THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 2000, *53A* (1), 211–233

Low Reliability of Perceptual Priming: Consequences for the Interpretation of Functional Dissociations Between Explicit and Implicit Memory

Beat Meier and Walter J. Perrig
University of Bern, Bern, Switzerland

In this study, three experiments are presented that investigate the reliability of memory measures. In Experiment 1, the well-known dissociation between explicit (recall, recognition) and implicit memory (picture clarification) as a function of age in a sample of 335 persons aged between 65 and 95 was replicated. Test–retest reliability was significantly lower in implicit than in explicit measures. In Experiment 2, parallel-test reliabilities in a student sample confirmed the finding of Experiment 1. In Experiment 3, the reliability of cued recall and word stem completion was investigated. There were significant priming effects and a dissociation between explicit and implicit memory as a function of levels of processing. However, the reliability of implicit memory measures was again substantially lower than in explicit tests in all test conditions. As a consequence, differential reliabilities of direct and indirect memory tests should be considered as a possible determinant of dissociations between explicit and implicit memory as a function of experimental or quasi-experimental manipulations.

In this paper we focus on the reliability of implicit and explicit memory measures. In the last two decades there has been continuous interest in the study of implicit and explicit memory. Tests of explicit memory or direct tests of memory require subjects to retrieve prior information in a deliberate manner. By contrast, in a test of implicit memory no intentional recollection of prior experiences is required, and memory is measured indirectly. Retention is indicated when performance on studied items exceeds that of similar items not presented in a prior study phase, a phenomenon commonly referred to as priming (Schacter, 1987). The distinction between these two types of memory tasks has been justified by empirical evidence that seems to be reliably grounded (Graf & Masson, 1993; Lewandowsky, Dunn, & Kirsner, 1989; Perrig, Wippich, & Perrig-Chiello,

Requests for reprints should be sent to Beat Meier, Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, B.C., Canada V6T 1ZA. Email: beat@cortex.psych.ubc.ca

This research was supported by Grant 40–35642.93 of the Swiss National Science Foundation and by the Foundation for the Advancement of Scientific Research at the University of Bern. We thank P. Perruchet and R. Russo for their comments on an earlier draft of this article.

1993; Richardson-Klavehn & Bjork, 1988; Roediger & McDermott, 1993). Additionally, as is suggested by the transfer-appropriate framework (Roedinger & Blaxton, 1987), direct as well as indirect memory tests can be differentiated into perceptual and conceptual tasks. In this paper, however, we focus on perceptual priming tasks.

One of the original arguments supporting the distinction between implicit and explicit memory draws on the logic of functional dissociation. Evidence for a dissociation between performance in these two kinds of test has been based on demonstrations of functional and statistical independence between the two types of test. The most compelling evidence for functional independence comes from testing amnesic patients whose performance on explicit memory tests is severely impaired whereas their performance on some implicit tests is normal (Moscovitch, Winocur, & McLachlan, 1986; Squire, Knowlton, & Musen, 1993). There are other findings that suggest that perceptual priming is not influenced by development (Carroll, Byrne, & Kirsner, 1985; Parkin & Streete, 1988), by mental retardation (Lorsbach & Worman, 1989, 1990; Perrig & Perrig, 1995), by normal ageing (Light & Singh, 1987; Mitchell, 1989; Perrig & Perrig, 1993), by schizophrenia (Bazin & Perruchet, 1996; Schwartz, Rosse, & Deutsch, 1993), or by depression (Denny & Hunt, 1992; Hertel & Hardin, 1990).

Additionally, there are several experimental manipulations that affect explicit but not implicit memory. Perceptual priming seems not to be influenced by depth of processing (Jacoby & Dallas, 1981; Roediger, Weldon, Stadler, & Riegler, 1992), length of retention intervals (Kolers, 1976; Tulving, Schacter, & Stark, 1982), division of attention (Eich, 1984; Parkin & Russo, 1990), or intake of alcohol (Hashtroudi, Parker, deLisi, Wyatt, & Mutter, 1984). On the other hand, changes in presentation format or modality have an effect on perceptual priming but not on explicit memory measures (Berry, Banbury, & Henry, 1997; Weldon & Roediger, 1987). If there is no overlap of perceptual characteristics between study and test, perceptual priming is eliminated or substantially decreased. Therefore, many implicit memory tasks such as picture identification, fragment completion, or word stem completion are assumed to rely on a perceptual representation.

In sum, there is broad evidence that there are many individual difference variables and experimental factors that influence memory performance without altering perceptual implicit memory performance, whereas there is only the change of perceptual characteristics between study and test that shows an effect on perceptual priming. The explanation for these dissociation effects is still a matter of unsettled debate. Dissociation effects are assumed to reflect different memory systems (Squire et al., 1993; Tulving & Schacter, 1990) or different types of process (Blaxton, 1989; Roediger & Blaxton, 1987).

In this paper we report data that demonstrate substantially lower reliability for implicit memory tests than for explicit memory measures. This forces a different perspective on the discussion of dissociation effects. Most evidently, it will be more difficult to show an effect of an experimental manipulation for any measure with a lower reliability. Therefore we are confronted with the possibility that with measures of perceptual priming, potentially existing effects cannot be found due to low reliability. This does not contradict the fact that priming as a group effect can be reliably demonstrated. The impact of our argument can be made clear as soon as one considers the advance of the theoretical and practical development of measurement of memory with a view to integrating the two areas of experimental research and the quasi-experimental or individual-difference approach (e.g. Geiselman, Woodward, & Beatty, 1982). For instance, there seem to be reasons to expect that it will be possible to discriminate amnesic patients with different aetiologies by the use of different implicit memory tests (Butters, Heindel, & Salmon, 1990; Gabrieli et al., 1994). Butters et al. (1990) presented evidence that patients in the early stages of Alzheimer's disease and Huntington's disease both being severely impaired on explicit memory tasks can be dissociated by implicit tests. The Alzheimer's disease patients showed little perceptual priming, but performed like control subjects on a skill-learning task. In contrast, the Huntington's disease patients were unable to acquire the skill-learning task, but showed normal priming effects. At this point, it is mandatory that memory tasks be evaluated not only on the basis of experimental criteria but also according to the psychometric standards of the individual-difference approach. The critical question concerns the reliablity of an individual's characterization, or the reliability of the assignment to one group or another. In the light of these considerations, it is essential to assess reliability of tests and validity of constructs.

In this study we present three experiments that demonstrate the differences of reliability between explicit memory performances and implicit memory measures that are assumed to measure perceptual priming. In Experiments 1 and 2, the reliabilities of perceptual clarification, free recall, and recognition are investigated. In Experiment 3, the reliabilities of word stem completion and cued recall are assessed. Experiment 1 was conducted as part of a longitudinal study with a test–retest interval of 2 years. In Experiment 2, we considered the parallel-test reliability of a student sample. Finally, in Experiment 3, a stem completion task was used with another student sample in order to generalize our findings to another perceptual priming measure. From both the review of the literature and our data, we conclude that it is timely to concentrate more systematically on this research question.

In experimental memory research, little time and effort have been devoted to the question of how reliable experimental tasks are. If theoretically predicted effects can be replicated, this is taken as an evidence for the sufficient reliability of the tasks. As far as we know, no study has yet been published that addresses the issue of reliability of implicit and explicit memory measures. However, there is some related work where reliability estimates are reported, which will be reviewed here.

