A Framework for Analyzing and Interpreting Differential Aging Patterns: Application to Three Measures of Implicit Learning*

Timothy A. Salthouse, Katheryn E. McGuthry, and David Z. Hambrick
School of Psychology, Georgia Institute of Technology

ABSTRACT

At least four distinct explanations can be proposed to account for patterns of spared and impaired performance in which the age-related effects on some variables are weaker than those on other variables. These are: (a) some variables lack sufficient reliability to exhibit relations with other variables; (b) the variables differ in their dependence on what many of the variables have in common; (c) variables with little or no age differences reflect a qualitatively different form of cognition than variables with moderate to large age differences; and (d) variables with little or no age differences have independent positive age-related influences that offset the negative age-related effects shared with other cognitive variables. These interpretations were examined with three different variables hypothesized to reflect implicit learning obtained from a sample of 183 adults ranging from 18 to 87 years of age. Only an implicit learning measure derived from a sequential reaction time task had acceptable reliability at the level of individual participants, and it was negatively related to age and positively related to variables reflecting fluid cognition. These results therefore suggest that typical measures of implicit learning, when they can be reliably assessed, do not reflect a qualitatively distinct type of cognitive processing nor do they seem to exhibit additional compensatory age-related influences.

Some variables are interesting from an age-comparative perspective because although they are similar in certain respects to other variables, they do not exhibit the same type of age-related differences. However, two important issues need to be considered when examining and interpreting the relations of age on different variables. One concerns the reliability of the variables because unless reliability is moderately high the variable cannot be expected to exhibit systematic relations with other variables. Unfortunately, the term reliability has been used in a number of different ways in cognitive psychology, and this has led to some confusion with respect to its role in research. *Webster’s Ninth New Collegiate Dictionary* (1991) defines reliability as “the extent to which an experiment, test, or measuring procedure yields the same results on repeated trials.” This definition implies a sense of consistency, but in various usages the consistency has been interpreted as referring to: (a) the probability that the results are not attributable to chance, as when reliable is used as a synonym for statistically significant; (b) the robustness of a phenomenon, related to the ease with which it has been replicated; and (c) the degree to which research participants will maintain a similar rank ordering of scores on repeated assessments. Although each of these usages of the term can be justified, it is impor-

---

*This research was supported by NIA Grant R37 AG06826 to Timothy A. Salthouse. We would like to acknowledge the valuable assistance in testing participants and scoring of data of Nate Fristoe, Holly Hancock, Kellie Hocking, Stephanie Sherwood, and Michelle Wolbretti. Address correspondence to: Timothy A. Salthouse, School of Psychology, Georgia Institute of Technology, Atlanta, GA 30332-0170, USA. Accepted for publication: March 23, 1999.
tant to distinguish among them because they are not equivalent and their implications can be quite different in certain situations.

For example, an empirical result or phenomenon could be significantly different from chance, and robust in the sense that it has been frequently replicated, and yet measures of the phenomenon may not be consistent at the level of individuals. A lack of consistency in terms of this latter, psychometric meaning of reliability could occur because of a variety of factors such as an insufficient number of observations for each individual, or lack of variation across people in the magnitude of the effect. Regardless of the reasons, however, even if a phenomenon has been well established in experimental studies, its assessment may not be meaningful in individual differences studies (including those comparing people of different ages) if the measures of the phenomenon are not reliable at the level of individuals. In fact, instances in which a robust phenomenon has been found to be unreliable at the level of individuals have recently been reported with semantic priming (Madden, Pierce, & Allen, 1993) and with negative priming (Park et al., 1996). This apparent paradox is understandable by considering that statistical significance of within-subjects condition effects (e.g., priming) are typically evaluated relative to the subjects by condition error term, and thus the likelihood of detecting a significant condition effect is greater when there is relatively little across-subjects variability in the magnitude of the condition effect. However, low variability tends to reduce the magnitude of between-person correlations that form the basis of estimates of reliability. There is consequently a sense in which the more robust a phenomenon, the lower the variation in the magnitude of the effect across people, and the lower the resulting estimates of reliability.

A second issue that needs to be considered when examining the influence of age on a variable is the relation of the age-related effects on that variable to the age-related effects evident on other variables. That is, because a large number of variables have been found to be significantly related to age in adulthood, it is important to determine the extent to which the effects on the target variable are distinct from, or independent of, the age-related effects on other variables. Several recent studies (e.g., Salthouse, Fristoe, & Rhee, 1996; Salthouse, Hambrick, & McGuthry, 1998; Salthouse, Hancock, Meinz, & Hambrick, 1996; Salthouse, Toth, Hancock, & Woodard, 1997; Verhaeghen & Salthouse, 1997) have revealed that many variables share surprisingly large proportions of age-related variance with one another. It is therefore possible that even though a variable was reliable, and was found to exhibit much smaller relations to age than several other variables, it might still share moderate to large amounts of variance with other variables.

