintosh computer for experimental and clinical applications. *Behavior Research Methods, Instruments, & Computers*, **19**, 270-274.
- SNODGRASS, J. G., & SURPRENANT, A. (1989). Effect of retention interval on implicit and explicit memory for pictures. *Bulletin of the Psychonomic Society*, **27**, 395-398.
- SRIVINAS, K. (1993). Perceptual specificity in nonverbal priming. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 582-602.
- TAKEGATA, R., & FURUTUKA, T. (1993). Perceptual priming effect in mentally retarded persons: Implicit and explicit remembering. *Japanese Journal of Educational Psychology*, **41**, 176-182.
- WARRINGTON, E. K., & WEISKRANTZ, L. (1968). New method of testing long-term retention with special reference to amnesic patients. *Nature*, **217**, 972-974.
- WELDON, M. S., ROEDIGER, H. L., III, BEITEL, D. A., & JOHNSTON, T. R. (1995). Perceptual and conceptual processes in implicit and ex-

- plicit tests with picture fragment and word fragment cues. *Journal of Memory & Language*, **34**, 268-285.
- WOLTZ, D. J. (1988). An investigation of the role of working memory in procedural skill acquisition. *Journal of Experimental Psychology: General*, **117**, 319-331.
- WOLTZ, D. J., & SHUTE, V. J. (1993). Individual difference in repetition priming and its relationship to declarative knowledge acquisition. *Intelligence*, **17**, 333-359.
- WYATT, B. S., & CONNERS, F. A. (in press). Implicit and explicit memory in individuals with mental retardation. *American Journal on Mental Retardation*.

NOTE

1. This negatively accelerating learning function, expressed in proportions (or probabilities) is $P(t+1) = P(t) + \phi[1 - P(t)]$. The

probability of learning on the next trial is equal to the probability of learning on the present trial plus the learning parameter times the probability of not learning on this trial. In terms relevant to the picture fragment completion test, the proportion of old pictures identified is equal to the proportion of baseline pictures identified plus the learning parameter times the proportion of baseline pictures not identified. This function can be used to solve for the learning parameter: $\phi = [P(t+1) - P(t)]/[1 - P(t)]$. Converted for use with means rather than proportions, $\phi = [M(t+1) - M(t)]/M(t) - 1$ (see Snodgrass, 1989a). Thus, relative perceptual learning = $[M(\text{new}) - M(\text{old})]/M(\text{new}) - 1$, and relative savings = $[M(\text{train}) - M(\text{old})]/M(\text{train}) - 1$.

(Manuscript received July 12, 1995;
revision accepted for publication December 11, 1996.)