Salthouse and Meinz (1995) estimated a reliability of r = .78 and r = .83 for a computational span and a reading span measure, boosting the correlation with the Spearman–Brown formula. Using this same formula, Salthouse and Babcock (1991) estimated split-half reliabilities of computational span and listening span as r = .90 and r = .86, respectively. Finally, Salthouse (1992a, 1992b) reported intercorrelations between a digit span and a listening span of r = .50 and r = .62, respectively. All these span measures are supposed to measure working memory and are comparable to memory tests that are used in many test batteries of intelligence and cognitive abilities.

McDonald-Miszczak, Hertzog, and Hultsch (1995) report two longitudinal studies where more typical explicit memory measures were used. In Study 1, three word list free recall tasks and three story free recall tasks were administered to a sample of 231 adults (age range 22 to 78 years) with a retest interval of 2 years. A longitudinal factor analysis using all recall measures as multiple indicators of two latent variables "word recall" and "text recall" revealed disattenuated stability estimates of .92 and .94, respectively. In study 2, 234 adults (55–86 years) were tested three times over 6 years with two word recall tasks and two text recall tasks similar to those in Study 1. Additionally, two sets of general knowledge tasks were administered at each time of testing as a specification of "fact recall". Longitudinal structural equation models produced 6-year stability coefficients of .87, .89, and .88 for word recall, text recall, and fact recall, respectively. This indicates that, in this study, stability of individual differences in explicit memory was quite high with subjects over a broad range of age. Although we cannot compare the stability scores, the corrected split-half correlation scores, and the retest-correlation coefficients directly, these reviewed studies show that explicit memory tasks, in particular memory span and recall tasks, are adequately reliable.

To our knowledge, there is no study assessing the reliability of an implicit memory test. However, Perruchet and Baveux (1989) used different implicit and explicit memory measures in a correlational study. Therefore, the intercorrelations among implicit and among explicit memory tasks can be used as indicators for the reliability and validity of each construct. They found a highly significant correlation of $r = .50$ between recall and recognition, whereas the correlations between four implicit memory measures (perceptual clarification, word-fragment completion, perceptual identification, and anagram solution) were lower and ranged from $r = -.25$ between perceptual clarification and perceptual identification to $r = .31$ between perceptual clarification and word-fragment completion. In a factor analysis, Perruchet and Baveux obtained a factor structure where two indirect tests of memory (clarification and word stem completion) loaded on the same factor as recall and recognition. They interpreted this result as a dissociation between implicit memory measures. Clarification and completion were interpreted to be susceptible to contamination of explicit memory strategies, in contrast to identification and anagram solution, which are ubiquitous indicators of implicit memory. On the other hand, in a factor analysis by Mitchell (1989), who assessed several explicit and one implicit measure, the implicit picture-naming task loaded on a factor that was distinct from the other two factors characterized as episodic and semantic memory. However, the implicit measure as a difference score based on reaction times loaded on the same factor as another speed-related measure (semantic retrieval efficiency). Therefore it seems possible that Mitchell obtained a separate processing speed factor.

Hultsch, Masson, and Small (1991) tested subjects over a broad range of age with a word stem completion task as indirect test. Fact recall, story recall, and word recall were used as direct tests. In all age groups they found significant relationships between the different explicit memory scores, between $r = .35$ and $r = .54$, whereas the correlation between word stem completion and any of the explicit measures failed to reach significance, $r < .14$, in each of the age groups. To exclude unreliability of this measure, Hultsch et al. computed Cronbach alphas on randomly selected halves of the stem completion list. With one exception, the internal consistencies were within a range of .61 to .69. Therefore the stem completion measure was judged to be acceptably reliable. However, Hultsch et al. gave no information about the consistency of the explicit measures. Therefore it might well be that consistency of the explicit measures is much higher. Interestingly, Hultsch et al. found small but significant effects of age in the implicit memory task.

To summarize, all these studies show that explicit memory has been measured quite reliably, while it would be premature to draw final conclusions about implicit memory measures. Although the procedures for estimating the reliability of the tests differ to some extent from study to study, there is a clear pattern in the reliability scores: Tasks where subjects are directly and explicitly encouraged to remember something show reliability coefficients in the range from .50 to over .90, whereas intercorrelations between different implicit memory measures show lower correlations. Given that the expected maximum correlation of a test with any other test is limited by its reliability, it is clear that the extent to which relationships with other variables can be established is restricted by the reliability of that measure itself. Seen in this light, the pattern of results from the studies reviewed earlier might reflect the fact that the low correlations of implicit memory measures with any other variable (e.g. Hultsch et al., 1991; Perruchet & Baveux, 1989) are at least partly due to the low reliabilities of these measures. There is a direct relationship between power and the reliability of a test, which is explained briefly. Reliability is defined by the proportion of true variance to observed variance. The observed variance can be expressed as the sum of the true variance and the error variance. Lower reliability leads to a higher proportion of the error variance in the observed variance. If the influence of an independent variable on two different kinds of tests is investigated, and one of these tests has a low reliability whereas the other test has a high reliability, then the test with low reliability must show a greater effect than the more reliable test to achieve significance. Therefore, for a given effect a measurement instrument with a higher reliability has more statistical power and a better chance of showing a significant effect of an independent variable than has a measurement instrument with low reliability (Hallahan & Rosenthal, 1996). For this reason differential reliability of implicit and explicit memory measures may be a determinant of the dissociations that have been found between implicit and explicit measures. These dissociations should be reduced when more powerful procedures are applied. Indeed, the results of meta-analyses, where variables like age or levels of processing are examined, show that these experimental and quasi-experimental manipulations affect implicit memory as well as explicit memory measures (Brown & Mitchell, 1994; Challis & Brodbeck, 1992; La Voie & Light, 1994).

Although the data pattern of the reported reliability scores is quite clear and suggestive, it is still small. Therefore, we conducted the following experiments to investigate the hypothesis that perceptual priming measures are less reliable than explicit memory tests. We selected a quasi-experimental approach with a large sample (Experiment 1), and experimental settings with standard sample sizes (Experiments 2 and 3).

EXPERIMENT 1

This study is part of the Basle interdisciplinary project on ageing (IDA-Project; Perrig-Chiello, Perrig, Stähelin, Krebs, & Ehrsam, 1996), which is a continuation of a medical longitudinal study that had collected mostly biomedical data from a sample of initially 6400 healthy people in 1960, 1965, and 1971 (Widmer, Stähelin, Nissen, & da Silva, 1981). From this sample, 335 persons were additionally tested in 1993 and 1995 with three direct (Free Recall 1, Free Recall 2, and recognition) and three indirect memory tasks (Perceptual Clarification 1, 2, and 3). The same tasks were used on both test occasions, 1993 and 1995, respectively, and the analyses presented here were designed to investigate the stability of implicit and explicit memory performances.

Method

Participants

For the Base IDA-Project, subjects had to be older than 65 years and were randomly selected from a larger sample (comprising 2959 men and 809 women) in 1993. Out of 848 invited persons, 442 healthy older persons (309 men and 133 women) who ranged in age from 65 to 93 years (M = 74.9, SD = 6.5) participated at the first test occasion in 1993. Out of these, 335 persons (236 men and 109 women) who ranged in age from 67 to 95 years (M = 76.1, SD = 6.2) also participated at the retest 2 years later. For this study we concentrated on the longitudinal data, which meant that we only analysed the data of the 335 persons who had participated at both test occasions (1993 and 1995).

Material

For memory assessment we used a computerized test, which was developed in our laboratory (Perrig et al., 1994). This 30-min procedure measures explicit and implicit memory. No demanding computer handling is required of the participants. The test consists of three direct memory tasks (Free Recall 1, Free Recall 2, and recognition) and three indirect memory tasks (Perceptual clarification 1, Perceptual clarification 2, and Perceptual clarification 3).