One analytical method that can be used to investigate the magnitude of shared age-related influences involves determining what all variables have in common, and then controlling an estimate of this common factor when examining the influence of age on individual variables. Although this shared influences or single common factor analytical procedure has been primarily applied to sets of variables that were all related to age, it can also be useful when some of the variables are expected to be unrelated to age because it may be informative about the particular manner in which the absence of an age relation occurs. Figure 1 illustrates three possible ways in which patterns of differential age relations could occur within a shared influences model. Note that four variables are represented, but that the age-related effects are much smaller on variable 4 (V4), than on variables 1 (V1), 2 (V2), or 3 (V3).

The pattern illustrated in panel (A) represents the possibility that the critical variable, V4, has a weaker relation to age than the other variables because it is less strongly related to what the variables all have in common. This outcome would imply that the variable is similar in certain respects to other variables, but that for some reason it is less dependent than the other variables upon whatever is shared among variables. The pattern illustrated in panel (B) represents the possibility that V4 is not related to age because it has little or nothing in common with the other variables, and consequently it does not share variance with other variables. If the vari-
able has been established to have sufficient reliability, this type of outcome would be consistent with an interpretation that the variable involves a qualitatively different form of processing than that involved with the other variables. However, a third possibility for the lack of an age relation is that the variable is not related to age because even though it shares variance with other variables, and presumably also shares many of the same age-related influences, it may also have an independent age-related influence in the opposite direction of that common to the other variables. This possibility is represented in panel (C) of Figure 1.

Previous research with the shared influences analytical procedure has revealed evidence for each of the outcomes portrayed in Figure 1. For example, the pattern in panel (A) was reported in a study by Salthouse et al. (1997) with a variable representing accuracy in the Judgment of Line Orientation Test. The median absolute correlation with age across all variables in this analysis was .46, whereas the absolute correlation for the Judgment of Line Orientation variable was only .13. Furthermore, most of the variables in that analysis had standardized relations with the factor representing what they all had in common of .5 or greater (median .65), but the standardized coefficient for the Judgment of Line Orientation variable was only .27.

The Salthouse et al. (1997) study also had an example of a pattern such as that represented in panel (B) of Figure 1. In this case the critical variable was an estimate of automatic processing in a stem completion memory task derived from a process-dissociation procedure. As would be expected from the assumption that measures of automatic processing reflect a qualitatively different type of processing than variables based on controlled processing, the automatic processing variable in the stem completion task was found to be reliable but did not have a significant relation to the common factor representing what the other variables had in common. Moreover, its absolute correlation with age was .07, compared to the overall median of .46.

Finally, a pattern similar to that portrayed in panel (C) of Figure 1 has been found with verbal fluency variables in studies by Salthouse, Fristoe, and Rhee (1996), and Salthouse et al. (1997). In both cases the age correlations for the fluency variables were relatively small, and a direct positive relation from age was evident in the shared influences analysis. Further analyses revealed that the direct relation from age to the
fluency variable could be interpreted as a positive influence of knowledge because the relation was found to be mediated by a measure of vocabulary or word knowledge. This pattern is consistent with the view that the age-related effects on the verbal fluency variable were small because increased knowledge served to compensate for the processes (e.g., speed of retrieval) that tended to decline with increasing age.

The procedures described above were applied in the current study to help understand why age-related effects are frequently small in measures of implicit learning. Implicit learning is inferred when improvements in performance occur without deliberate attempts to improve, and when there is minimal conscious awareness of the basis for the improvement (Seger, 1994). Or in the words of Reber, Walkenfeld, and Hernstadt (1991), “implicit learning is the process whereby a complex rule-governed knowledge base is acquired largely independently of awareness of both the process and the product of the acquisition” (p. 888).

Three different tasks were used to investigate implicit learning. In each task the measure of implicit learning was operationalized as the level of performance evident without any attempt by the experimenter to induce conscious effort to acquire the materials, or deliberate recollection of prior experiences at the time of testing. The three tasks used in the current study to assess implicit learning are described next.

The sequence learning task was originally introduced by Nissen and colleagues (e.g., Nissen & Bullemer, 1987), and has been used in numerous studies, including age-comparative studies by Cherry and Stadler (1995), Curran (1997), Frensch and Miner (1994, Expt. 3), and Howard and Howard (1989, 1992). The version of the task used in this project is a serial reaction time task with ten repetitions of a 10-element pattern in each of four blocks of trials. Reduction in reaction time as a function of repetition of the pattern, and slower responses after transfer to a random sequence, has been considered evidence of implicit learning in this task. Little or no age differences have been found in these implicit measures of learning (although see Curran, 1997), but older adults often perform at lower levels than young adults in an intentional measure of learning derived from a sequence generation task (e.g., Cherry & Stadler, 1995; Frensch & Miner, 1994, Expt. 3; Howard & Howard, 1989, 1992).