Four sets of 15 pictures of easy-to-name objects were selected from the Snodgrass and Vanderwart (1980) material.

For the study phase, the 15 pictures of the first set and additional elements (words, numbers, patterns, fragmented objects) were used to construct a complex scene. This scene was copied twice on the same screen, one on the left side, the other on the right side. However, in the left image 18 small parts were missing or changed. The pictures of the three other sets were used as new items in the clarification tasks.

Procedure

All participants were tested individually. They were instructed to sit in front of a computer screen and to make themselves comfortable. In the study phase of the experiment, participants saw a picture that contained two identical scenes on the computer screen. The left images differed from the right on 18 details for which the participants were asked to search and to show to the experimenter. They were also instructed to study the picture in such a way that they could talk about it later. After 3 mins the picture disappeared. The study phase was followed by the first perceptual clarification test where 15 pictures out of the study phase picture (Set 1) and 15 new pictures (Set 2) were presented in a clarification procedure, in which each picture was drawn on the screen pixel by pixel. From the subjective phenomenological perspective, the participants saw small black points appear on a white screen, continuously constructing a picture, which became clearer and clearer (Figure 1). The pictures were presented in a fixed random order. The task was to name each picture as quickly as possible. Before each clarification process a fixation cross appeared in the middle of the screen for 500 msec. Then the fixation cross disappeared and the clarification process began. Timing started with the beginning of the picture presentation and was stopped by the experimenter when the picture was named. If the participant did not name a picture within 20 sec, it disappeared and was coded as missing. Only the data of correctly named pictures were analysed. Because we expected that naming

FIG. 1. Example of the perceptual clarification procedure (test phase of Experiment 1).

times were positively correlated with age, relative priming scores were defined as the percentage gain in naming speed of the old pictures compared to the new pictures: Priming 1 = [(new − old)/new] × 100.

As a second task the picture-naming task was repeated with the old pictures from the original scene (Set 1) and another set of new pictures (Set 3). At the same time subjects had to respond to little suns flashing on either the left or the right side of the computer screen for 200 mecs. This dual task was considered as a measure of processing capacity. This allows measurement of Priming 2 under a condition of divided attention. However, it has to be considered that potentially different results in Priming 2 compared to Priming 1 cannot easily be attributed to divided attention, because the old pictures in the test phase were the same as those in Priming 1 (Set 1). Again, Priming 2 was computed as the percentage gain in naming speed of the old pictures. Nevertheless, and more important for the purpose of the present study, we included an additional indirect memory measure to analyse the reliability of perceptual priming tasks.

In the second test session (1995) a third priming task, an adaptation of the Gollin task (Gollin, 1960), was added. This task has been used in several quasi-experimental studies (e.g. Gabrieli et al., 1994; Grafman et al., 1990; Heindel, Salmon, & Butters, 1990). In this task the new items of the second perceptual clarification test served as the old items (Set 3). In the test phase fragments of old and new pictures (Set 4) were presented for 3 sec at five levels of fragmentation. The pictures consisted of 10%, 20%, 30%, 40%, and 50% of black pixels of the original at Levels 1, 2, 3, 4, and 5, respectively. When a picture was not named correctly at a previous stage it appeared again in a less fragmented state. If a picture was not named correctly at Level 5 it was coded as Level 6. Priming 3 was defined as the percentage of gain in level of fragmentation of previously presented pictures in picture naming: Priming 3 = [new − old)/new] × 100.

Two free recall tests and a recognition test were administered on both test occasions after the second clarification task. In the first free recall task participants were asked to remember every object that had been presented in the picture of the study phase. In the second free recall task they were asked to recall the differences that they had detected in the two scenes. Thus the second free recall task was more intensely related to the self-performed task of error scanning. In the recognition task, 15 objects from the initial scene in the study phase as well as 15 objects that were used in the picture-naming task later on in the test procedure were presented together on the computer screen. Participants were asked to identify only those elements that had been presented in the initial scene where they had to find the differences at the beginning. Because each object has been presented at least once in the whole test procedure this recognition task can be considered as a kind of an exclusion task in Jacoby's (1991) terms.

Analysis

For data analysis we used a correlational approach. The correlations between the indirect tests and the direct tests, respectively, as well as retest reliabilities for the different tasks were analysed.

Results and Discussion

Means, standard deviations, inter-task correlations, and retest correlations of all measures (in italics) are presented in Table 1. At the top of the matrix implicit and explicit tests from Session 1 are displayed, and beneath them implicit and explicit tests from Session 2. Finally age and its relation to the other variables are shown.

Performance in Indirect Memory Tests. Participants showed substantial priming in all tasks at both measurements. In Session 1, the mean naming times of the old pictures and the new pictures were 4.36 sec and 4.88 sec for Priming 1 and 4.41 sec and 5.59 sec for Priming 2. In Session 2, the mean naming times of the old pictures and the new pictures were 4.07 sec and 4.8 sec for Priming 1 and 4.2 sec and 5.49 sec for Priming 2. The mean fragmentation levels at which pictures were named correctly were 2.17 and 2.64 for the old pictures and the new pictures, respectively. *T* tests of the naming times (for Priming 1 and Priming 2) and the fragmentation level (Priming 3) for old and new pictures were highly significant for all old–new comparisons, $p < .01$ with t values between $t(334) = 19.5$ and $t(334) = 35.8$.

Further, a correlational analysis showed that the priming scores did not correlate strongly with age. The correlations in Session 1 were $r = -.06$, $p > .05$, for Priming 1, and $r = -.03$, $p > .05$, for Priming 2; correlations in Session 2 were $r = -.19$, $p < .01$, for Priming 1; $r = .01$, $p > .05$, for Priming 2, and $r = -.15$, $p < .01$, for Priming 3.

The intercorrelations between the different priming tasks within each measurement were generally low, and two out of four correlations were not even significant. Test–retest correlations over the period of 2 years were $r = .13$, $p < .05$, for Priming 1, and $r = .39$, $p < .01$, for Priming 2. Scatterplots of these correlations are depicted in Figure 2. From these plots it can be seen that the low reliabilities are not due to truncated range.

Performance in Direct Memory Tests. The performances of all free recall and recognition tasks were significantly related to age, showing correlations between $r = -.37$, $p < .01$, and $r = -.49$, $p < .01$.

The intercorrelations between the explicit memory tasks turned out to be highly significant on both test occasions, showing correlations between $r = .45$, $p < .01$, and $r = .72$, $p < .01$. Furthermore, test–retest reliabilities were highly significant, with

TABLE 1

Means, standard deviations, inter-task correlations, and reliability coefficients for explicit and implicit memory measures in Experiment 1

	M	*SD*	*(1)*	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*	*(11)*
Session 1													
Implicit Tests													
(1) Priming 1	10.6	10.74	1.00										
(2) Priming 2	21.3	8.78	.22	1.00									
Explicit Tests													
(3) Free Recall 1	0.25	0.12	.18	.05	1.00								
(4) Free Recall 2	0.21	0.14	.09	.06	.62	1.00							
(5) Recognition	1.45	0.83	.18	.06	.61	.45	1.00						
Session 2													
Implicit Tests													
(6) Priming 1	15.3	10.5	*.13*	.01	.13	.15	.16	1.00					
(7) Priming 2	23.5	10.9	.12	*.39*	.08	.01	.01	.08	1.00				
(8) Priming 3	17.8	11.1	.06	.04	.11	.14	.07	.13	.13	1.00			
Explicit Tests													
(9) Free Recall 1	0.28	0.15	.09	.08	*.64*	.62	.49	.17	.04	.10	1.00		
(10) Free Recall 2	0.23	0.16	.11	.04	.55	*.63*	.51	.11	.05	.15	.72	1.00	
(11) Recognition	1.55	0.91	.15	.09	.50	.54	*.49*	.12	.05	.07	.58	.54	1.00
(12) Age	76.1	6.2	−.06	−.03	−.42	−.43	−.37	−.19	.01	−.15	−.49	−.44	−.38

Note: Free recall is measured as proportion of correctly recalled items. Recognition is measured as d' from signal-detection theory. Reliability coefficients are printed bold and in italics. Correlations $r > .15$ are significant ($p < .01$).

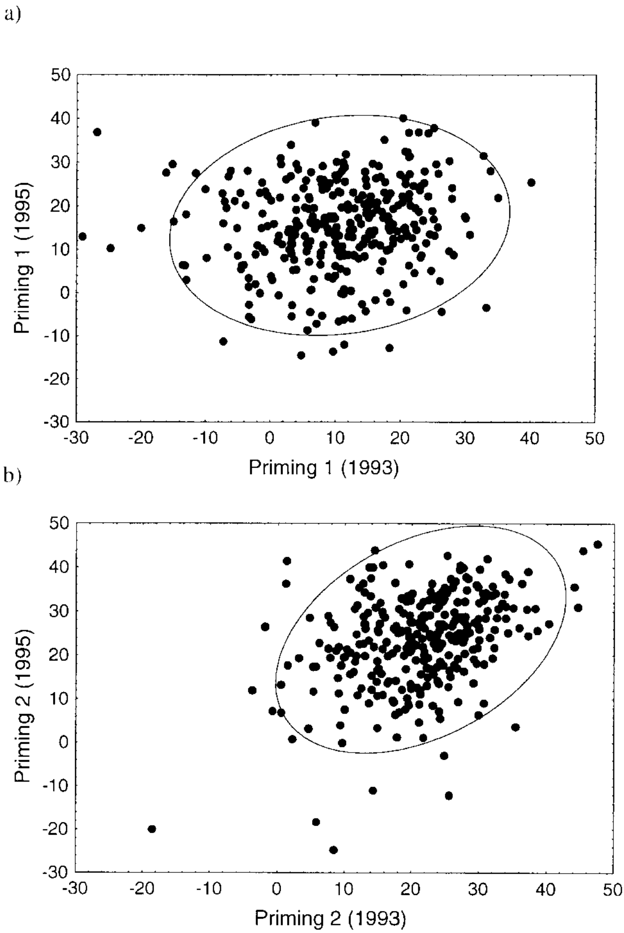

FIG. 2. Point plots of the retest reliabilities (a) for Priming 1 and (b) for Priming 2 in Experiment 1.

correlations of $r = .64$, $p < .01$, for Free Recall 1, $r = .63$, $p < .01$, for Free Recall 2, and $r = .49$, $p < .01$, for recognition.

Comparison of the Reliability Estimates. In all three clarification tasks we found substantial priming effects. Moreover, the well-established finding that explicit memory performance is highly correlated with age was replicated, whereas the correlation between age and implicit memory was much lower or absent. Therefore, the selected tasks can be considered as valid and representative operationalizations of implicit and explicit memory, and hence the preconditions for our primary goal, the analysis of the reliability of the tasks or the stability or individual differences in these measures, are met. The inter-task correlations of the implicit tasks were not significant, or were substantially lower than the corresponding correlations of the explicit tasks. Testing the highest correlation within indirect tests ($r = .22$) and the lowest correlation within direct tests ($r = .45$) with the procedure for the comparison of dependent correlations proposed by Raghunathan, Rosenthal, and Rubin (1996) revealed a highly significant difference, $Z = 3.78$, $p < .01$. This finding is astonishing, considering the very similar demands of the three priming tasks. Test–retest correlations showed the same pattern: Retest reliabilities of the implicit tasks, though significant, were much lower than those of the explicit tasks. Again, directed comparisons of the retest reliabilities between direct and indirect tests with the Raghunathan et al. procedure revealed that all the reliability scores of the priming tasks differed significantly from the explicit memory measures, with Z-values between $Z = 1.77$, $p < 0.05$, for the comparison of Priming 2 and recognition, and $Z = 7.95$, $p < 0.01$, for the comparison of Priming 1 and Free Recall 1.

If implicit memory testing is much less reliable than explicit memory examination, it might well be that the functional dissociation between explicit and implicit memory as a function of age could be due to differential reliability of the tests. The fact that we found an age effect on at least one priming measure in this study might corroborate this interpretation. It is possible that due to our large sample, effects can be shown that are otherwise hidden. This observation is in accordance with other studies using large samples (e.g. Hultsch et al., 1991) or meta-analysis (La Voie & Light, 1994).

Of course, there may be other reasons for the generally low correlations in the implicit memory measures, and the actual relationships between the variables may be far stronger then the coefficient values suggest. Considering the retest interval of 2 years it might be that we assessed low stability over a long period of time, whereas the reliability of the tasks is much higher. Therefore, in Experiment 2 we administered a parallel-test design to a sample of students. Two different test versions were applied within one session.

EXPERIMENT 2

Method

Participants

Forty undergraduate students participated in this experiment as a partial fulfilment of a course requirement. Their age ranged from 19 to 46 years with a mean age of 25.5 years.

Material

Test Version A was composed of the same material as the test in Experiment 1 except that the third perceptual clarification task was omitted. Version B was constructed in the same way as Version A, but different pictures from the Snodgrass and Vanderwart (1980) material was used. Additionally, the program was adapted for IBM-compatible computers. Again all relevant data were directly saved on the computer.

Procedure

The procedure was the same as that in Session 1 of Experiment 1, except that the subjects were administered the parallel version immediately after the initial test version. Half of the subjects worked first through Test Version A and then through Test Version B; the other half was tested with Version B first, before Version A was administered.

Analysis

Again, we used a correlation approach. The intercorrelations and parallel test reliabilities between the indirect tests and direct tests were analysed.

Results and Discussion

Means, standard deviations, inter-task correlations, and retest reliabilities of all measures are presented in Table 2. In the upper half of the matrix the first implicit and explicit tests are displayed and beneath them implicit and explicit parallel tests are shown.

Performance in Indirect Memory Tests. Mean naming times of the old pictures and the new pictures were 3.93 sec and 4.67 sec for Priming 1 and 3.51 sec and 5.32 sec for Priming 2 in the first test, and 3.97 sec and 4.95 sec for Priming 1 and 3.61 sec and 5.41 sec for Priming 2 in the second test. *T* tests of the reaction times of old and new pictures confirmed a strong priming effect in both conditions and in both test versions, $p < .01$, with t values between $t(39) = 7.43$ and $t(39) = 20.64$.

The relationships between the indirect tests were comparable with the results of Experiment 1. The intercorrelations of the different priming tasks within each measurement were generally low, and three out of four coefficients were not even statistically significant. The parallel test reliabilities between Test Versions A and B were also low, $r = .29, p > .05$, for Priming 1, and even negative for Priming 2, $r = -.08$, $p > .05$. Thus again, we found insufficient reliability for Priming 1 and no reliability at all for Priming 2.

Performance in Direct Memory Tests. In contrast to the indirect tasks, all explicit memory measures showed significant parallel test reliabilities between $r = .43$, $p < .01$, and $r = .65$, $p < .01$. These data replicate the findings of Experiment 1.