The artificial grammar learning task is based on tasks developed by Reber and colleagues (e.g., Reber et al., 1991). The version used in this project is very similar to that recently described by McGeorge et al. (1997). The initial requirement for the participants is to try to remember 20 letter strings that are presented three times each. Next 40 new letter strings are to be classified, without feedback, as grammatical (well-formed) or nongrammatical (not well-formed) according to the rules used to create the original letter strings. Accuracy in the classification test can be considered a measure of implicit learning of the rules governing the arrangement of letter strings studied during the memorization stage. That is, although the participants attempted to remember the letter strings, they may have been implicitly learning something about the rules used to generate those strings, and that is what is presumably tested with the novel stimuli in the classification phase. McGeorge et al. (1997) reported that there were no age differences in the implicit measure (i.e., judgments of well-formedness), but moderate to large age differences favoring younger adults were found in a measure of explicit learning derived from a separate task.

The associative learning task involved speeded decisions about pairs of unfamiliar symbols. A code table containing eight pairs of symbols is presented at the top of the computer screen, and a pair of probe items is presented in the middle of the screen. The task is to decide, as rapidly and accurately as possible, whether the probe symbol pairs match according to the code table. The arrangement of pairs in the code table varies from trial to trial, but four symbol pairs are always consistently mapped (CM) with one another, and four symbol pairs are variably mapped (VM) to one another. The difference in reaction time between CM and VM pairs after practice can serve as an index of implicit learning of the pairings.
Each of these tasks involves learning relationships among stimuli—either sequences of spatial positions, sequences of letters, or pairs of symbols. The schematic structure within each task is illustrated in Figure 2. The pattern of sequence stimuli (labeled A through D for the four positions from left to right) is represented in the top panel, and the rule system used for the generation of letter strings in the artificial grammar task is illustrated in the middle panel. The bottom panel portrays stimuli in the associative learning task. The symbol pairs in the bottom left panel were CM, and those in the bottom right were VM.

Two assumptions of the current approach should be mentioned. The first is that the discrepancy between the median time for repeated and random patterns (in the sequence learning task), between the median time for varied and consistent pairings (in the associative learning task), and between classification accuracy and the chance level (in the grammar learning task), reflects the amount of implicit learning that has occurred. Although most researchers investigating implicit learning have made this assumption, it may not always be completely valid. For example, in the patterned sequence task one individual may anticipate the occurrence of the target signal and thus have response times close to zero, whereas another individual with the same level of learning may wait until the imperative signal occurs before initiating the response. The former individual would have a higher score on the measure used as an index of implicit learning when the actual amount of learning may not differ between the two people. (Note that excluding response times below a particular value does not eliminate this problem, but merely ensures that the analyses do not include responses in which the most blatant anticipations were detected and deleted.)

The second assumption of the current approach is that the research participants will not become aware of the stimulus structure, and possibly change their approach to the task such that the learning is no longer implicit. This is a potentially serious problem when several implicit learning tasks are performed in the same session. There are at least two methods of investigating this possibility. One is to ask participants if they were aware of the structure in each task, and then determine whether there were significant differences between aware and unaware participants in the relevant measures of performance. This method is not necessarily optimal because the validity of these types of self-reports is difficult to assess, and they may be especially inaccurate when the questions are delayed until after completion of all three tasks. A second method of examining possible reactive effects consists of comparing the results in these studies with those of other studies in which only a single task was performed, and thus one can have more confidence that awareness was not a problem.

To summarize, a major focus of the current study was to investigate alternative explanations for the weak to nonexistent age-related differences often found in measures of implicit learning. First, the possibility that the measures fail to exhibit relations with age because of low reliability will be examined by obtaining estimates of reliability. Second, shared influence analyses will be conducted to examine the implicit learning variables in the context of other variables known to be age-sensitive (i.e., measures of perceptual speed, reasoning, memory, and spatial abilities).

METHOD

Participants
Descriptive characteristics of the 183 participants are summarized in Table 1. The participants were recruited from newspaper advertisements, and they received nominal monetary compensation for their time and travel. It can be seen that the participants in each age group averaged over three years of college, but that the level of vocabulary was greater with increased age. An additional 9 individuals (mean age = 60.9 years) did not complete all of the tasks and thus their data were not included in the analyses.

Procedure
All participants performed the tasks in the following order within a single session: background demographic questionnaire, synonym and antonym vocabulary, pattern comparison, letter comparison,
**Sequence**

D - B - C - A - C - B - D - C - B - A

**Grammar**

![Diagram of Grammar]

**Associative**

![Diagram of Associative Learning]

Fig. 2. Illustration of the structure among the stimuli within each task. The series of letters in the sequence task refers to the order in which the stimuli occurred, with the letters signifying boxes from left to right on the screen. The transition diagram for the grammar task indicates the rules used to generate the strings of letters. In the associative learning task the pairs in the left column were consistently paired with one another, whereas those in the right column were paired equally often with each of the other items in that set.
Differential Aging Patterns

Table 1. Descriptive Characteristics of Research Participants.

<table>
<thead>
<tr>
<th>Age group</th>
<th>18–39</th>
<th>40–59</th>
<th>60–87</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>(SD)</td>
<td>Mean</td>
</tr>
<tr>
<td>N</td>
<td>67</td>
<td></td>
<td>71</td>
</tr>
<tr>
<td>Age</td>
<td>30.3</td>
<td>(5.7)</td>
<td>49.6</td>
</tr>
<tr>
<td>Percent females</td>
<td>68.7</td>
<td></td>
<td>70.4</td>
</tr>
<tr>
<td>Years of education</td>
<td>15.1</td>
<td>(2.7)</td>
<td>15.7</td>
</tr>
<tr>
<td>Self-rated health</td>
<td>2.1</td>
<td>(0.9)</td>
<td>1.9</td>
</tr>
<tr>
<td>Synonym vocabulary</td>
<td>4.5</td>
<td>(2.6)</td>
<td>7.1</td>
</tr>
<tr>
<td>Antonym vocabulary</td>
<td>4.2</td>
<td>(2.9)</td>
<td>6.4</td>
</tr>
</tbody>
</table>

Note. Health was assessed on a 5-point scale ranging from 1 for excellent to 5 for poor.

number matching, pattern matching, free recall, matrix reasoning, cube assembly, sequence learning, artificial grammar learning, digit digit reaction time, digit symbol reaction time, associative learning, and awareness questionnaire. Because the paper-and-pencil tests were administered in small groups, they were presented before the self-paced computer tests to allow participants to complete the sessions at their own rates. In addition to a break between the paper-and-pencil and computer tests, participants were allowed to rest between each block of computer-administered trials. Each of the three implicit learning tasks also had another very brief phase intended to assess explicit learning, but those data are not reported here because both the absolute levels of performance and the estimated reliabilities were quite low.

Sequence Learning

The sequence learning task consisted of five blocks of 100 trials each. The first four blocks contained 10 repetitions each of the same 10-element pattern (i.e., D-B-C-A-C-B-D-C-B-A, with the stimuli labeled A through D in left-to-right order). The fifth trial block consisted of randomly arranged stimuli (with the constraint that the same stimulus could not immediately follow itself), and was presented immediately following the patterned blocks without any notification to the participant of the change. The stimuli consisted of four boxes each subtending approximately 1.1°, with an asterisk (approximately 0.4°) serving as the imperative stimulus. An interval of 500 ms elapsed between the response to one stimulus and the presentation of the next stimulus. Responses consisted of the Z, X, , and / keys, with the first two fingers of each hand, for the boxes from left to right, respectively. The instructions were to respond by pressing the appropriate key as rapidly as possible after the asterisk appeared.

Artificial Grammar Learning

Stimuli in the grammar learning task consisted of strings of between three and six letters, with each letter subtending approximately 1.7° at a viewing distance of 40 cm. In the study phase 20 letter strings were presented three times each, with each presentation lasting 6 s. The instructions were to study the individual strings so that they would be remembered later.

The participants were next told that the previous strings had been generated by a set of rules, and that they should use those rules to classify 40 new letter strings as well-formed according to the rules, or as violating the rules. The / key was to be pressed for letter strings that followed the rules, and the Z key was to be pressed for letter strings that violated the rules. Accuracy was stressed more than speed, but no feedback was provided after the responses.

The materials and procedures for the artificial grammar task were very similar to those used in McGeorge et al. (1997) and Dulaney et al. (1984). The acquisition set of 20 items was identical to that in the earlier studies, and the 40 grammatical test items were identical except that items repeated from the acquisition set were not used, and five items (i.e., MTRVXR, RRXV, XRVX, TVXTV, and XTIVTV) were deleted from the nongrammatical set used in the earlier studies to have the same number of grammatical and nongrammatical test items.
Associative Learning

Stimuli in the associative learning task consisted of pairs of symbols resembling those in the bottom panel of Figure 2. A code table containing all of the symbol pairs was presented at the top of the computer display, and a probe pair of symbols appeared in the middle of the screen. The instructions were to decide as rapidly as possible whether the probe items matched according to the pairing of items in the code table on that particular trial. The stimuli were presented in boxes subtending visual angles of approximately 2.6° at a viewing distance of 40 cm, and the / key was to be pressed for a YES decision and the Z key for a NO decision. Four pairs of symbols were consistently paired with one another across all trials, and four pairs were variably mapped such that symbol A-1 might be associated with symbol B-1 on one trial, with symbol B-2 on another trial, etc. Furthermore, the order in which the symbol pairs appeared in the code table varied randomly from one trial to the next. Each of five trial blocks contained 48 trials, three correct and three incorrect mappings of each CM symbol and of each VM symbol.