Comparison of the Reliability Estimates. In Experiment 2, the findings of Experiment 1 were replicated. Again we found stable priming effects across tests, whereas the priming score of each subject was not reliable. Therefore the length of the test–retest interval in Experiment 1 cannot be the cause of the low reliability in perceptual priming. A directed comparison of the parallel test reliabilities of implicit and explicit memory measures with the test Raghunathan et al. (1996) revealed that four of the six comparisons between implicit and explicit memory were significant (Priming 1 vs. d' and Priming 2 vs. Free Recall 1, Free Recall 2, and recognition) with Z-values between 2.05 and 3.69, whereas the comparisons of Priming 1 with Free Recall 1 and Free Recall 2 failed to reach significance, $Z = .70$, $p = .24$, and $Z = 1.04$, $p = .15$, respectively. To find significant differences for these two comparisons, sample sizes of 207 and 98 subjects, respectively, would have been necessary. Therefore, one could argue that our sample size was too small. However, the pattern of the data completely replicates the findings of Experiment 1.

TABLE 2

Means, standard deviations, inter-task correlations, and reliability coefficients for explicit and implicit measures in Experiment 2

	M	*SD*	*(1)*	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*
First measurement											
Implicit Tests											
(1) Priming 1	16.05	11.94	1.00								
(2) Priming 2	34.01	8.59	.12	1.00							
Explicit Tests											
(3) Free Recall 1	0.49	0.14	.12	−.05	1.00						
(4) Free Recall 2	0.53	0.16	.06	.19	.62	1.00					
(5) Recognition	2.80	0.65	.07	−.14	.63	.40	1.00				
Parallel test											
Implicit Tests											
(6) Priming 1	19.82	13.87	*.29*	−.15	.09	−.08	.31	1.00			
(7) Priming 2	33.24	10.50	.12	−*.08*	−.06	−.10	.15	.41	1.00		
Explicit Tests											
(8) Free Recall 1	0.50	0.13	.10	−.06	*.43*	.17	.44	.16	.10	1.00	
(9) Free Recall 2	0.57	0.17	.04	.14	.45	*.49*	.36	.16	.30	.57	1.00
(10) Recognition	2.86	0.71	.12	.07	.54	.39	*.65*	.42	.09	.56	.40

Note: Free recall is measured as proportion of correctly recalled items. Recognition is specified with discrimination score d' from signal-detection theory. Reliability coefficients are printed bold and in italics. Correlations $r > .31$ are significant ($p < .05$), correlations $r > .41$ are significant ($p < .01$).

EXPERIMENT 3

Experiment 3 was designed to investigate whether our findings from the perceptual clarification task could be generalized to a more typical measure of indirect memory, the word stem completion task, and another explicit memory measure, cued recall. Again we used a parallel test design with an interval of 2 weeks between the test occasions. Additionally, the effect of levels of processing manipulation was studied.

Method

Participants

Fifty undergraduate students participated as subjects in return for course credit. Their age ranged from 20 to 37 years with a mean age of 24.5 years.

Material

A total of 140 word stems from the standardization of Meier and Eckstein (1998) were used. The critical stimulus set consisted of 70 five-letter words in Session 1 and 70 six-letter words in Session 2. These words were selected in such a way that each of the word stems was unique and could be completed with at least three different nouns. The stimulus sets were divided into 10 sublists of seven words. The proportion of completion with a target word was equated so that each sublist had a baseline of $p = .12$.

For all subjects, four sublists were used for each of the two study conditions, and two lists were used in the new condition of the implicit test. Sublists were rotated through conditions so that each word appeared equally often in each condition (i.e. studied and non-studied, semantically and graphemically processed).

Word stems for the test list were created by cutting off the last two letters of each word in Session 1 and cutting of the last three letters of each word in Session 2. Study words and test cues were presented in the centre of the screen of a VGA-Monitor, which was controlled by a Pentium-PC. The program was developed with the Micro-Experimental Laboratory software package (Schneider, 1988). As a distractor task the short version of the Stroop Interference Test (Regard, 1981; Spreen & Strauss, 1991) was used.

Procedure

Subjects were tested individually with a retest interval of 2 weeks. On both occasions the experiment was conducted in three phases—study, distraction, and test. The procedure was identical for all subjects. They were informed that they had to perform several tasks, and that the difficulty of the tasks would increase towards the end of each session.

The study phase consisted of a semantic and a graphemic condition. For the semantic block, subjects judged the pleasantness of each word on a 5-point scale. For the graphemic block, subjects counted the number of enclosed areas within each word (e.g. four in the word "paper"). The order of these tasks was fixed, and each subject performed the judgement task before the counting task. Presentation order of words within each block was random for each subject. In both study phases each trial consisted of a 2 sec presentation of the word, then the screen was cleared and the subject had 3 sec to type in the answer (a digit) before the next word was presented. After the study phase the short version of the Stroop test was administered as a distractor task.

The test phase of the experiment consisted of the presentation of 70 word stems. In the implicit test 42 word stems were presented as stimuli, out of which 14 corresponded to words presented in the semantic study phase, 14 corresponded to the words shown in the graphemic study phase, and 14 had no relation to words presented earlier. In the explicit test, 28 word stems were presented: Of these, 14 corresponded to words presented in the semantic study phase, and 14 corresponded to words of the graphemic study phase. In the implicit test, the instruction was to complete the word stem with the first five-letter noun that came to mind in Session 1, and the first six-letter noun in Session 2. Subjects had to type the word on the keyboard and to confirm their response with the "return" key. After the subjects had entered a word, or if no answer was given within 10 sec, the stimulus disappeared and the next word stem was presented. In the explicit test condition, the instruction was to use the word stem as a cue to retrieve a word that had been presented earlier in the experiment either in the semantic or the graphemic study phase. Again subjects had to type the word into the computer and to confirm their answer with the "return" key. They were instructed not to guess and only to type in a word when they actually remembered that it had already been presented. After the subjects had entered a word, or if no answer was given within 10 sec, the stimulus disappeared and the next word stem was presented.

Analysis and Design

For each session, we first analysed the indirect test condition with a one-way analysis of variance (ANOVA) separately to look for priming effects. Then a 2 × 2 ANOVA design was used with study condition and test instruction to establish the dissociation between explicit and implicit memory as a function of levels of processing. Both experimental manipulations, study instruction (semantic vs. graphemic) and test instruction (direct vs. indirect) were varied within subjects. Subsequently, the retest reliabilities for the different tasks and conditions were analysed.

Results and Discussion

Table 3 shows the mean proportions of stems completed with critical words under each experimental condition for both sessions, the priming effects, the inter-task correlations and the reliability coefficients. Responses were classified as target items only if the target word was typed correctly.

First, the proportions of stems completed with target words in the indirect test conditions were analysed. In Session 1 these proportions were $p = .33$ for the semantic study condition, $p = .27$ for the graphemic study condition, and $p = .12$ for the unstudied condition. In Session 2 proportions were $p = .30$, $p = .24$, and $p = .12$, for the semantic, graphemic, and unstudied conditions, respectively. Separate one-way ANOVAs for both test occasions revealed highly significant differences between these study conditions, $F(2, 98) = 38.95$ for Session 1, and $F(2, 98) = 27.21$, for Session 2. According to a Scheffé test the sources of these effects were located for both sessions as a difference between the two old study conditions (graphemic and semantic) and the unstudied condition, $p < .01$, whereas the difference between graphemic and semantic study conditions did not reach significance, $p > .05$. Thus, we have found again consistent priming effects for both study conditions (graphemic and semantic) at both test occasions.