The digit digit and digit symbol reaction time tasks were also administered on computers, and were presented in the following sequence: 18 trials of practice on digit digit, 90 trials on digit digit, 18 trials of practice on digit symbol, two blocks of 90 trials each on digit symbol, followed by a final 90 trials on digit digit. The stimulus displays in both tasks consisted of a code table at the top of the screen containing nine pairs of items, and a probe pair in the middle of the screen. The pairs of items in the digit symbol task consisted of digits and symbols, and thus the research participant had to refer to the code table to determine whether the items in the digit symbol probe pair were associated with one another. All items in the digit digit task consisted of digits, and thus the research participant merely had to decide whether the probe digits were physically identical. The instructions were that responses (i.e., / for same, and Z for different) should be made as rapidly and accurately as possible. Because accuracy averaged over 95%, performance in these tasks was summarized in terms of the median reaction times.

The remaining tasks were administered with paper-and-pencil procedures. Because the same procedures were used in another recent study (with an independent sample of participants), much of the following description is repeated from the other study (i.e., Salthouse, Fristoe, McGuthry, & Hambrick, 1998). The vocabulary tests consisted of 10 5-alternative multiple-choice questions each for the antonym and synonym portions, with 5 min allowed to complete both parts. Each of the perceptual speed tests (i.e., pattern comparison, letter comparison, number matching, and pattern matching) involved an instruction page with several examples, and two test pages for which participants were allowed 30 s each.

Items in the pattern comparison test consisted of pairs of patterns composed of three to nine line segments, and items in the letter comparison test consisted of pairs of three to nine letters. Approximately one half of the pairs in each test page were identical and one half differed in the identity of a single element. Participants were instructed to write an S between the pairs that were the same, and to write a D between the pairs that were different, and to do so as rapidly and accurately as possible.

Items in the number matching and pattern matching tests consisted of a two-digit target number or target pattern on the left, and five alternative numbers or patterns on the right. The task was to circle the alternative on the right that matched the target on the left, and to do so as rapidly and accurately as possible.

The free recall test was based on the Rey Auditory Verbal Learning Test (Schmidt, 1996), and consisted of five study-test trials of a list of 15 words presented in the same order (i.e., drum, curtain, bell, coffee, school, parent, moon, garden, hat, farmer, nose, turkey, color, house, river). The words were read by the examiner at a rate of about one word every 2 s, and 45 s was allowed for written recall after each list. Each recall attempt was written on a separate page in a booklet.

The matrix reasoning test resembled the Raven's Progressive Matrices Test in that items in the test contained a 3 × 3 matrix of geometric patterns with the lower right cell blank. Immediately below the matrix was a set of eight alternatives representing possible completions of the matrix. The task for the participant was to mark the best completion of the matrix from the set of alternatives. The test began with two practice problems with the answers provided, and then participants were allowed 10 min to complete as many of the 20 problems as possible.

The cube assembly test was based on a test originally described by Shepard and Feng (1972). Items in the test consisted of displays of six connected squares which were to be assembled to form a cube. One of the squares was shaded to represent the base of the cube, and two of the squares contained arrows. The task for the participant was to decide whether the arrows would point at one another when the cube was assembled, and to indicate the decision by marking Y for yes, or N for
no. Two illustrated practice problems preceded the test of 24 items, for which participants were allowed 10 min.

RESULTS

Many of the results are reported in the form of correlation coefficients. Rather than report the significance level after each correlation, it should be noted that with the current sample size an absolute correlation greater than .15 is significant at $p < .05$, and an absolute correlation of greater than .19 is significant at $p < .01$.

The data used to assess implicit learning in the three tasks are illustrated in Figure 3 as a function of age group. It can be seen that increased age was associated with slower responses in the sequence and associative tasks, but that there was little relation between age and accuracy in the artificial grammar task. Accuracy in the five reaction time blocks in the sequence task averaged more than 97% and was not significantly related to age. Accuracy in the reaction time blocks in the associative task averaged more than 93%, but the younger adults decreased their accuracy somewhat across successive trial blocks whereas the older adults did

![Sequence Learning Graph](image)

![Artificial Grammar Graph](image)

![Associative Reaction Time Graph](image)

Fig. 3. Top – mean (and standard error) of the median reaction times for the four blocks of patterned trials (1-4) and the block of random trials (5); lower left – mean (and standard error) of percentage correct judgments in the classification phase of the artificial grammar task; lower right – mean of the median reaction times across successive trial blocks for the varied mapping and consistent mapping pairs in the associative learning task.
not. This pattern was manifested in larger correlations between age and accuracy in the fifth trial block (i.e., \( r = .26 \) for VM and \( r = .23 \) for CM) than in the first trial block (i.e., \( r = .06 \) for VM and \( r = .08 \) for CM). However, because the shifts in accuracy were similar for the VM and CM trials, this possible age difference in speed-accuracy emphasis is unlikely to affect the interpretation of the relations of age to the measures of implicit learning in the associative task.

The measure of implicit learning in the sequence task was the difference in reaction time between the random (block 5) and the last pattern (block 4) block. Another possible measure is the decrease in reaction time in the pattern phase from blocks 1 to 4. The two measures were significantly correlated with one another (\( r = .34 \)), and had similar relations with age (i.e., \( r = -.11 \) for the block 4 minus block 1 difference, and \( r = -.17 \) for the random minus pattern difference). Because the decrease in reaction time from blocks 1 to 4 reflects general factors associated with familiarity with the task and with the manner of responding in addition to processes specific to the presented sequence, only the random minus pattern measure was used in subsequent analyses. The measure of implicit learning in the artificial grammar task was the level of accuracy in the classification phase when the participants were instructed to categorize novel letter strings according to the rules used to generate the previously studied letter strings. The implicit learning measure in the associative learning task corresponded to the difference in reaction time between VM stimuli and CM stimuli on the fifth trial block.

Table 2 contains summary statistics for the primary implicit learning variables in the three tasks as a function of age. The means for the implicit learning measure in the artificial grammar task (i.e., 58.6% to 60.0%) were not particularly high, but they were similar to the values of 60.9% in Reber et al. (1991) and the values between 58% and 62% in McGeorge et al. (1997).

Analyses of variance were also conducted on the implicit learning measures (i.e., difference scores in the sequence and associative tasks), with age group (young, middle, and old) as the only factor. The age effects were not significant for the sequence measure, \( F(2, 180) = 1.80, MSE = 11,786.59 \), or for the artificial grammar measure, \( F(2, 180) = 0.31, MSE = 93.58 \). However, there was a significant age difference for the associative measure, \( F(2,180) = 4.07, MSE = 126,636.78 \), and inspection of Table 2 reveals that the difference between CM and VM reaction times was larger with increased age. Note that this pattern implies that the degree of implicit learning was actually greater with increased age.

Increased age was associated with slower baseline reaction time in the sequence and associative tasks, and thus the implicit measures were also examined in terms of relative differences (i.e., the ratio of the block 5 minus block 4 difference to reaction time in block 4 of the sequence task, and the ratio of the VM minus CM difference to reaction time for CM trials in the associative task). Use of this relative measure changed the age correlation from \( r = -.17 \) to \( r = -.22 \) for the sequence implicit learning

<table>
<thead>
<tr>
<th>Task</th>
<th>18–39 Mean (SD)</th>
<th>40–59 Mean (SD)</th>
<th>60–87 Mean (SD)</th>
<th>Estimated reliability coefficient alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence</td>
<td>85.4 (89.7)</td>
<td>84.5 (83.1)</td>
<td>49.6 (158.7)</td>
<td>.74</td>
</tr>
<tr>
<td>Grammar</td>
<td>60.0 (8.8)</td>
<td>59.7 (9.5)</td>
<td>58.6 (11.1)</td>
<td>.40</td>
</tr>
<tr>
<td>Associative</td>
<td>135 (304)</td>
<td>226 (386)</td>
<td>330 (377)</td>
<td>.49</td>
</tr>
</tbody>
</table>

Note. Entries in the sequence and associative rows are in ms, and those in the grammar row are in units of percentage correct.
measure, indicating a slightly larger age difference favoring younger adults. However, changing from the absolute to the relative method of assessment decreased the correlation for the associative implicit learning measure from $r = .19$ to $r = .06$, indicating that there were no longer significant relations to age when the implicit learning measure was considered in proportional terms.

$t$ tests were also conducted to determine whether the measures of implicit learning were significantly greater than chance, which was zero for the sequence and associative tasks and 50% for the grammar task. All of the differences were significant ($p < .05$), with the following $t$ values for the young, middle, and old groups: 9.24, 8.64, and 5.18, respectively, for the grammar task; 7.80, 8.57, and 2.10, respectively, for the sequence task; and 3.64, 4.93, and 5.88, respectively, for the associative task.

Reliability Analyses
Reliability of the implicit learning variables was estimated with an internal consistency method based on correlations among the items comprising each measure. In the sequence learning task the individual stimulus positions and corresponding responses can be considered items, and in the artificial grammar task each letter string in the classification phase can be considered an item. Designation of items in the implicit learning phase of the associative learning task is somewhat arbitrary because the measure of learning in this task was the difference in average reaction time between the four VM and four CM stimulus pairs, and a given CM stimulus was not linked to any particular VM stimulus. The following procedure was therefore used to create items in this task. First the average reaction time was determined separately for each VM stimulus and for each CM stimulus. Next differences were computed between each pair of VM and CM stimuli. Finally, the four differences for each CM stimulus were averaged to serve as the individual items in the implicit learning phase of the associative task.

Cronbach’s alpha was then used to obtain estimates of reliability from the block 5 minus block 4 differences for the four items in the sequence task, from the 40 items in the grammar task, and from the average differences for the four items in the associative task. The estimates derived from this procedure are summarized in the last column of Table 2, where it can be seen that they are rather low. The implicit learning measure in the sequence task was an exception because the coefficient alpha estimate of reliability for that measure was a respectable .74. These results therefore indicate that even though the implicit learning phenomenon was evident at the group level in both the artificial grammar and associative tasks, the measures of the phenomenon had low reliability at the level of individuals.