A further 2 (semantic vs. graphemic study condition) × 2 (indirect vs. direct test instruction) ANOVA for repeated measures revealed a significant main effect of study

TABLE 3
Mean proportions and standard deviations of items correct, inter-task correlations and reliability coefficients in Experiment 3

	M	*SD*	*(1)*	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*
Session 1											
Implicit Tests											
(1) Stem Completion (semantic)	0.33	0.15	1.00								
(2) Stem Completion (graphemic)	0.27	0.13	−.02	1.00							
(3) Stem Completion (baseline)	0.12	0.09	−.07	.32	1.00						
Explicit Tests											
(4) Cued Recall (semantic)	0.39	0.16	.27	.23	.32	1.00					
(5) Cued Recall (graphemic)	0.15	0.13	.19	.21	.35	.45	1.00				
Session 2											
Implicit Tests											
(6) Stem Completion (semantic)	0.30	0.15	*.13*	.20	.15	.09	.19	1.00			
(7) Stem Completion (graphemic)	0.24	0.14	.08	*.25*	.05	.15	.24	.16	1.00		
(8) Stem Completion (baseline)	0.12	0.09	.16	.01	*.20*	.19	.09		−.07	.02	1.00
Explicit Tests											
(9) Cued Recall (semantic)	0.34	0.18	.06	.31	.30	*.48*	.49	.18	.16	−.11	1.00
(10) Cued Recall (graphemic)	0.13	0.12	.22	.02	.03	.46	*.52*	.02	.02	.21	.30

Note: Reliability coefficients are printed bold and in italics. Correlations $r > .28$ are significant ($p < .05$), correlations $r > .39$ are significant ($p < .01$).

condition, $F(1, 49) = 63.95$, $p < .01$ in Session 1 and $F(1, 49) = 53.25$, $p < .01$ in Session 2, whereas the main effect of test instruction just missed significance, $F(1, 49) = 3.81$, $p > .05$ in Session 1 and $F(1, 49) = 2.54$, $p > .1$ in Session 2. More important, we found significant Study Condition × Test Instruction interactions in both sessions, $F(1, 49) = 28.56$, $p < .01$, in Session 1 and $F(1, 49) = 18.21$, $p < .01$, in Session 2. Thus, the manipulation of levels of processing did not affect the word stem completion, whereas semantic encoding led to better cued recall performance than did graphemic encoding. Therefore the dissociation between direct test and indirect test as a function of a manipulation of levels of processing at encoding was replicated, and the requirements were met for the analysis of reliability.

Retest reliability for both implicit test conditions was rather low, with $r = .13$, $p > .05$, for the semantic study condition and $r = .25$, $p > .05$, for the graphemic study condition. On the other hand, the retest reliabilities for the explicit measures were $r = .48$, $p < .05$, for the semantic condition and $r = .52$, $p < .05$, for the graphemic condition. We did not calculate difference scores for the implicit measure, because the baseline rate is only useful to compare group means. However a substraction of the group mean from the individual test score is a linear transformation, which does not lead to different correlation coefficients.

Comparison of the Reliability Estimates. In Experiment 3, we replicated the dissociation between implicit and explicit memory as a function of levels of processing. Consistent with our findings from the previous experiments, the reliability of explicit memory measures was found to be higher than the reliability of implicit measures. A directed comparison of the parallel test reliabilities of implicit and explicit memory measures with the test of Raghunathan et al. (1996) revealed that the reliability of implicit memory after a semantic study phase was significantly different from both explicit memory conditions, with Z-values of 1.95 and 2.16, $p < .05$, for the semantic and the graphemic explicit test, whereas the reliability of implicit memory after a graphemic study phase failed to reach significance, with explicit memory performance after the semantic study condition, $Z = 1.29$, $p = .10$, and with explicit memory performance after the graphemic study condition, $Z = 1.57$, $p = .06$. To find significant differences for these two comparisons, sample sizes of 80 and 57 subjects respectively, would have been necessary.

Thus, as in Experiment 1 and 2, a precondition for an adequate interpretation of the functional dissociation between implicit and explicit was not fulfilled, because implicit and explicit memory measures had different reliabilities.

GENERAL DISCUSSION

The results of our experiments revealed that reliability of implicit memory measures like picture clarification and word stem completion is much lower than reliability of explicit memory measures like free recall, recognition, and cued recall. Additionally, the intercorrelations within the domain of explicit memory are higher than those within the domain of implicit memory. The tasks, and the experimental and the quasi-experimental settings that we used in this study can be considered as representative for investigations within the field of implicit memory research. In each condition of our experiments we found significant priming in the indirect memory tests. Moreover, we replicated two functional dissociations: Age and levels of processing had effects on explicit tests, but little or no effect on implicit tests. Despite this, the implicit memory tests always showed low retest reliability as well as low parallel test reliability. We believe that this finding is important, because it demands a new perspective in the evaluation of experimental and individual difference research in the domain of implicit memory.

If implicit memory testing is much less reliable than explicit memory examination with comparable sizes of item materials, one needs to be aware that some of the reported functional dissociations between explicit and implicit memory performance in experimental and quasi-experimental research might be caused by the different reliabilities of the instruments. These dissociations usually show strong effects of factors like attention, levels of processing, or age on explicit measures, but not on implicit measures. However, lower reliability leads to a higher proportion of the error variance. A test with low reliability must show a greater effect to achieve significance. Therefore, small effects of implicit memory probably can not be shown due to the low reliability of the measurement instruments. These considerations can be extended to the issue of stochastic independence. Unlike functional independence, stochastic independence is not based on comparing the average performance of two tests, but rather on determining whether performance on a particular item in one test predicts performance on the same item in another test. If the reliability of one of these measures is low, we cannot expect a systematic relationship between the test items. As suggested by Ostergaard (1992) only a small proportion of the variance of implicit memory measures may be due to memory processes. Other factors like cue characteristics, response biases, and pre-experimental familiarity with the test items can also contribute to the observed variance. Therefore, estimates of stochastic independence between implicit and explicit memory tests may have little relation because only a small proportion of the variance in implicit measures is related to memory. Considering the low reliability of indirect memory measures, at least the perceptual ones analysed in this study, one would also expect stochastic independence between different indirect memory tests. This is exactly what has been found (Cabeza & Ohta, 1993; Witherspoon & Moscovitch, 1989). Additionally, in the study of Witherspoon and Moscovitch, stochastic dependency was not very high even when the same persons were tested with the same test and the same item material within one test session ($\phi = .37$ for perceptual identification and $\phi = .54$ for word fragment completion). In a study by Hayman and Tulving (1989) successive tests of fragment completion with different fragments were also not highly associated.

In experimental memory research the focus usually lies on the comparison of direct and indirect tests as a function of experimental variables. However, an increasing number of studies focus on developmental or individual difference comparisons of several groups. As a result, psychometric criteria—in particular, reliability of the tasks—become more important. Furthermore, for individual difference and neuropsychological research our findings show that priming measures need to be used very cautiously as tools for measuring individual differences or even as diagnostic tools. As our data suggest, much work remains to be done in order to clarify the question of whether priming tasks can be used as measures of an individual difference variable (in diagnostic research), and whether they are mediated by certain brain structures (in neuropsychological research). Given the low reliability of implicit memory measures and the low intercorrelations between different implicit memory meaures, the practice of associating performance in these tasks with lesions in certain brain areas seems doubtful, as it is doubtful to infer different processes from functional dissociations in general (Dunn & Kirsner, 1988; Hintzman, 1990).

Given our findings, one has to ask what reasons might cause them, and what could be done to enhance reliability of implicit memory tests. To answer these questions much more research will have to be done. Nevertheless, some general considerations can be made here.

We can imagine several reasons for the low reliability of implicit memory measures. First, task instructions are less specific in many implicit measures than in explicit measures. This is true especially for a task like word stem completion where the subjects have to complete a word stem "with the first word that comes to mind". This task instruction offers more degrees of freedom for finding an answer that is correct according to the instructions. However, not every "correct" answer is a correct answer in terms of an indirect memory test because only previously presented words are coded as correct. In contrast, the explicit instruction to "complete with a word that has been presented earlier" is much clearer with respect to a correct answer. Basically, there is a much larger search domain in implicit tests than in explicit tests. Therefore, the vaguer instruction could be a reason for the lower reliability of implicit memory measures. Second, most implicit measures are difference scores. There has been a long controversy about the usefulness of difference scores in scientific research because they tend to be unreliable (see Williams & Zimmermann, 1996, for an overview). However, explicit memory measures like recognition (which is also a difference score) demonstrate that difference scores can be reliable. Additionally, by means of a thorough standardization of the test material, which leads to a prior knowledge of the baseline value (e.g. the probability of completing a new word stem with a target word) this problem can be avoided. In this case, performance in the indirect test condition can be used as an implicit measure without subtracting the baseline performance. However, as can be seen in Experiment 3, where test material was composed in such a way, even the reliabilities of the implicit word stem completion tests were low. Therefore the possibility that difference scores are the main reason for the lower reliability of perceptual priming can be rejected. Further, it has been argued that, in contrast to explicit memory measures, priming is a measure of automatic processes. Automatic processes are assumed to be fast, to require no awareness, and to reflect a general phenomenon with little individual variance (Hasher & Zacks, 1979). Because of the small inter-individual variation of automatic processes, high reliability would not be expected (Reber, Walkenfeld, & Hernstadt, 1991). Our data suggest that low reliability is not caused by a lack of variation in implicit measures as can be seen in Experiment 3, where the variation of explicit measures can be compared directly. However, given the relationship between reliability and variance there must be more error variance in the observed variance of the implicit measures.

Additionally, it could well be that priming fluctuates in a way that makes it unlikely to be measured as a stable feature. Finally, possibly some people show no priming, for reasons that are not related to the reliability of a test. Therefore, we performed additional analysis of our data in which only those subjects were included who showed consistent priming effects in both test conditions. However, over all experiments, this restriction had little, no consistent, and in fact no increasing effect on the reliability of the measures.

How then can the reliability of the implicit tests be enhanced? From a theoretical point of view there are two aspects that directly influence reliability. First, reliability increases with test length, and second, reliability increases with the reliability of each item (but see Li, Rosenthal, & Rubin, 1996 for exceptions). Indeed, reliability can be enhanced when more items are used. However, enhancing the number of items increases reliability of explicit and implicit memory tests. Additionally, with certain populations this criterion cannot be easily fulfilled because of the demanding aspects of long tasks. Especially with children, elderly participants, or patients, a prolongation of the test procedure may be problematic or even impossible. Also, the possibility of increasing the reliability of each test item cannot easily be fulfilled. Usually priming is not investigated at item level, but the performance on several items is accumulated before the difference score is calculated. Therefore, dedicating more time and effort to the standardization of item material may increase reliability.

We conclude that when the focus of research lies on the comparison of several manipulations on different tests, as in implicit memory research, comparable reliabilities for implicit and explicit tests should be established first. Only by these means can we be sure that dissociations are not an artefact due to differential reliability of tests. Therefore, we believe that the findings of this study justify the claim that future studies comparing different tests of memory, in particular tests of explicit memory and implicit memory, should be concerned with reliability.

REFERENCES

Bazin, N., & Perruchet, P. (1996). Implicit and explicit associative memory in patients with schizophrenia. *Schizophrenia Research*, *22*, 241–248.

Berry, D.C., Banbury, S., & Henry, L. (1997). Transfer across form and modality in implicit and explicit memory. *Quarterly Journal of Experimental Psychology*, *50A*, 1–24.

Blaxton, T.A. (1989). Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 657–668.

Brown, A.S., & Mitchell, D.B. (1994). A reevaluation of semantic versus nonsemantic processing in implicit memory. *Memory & Cognition*, *22*, 533–541.

Butters, N., Heindel, W.C., & Salmon, D.P. (1990). Dissociation of implicit memory in dementia: Neurological implications. *Bulletin of the Psychonomic Society*, *28*, 359–366.

Cabeza, R., & Ohta, N. (1993). Dissociating conceptual priming, perceptual priming, and explicit memory. *European Journal of Cognitive Psychology*, *5*, 35–53.

Carroll, M., Byrne, B., & Kirsner, K. (1985). Autobiographical memory and perceptual learning: A developmental study using picture recognition. *Memory & Cognition*, *13*, 273–279.

Challis, B.H., & Brodbeck, D.R. (1992). Level of processing affects priming in word-fragment completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 595–668.

Denny, E.R., & Hunt, R.R. (1992). Affective valence and memory in depression: Dissociation of recall and fragment completion. *Journal of Abnormal Psychology*, *101*, 575–580.

Dunn, J.C., & Kirsner, K. (1988). Discovering functionally independent mental processes: The principle of reversed association. *Psychological Review*, *95*, 91–101.

Eich, J.E. (1984). Memory for unattended events: remembering with and without awareness. *Memory & Cognition*, *12*, 105–111.

Gabrieli, J.D.E., Keane, M.M., Stanger, B.Z., Kjelgaard, M.M., Corkin, S., & Growdon, J.H. (1994). Dissociations among structural-perceptual, lexical-semantic, and event-fact memory systems in Alzheimer, amnesic, and normal subjects. *Cortex*, *30*, 75–103.

Geiselman, R.E., Woodward, J.A., & Beatty, J. (1982). Individual differences in verbal memory performance: A test of alternative information-processing models. *Journal of Experimental Psychology: General*, *111*, 109–134.

Gollin, E.S. (1960). Developmental studies of visual recognition of incomplete objects. *Perceptual and Motor Skills*, *11*, 289–298.

Graf, P., & Masson, M.E.J. (1993). *Implicit memory: New directions in cognition, development, and neuropsychology*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Grafman, J., Weingartner, H., Newhouse, P.A., Thompson, K., Lalonde, F., Litvan, I., Molchan, S., & Sunderland, T. (1990). Implicit learning in patients with Alzheimer's disease. *Pharmacopsychiatry*, *23*, 94–101.

Hallahan, M., & Rosenthal, R. (1996). Statistical power: Concepts, procedures, and applications. *Behavior Research and Therapy*, *34*, 489–499.

Hasher, L., & Zacks, R.T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General*, *108*, *3*, 356–388.

Hashtroudi, S., Parker, E.S., deLisi, L.E., Wyatt, R.J., & Mutter, S.A. (1984). Intact retention in acute alcohol amnesia. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 156–163.

Hayman, C.A.G., & Tulving, E. (1989). Is priming in fragment completion based on a ''traceless'' memory system? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 941–956.

Heindel, W.C., Salmon, D.P., & Butters, N. (1990). Pictorial priming and cued recall in Alzheimer's and Huntington's disease. *Brain and Cognition*, *13*, 282–295.