Awareness Analyses
The next set of analyses was conducted to determine whether reports of awareness of the stimulus structure in the initial phases of the tasks were related to age, or to any of the learning measures. The specific questions and the percentage of positive responses to each were: (sequence) Did you realize that the asterisks were appearing in a regular pattern when you were doing the task? 69.2%; (grammar) Did you realize that there was a rule that was used to determine the order of the letters? 48.4%; and (associative) Did you realize that some of the symbols were always paired with one another when you were doing the task? 65.9%. These values thus indicate that between 1/2 and 2/3 of the participants reported that they were aware of the structure in each of the implicit learning tasks.

Point-biserial correlations were next examined in which aware responses were coded as 1 and unaware responses as 0. The correlations between these awareness reports and age were, respectively: sequence = -.24; grammar = -.08; and associative = -.13. These results indicate that increased age was associated with a somewhat lower likelihood of reporting awareness of the structure in each task, but only in the sequence learning task was the age relation significantly different from zero. This task was also the only one in which there was a significant correlation between self-reported awareness and implicit learning performance (i.e., sequence, $r = .35$; grammar, $r = .06$; and associative, $r =$
However, additional analyses revealed that none of the interactions of age and awareness on the measures of implicit learning were statistically significant. For example, the correlations between age and the measure of implicit sequence learning were -.02 for the aware individuals and -.24 for the unaware individuals, but these values did not differ significantly from one another. Correlations were also examined between the self-reports of awareness among the three tasks. These correlations were all small but positive (i.e., sequence – grammar = .21; sequence – associative = .35, and grammar – associative = .20), indicating that someone who reported awareness of the structure in one task was also likely to report awareness of the structure in other tasks.

Next, estimates of reliability for the implicit measures were computed separately for the aware and unaware individuals. The estimates were very similar for the two groups with the sequence implicit measure (i.e., unaware = .74, aware = .69), but they were lower for unaware individuals than for aware individuals with the grammar (i.e., unaware = .09, aware = .57) and associative (i.e., unaware = .28, aware = .56) measures. This pattern suggests that the reliability of implicit learning measures in the artificial grammar and associative tasks was even lower than that reported in Table 2 for individuals who reported that they were unaware of the stimulus structure, and presumably were least likely to have been influenced by explicit learning.

Finally, correlations between the implicit learning variables and other cognitive variables were computed separately for aware and unaware individuals. The other cognitive variables were the same as those used in the shared influence analysis to be described later (see Figure 3 for a listing). The differences between the correlations for aware and unaware individuals were tested in the context of an interaction of the implicit learning variable and a dummy variable representing awareness status in a regression equations predicting each cognitive variable. With the sequence implicit learning variable the median absolute correlation for aware participants was .29 and that for unaware participants was .20, but none of the differences between correlations was statistically significant. The medians for the implicit learning variable in the grammar task were .12 for aware and .12 for unaware participants, and again none of the correlations differed significantly from one another. Medians for the absolute correlations with the implicit associative learning variable were .13 for aware and .07 for unaware participants, with only the correlations for the letter comparison variable differing significantly from one another.

Although a substantial proportion of the research participants reported being aware of the stimulus structure in the implicit learning phase of the tasks, the analyses summarized above revealed few significant differences between aware and unaware individuals. Subsequent analyses were therefore based on the data from all participants regardless of their self-reported awareness status.

**Correlation Matrix**

The correlation matrix for all of the primary variables is contained in Table 3. Inspection of the table reveals the expected pattern of positive correlations with age for the vocabulary measures, and correlations indicative of poorer performance with increased age in the reaction time speed, paper-and-pencil perceptual speed, matrix reasoning, and free recall variables. Surprisingly, performance on the cube assembly test of spatial visualization was not significantly related to age; this may be attributable to the relatively low level of performance on this test (i.e., a mean of only 3.9 items correct of a possible 24). The correlations between age and the implicit learning measures were quite small (i.e., -.03, -.17, and .19 for grammar, sequence, and associative, respectively).