Hertel, P.T., & Hardin, T.S. (1990). Remembering with and without awareness in a depressed mood: Evidence of deficits in initiative. *Journal of Experimental Psychology: General*, *119*, 45–59.

Hintzman, D.L. (1990). Human learning and memory: Connections and dissociations. *Annual Review of Psychology*, *41*, 109–139.

Hultsch, D.F., Masson, M.E.J., & Small, B.J. (1991). Adult age differences in direct and indirect tests of memory. *Journal of Gerontology: Psychological Sciences*, *46*, 22–30.

Jacoby, L.L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513–541.

Jacoby, L.L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, *110*, 306–340.

Kolers, P.A. (1976). Reading a year later. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 554–565.

La Voie, D., & Light, L.L. (1994). Adult age differences in repetition priming: A meta-analysis, *Psychology & Aging*, *9*, 539–553.

Lewandowsky, S., Dunn, J.C., & Kirsner, K. (1989). *Implicit memory. Theoretical issues*. Hillsdale, NJ: Lawrence Erlbaum, Associates, Inc.

Li, H., Rosenthal, R., & Rubin, D.B. (1996). Reliability in psychology: From Spearman-Brown to maximal reliability. *Psychological Methods*, *1*, 98–107.

Light, L.L., & Singh, A. (1987). Implicit and explicit memory in young and old adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 531–541.

Lorsbach, T.C., & Worman, L.J. (1989). The development of explicit and implicit forms of memory in learning disabled children. *Contemporary Educational Psychology*, *14*, 67–76.

Lorsbach, T.C., & Worman, L.J. (1990). Episodic priming in children with learning disabilities. *Contemporary Educational Psychology*, *15*, 93–102.

McDonald-Miszczak, L., Hertzog, C., & Hultsch, D.F. (1995). Stability and accuracy of metamemory in adulthood and aging: A longitudinal analysis. *Psychology & Aging*, *10*, 553–564.

Meier, B., & Eckstein, D.J. (1998). German material for studying implicit and explicit memory with the word stem completion task. *Sprache & Kognition*, *17*, 89–105.

Mitchell, D.B. (1989). How many memory systems? Evidence from aging. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 31–49.

Moscovitch, M., Winocur, G., & McLachlan, D. (1986). Memory as assessed by recognition and reading time in normal and memory-impaired people with Alzheimer's disease and other neurological disorders. *Journal of Experimental Psychology: General*, *115*, 331–347.

Ostergaard, A.L. (1992). A method for judging measures of stochastic dependence: Further comments on the current controvesy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 413–420.

Parkin, A.J., & Russo, R. (1990). Implicit and explicit memory and the automatic/effortful distinction. *European Journal of Cognitive Psychology*, *2*, 71–80.

Parkin, A.J., & Streete, S. (1988). Implicit and explicit memory in children and adults. *British Journal of Psychology*, *79*, 361–369.

Perrig, W.J., Hofer, D., Kling, V., Meier, B., Perrig, P., Ruch, M., & Serafin, D. (1994). *Computerunterstüzter Gedächtnis-Funktions-Test. Manual*. Basel: Forschungsberichte des Instituts für Psychologie.

Perrig, W.J., & Perrig, P. (1993). Implizites Gedächtnis: Unwillkürlich, entwicklungsresistent und altersunabhängig? *Zeitschrift für Entwicklungspsychologie and Pädagogische Psychologie*, *15*, 29–47.

Perrig, P., & Perrig, W.J. (1995). Implicit and explicit memory in mentally retarded, learning disabled, and normal children. *Swiss Journal of Psychology*, *54*, 77–86.

Perrig, W.J., Wippich, W., & Perrig-Chiello, P. (1993). *Unbewusste Informationsverarbeitung*. Bern: Huber.

Perrig-Chiello, P., Perrig, W.J., Stähelin, H.B., Krebs E., & Ehrsam R. (1996). Autonomie, Wohlbefinden und Gesundheit im Alter. Das Basler IDA-Projekt. *Zeitschrift für Gerontologie und Geriatrie*, *29*, 95–109.

Perruchet, P., & Baveux, P. (1989). Correlational analysis of explicit and implicit memory performance. *Memory & Cognition*, *17*, 77–86.

Raghunathan, T.E., Rosenthal, R., & Rubin, D.B. (1996). Comparing correlated but nonoverlapping correlations. *Psychological Methods*, *1*, 178–183.

Reber, A.S., Walkenfeld, F.F., & Hernstadt, R. (1991). Implicit and explicit learning: Individual differences and IQ. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 888–896.

Regard, M. (1981). *Cognitive rigidity and flexibility: A neuropsychological study*. Unpublished PhD dissertation, University of Victoria, British Columbia.

Richardson-Klavehn, A., & Bjork, R.A. (1988). Measures of memory. *Annual Review of Psychology*, *39*, 475–543.

Roediger, H.L., & Blaxton, T.A. (1987). Effects of varying modality, surface, features, and retention interval on priming in word-fragment completion. *Memory & Cognition*, *15*, 379–388.

Roediger, H.L., & McDermott. (1993). Implicit memory in normal human subjects. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology*: *Vol. 8* (pp. 63–131). Amsterdam: Elsevier.

Roediger, H.L., Weldon, M.S., Stadler, M.A., & Riegler, G.H. (1992). Direct comparison of word stems and word fragments in implicit and explicit retention tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 1251–1269.

Salthouse, T.A. (1992a). Influence of processing speed on adult age differences in working memory. *Acta Psychologica*, *79*, 155–170.

Salthouse, T.A. (1992b). Why do adult age differences increase with task complexity? *Developmental Psychology*, *28*, 905–918.

Salthouse, T.A., & Babcock, R.L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology*, *27*, 763–776.

Salthouse, T.A., & Meinz, E.J. (1995). Aging, inhibition, working memory, and speed. *Journal of Gerontology: Psychological Sciences*, *50b*, 297–306.

Schacter, D.L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *13*, 501–518.

Schneider, W. (1988). Micro experimental laboratory: An integrated system for IBM PC compatibles. *Behavior Research Methods, Instruments and Computers*, *20*, 206–217.

Schwartz, B.L., Rosse, R.B., & Deutsch, S.I. (1993). Limits of the processing view in accounting for dissociations among memory measures in a clinical population. *Memory & Cognition*, *21*, 63–72.

Snodgrass, J.G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for naming agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174–215.

Spreen, O., & Strauss, E. (1991). *A compendium of neuropsychological tests*. NY: Oxford University Press.

Squire, L.B., Knowlton, B., & Musen, G. (1993). The structure and organization of memory. *Annual Review of Psychology*, *44*, 453–495.

Tulving, E., & Schacter, D.L. (1990). Priming and human memory systems. *Science*, *247*, 301–306.

Tulving, E., Schacter, D.L., & Stark, H.A. (1982). Priming effects in word fragment completion are independent of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*, 336–342.

Weldon, M.S., & Roediger, H.L. (1987). Altering retrieval demands reverses the picture superiority effect. *Memory & Cognition*, *15*, 269–280.

Widmer, L.K., Stähelin, H.B., Nissen, C., & da Silva, A. (1981). *Venen-Arterien-Krankheiten, koronare Herzkrankheiten bei Berufstätigen*. Bern: Hans Huber.

Williams, R.H., & Zimmermann, D.W. (1996). Are simple gain scores obsolete? *Applied Psychological Measurement*, *20*, 59–69.

Witherspoon, D., & Moscovitch, M. (1989). Stochastic independence between two implicit memory tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 22–30.

Original manuscript received 8 May 1998
Accepted revision received 30 November 1998