Several additional correlations were computed to examine the effects of excluding outliers and individuals with negative or below chance estimates of implicit learning. Two older adults (ages 63 and 72) had extremely large negative values for the sequence implicit learning variable, and thus several of the analyses were repeated after deleting their data. The mean and standard deviation in the 60-87 age group changed substantially (to 74.9 and 101.7, respectively), and the age correlation decreased to
Table 3. Correlation Matrix.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>1</td>
<td>.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Synonym Vocabulary</td>
<td>.41</td>
<td></td>
<td>.81</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antonym Vocabulary</td>
<td>.26</td>
<td>.83</td>
<td></td>
<td>.84</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digit Digit RT</td>
<td>.48</td>
<td>-.02</td>
<td>-.10</td>
<td>(.90)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digit Symbol RT</td>
<td>.53</td>
<td>-.08</td>
<td>-.17</td>
<td>.77</td>
<td>(.94)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pattern Comparison</td>
<td>.43</td>
<td>.11</td>
<td>.15</td>
<td></td>
<td>-.49</td>
<td>-.53</td>
<td>.53</td>
<td>(.80)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Letter Comparison</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pattern Matching</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number Matching</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cube Assembly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Matrix Reasoning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Free Recall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grammar Implicit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sequence Implicit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Associative Implicit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>47.1</td>
<td>6.2</td>
<td>5.6</td>
<td>735</td>
<td>1449</td>
<td>15.8</td>
<td>10.3</td>
<td>15.2</td>
<td>26.7</td>
<td>3.9</td>
<td>8.6</td>
<td>10.7</td>
<td>59.5</td>
<td>76.3</td>
<td>218</td>
</tr>
<tr>
<td>SD</td>
<td>15.7</td>
<td>2.9</td>
<td>3.2</td>
<td>142</td>
<td>356</td>
<td>3.6</td>
<td>2.8</td>
<td>3.9</td>
<td>3.5</td>
<td>7.0</td>
<td>4.4</td>
<td>1.9</td>
<td>9.6</td>
<td>109</td>
<td>362</td>
</tr>
</tbody>
</table>

Note. N = 183. RT = reaction time. Correlations with an absolute value greater than .19 are significantly (p < .01) different from zero. Estimated reliabilities (in parentheses) were derived from boosting the correlation between the scores on the two parts by the Spearman-Brown formula, or from coefficient alpha for tests with only one administration. For the free recall score the reliability was estimated from the correlations among the five trials according to the formula: reliability = [n(avg. r)]/[1+(n-1)(avg. r)]. Estimates of reliability for the implicit learning measures were derived from the coefficient alpha procedure described in the text.
However, the estimated reliability was still in the moderate range \((r = .68)\), and correlations with other variables were nearly unchanged from the values in Table 3 (e.g., \(-.31\) for Digit Symbol Reaction Time (DSRT), \(.29\) for Recall, and \(.33\) for Matrix Reasoning).

Excluding data from participants with values of less than 50% (the chance value) in the artificial grammar implicit learning variable reduced the estimated reliability to \(.01\), but had little effect on the correlations with age, \(r = -.09\), or the correlations with other cognitive variables (e.g., DSRT, \(r = -.13\); Recall, \(r = .20\); and Matrix Reasoning, \(r = .27\)). Excluding data from participants with sequence implicit learning scores of less than zero reduced the estimated reliability to \(.61\), and reduced the correlations with age \((r = -.06)\) and with other variables (e.g., DSRT, \(r = -.23\); Recall, \(r = -.15\); and Matrix Reasoning, \(r = .18\)). Finally, excluding data from participants with associative implicit learning scores of less than zero reduced the estimated reliability to \(.19\) and changed the age correlation to \(-.18\) (from \(.19\)) but had relatively little effect on other correlations (e.g., DSRT, \(r = .19\); Recall, \(r = -.04\); and Matrix Reasoning, \(r = -.12\)).

These results suggest that although the magnitude of the correlations was reduced when the range of variation was restricted by eliminating individuals with negative implicit learning scores, the implicit learning variable from the sequence learning task continued to have the highest estimated reliability among the measures of implicit learning.

**Shared Influence Analyses**

The goal of the final set of analyses was to investigate the interrelation of the age relations on different variables in the context of a simple structural equation model. Even though two of the implicit learning variables had low estimates of reliability, and thus might not be expected to be related to other variables, all of the implicit learning variables were included in the analyses for the sake of completeness. The analytical procedure consisted of two steps. In the first step, all variables were specified to have relations to a common factor, which in turn was specified to be related to age. All relations that were not statistically significant in this analysis were deleted, and the coefficients for the remaining relations were then fixed to the estimated values. In the second step of the analysis, direct relations from age to the individual variables were examined after taking into consideration the relations of age to what all the variables had in common. Any statistically significant relation between age and a variable was retained in the model, and all others were deleted. The results of this procedure are summarized in Figure 4, which portrays the standardized regression coefficients for all statistically significant relations. Fit statistics are not particularly meaningful in this type of analysis because there is no attempt to model any relations between variables except those involving age. Nevertheless, the overall fit of this simple model was not bad, as revealed by several indexes (i.e., \(\chi^2 = 191\), \(df = 75\); Non-Normed Fit Index = .88, Comparative Fit Index = .88; Standardized Root Mean Residual = .07).

There are two important points to be noted about Figure 4. The first is that only the implicit learning variable from the associative task was not significantly related to the common factor representing the variance shared among all variables. As noted in the introduction, if the variable was reliable then a pattern such as this would be consistent with the interpretation that the variable represents a qualitatively different type of processing than that involved in the other variables. However, the reliability of the associative implicit learning measure was quite low (i.e., coefficient alpha of .49), and thus the absence of a loading on the common factor may simply be attributable to a lack of systematic variance in the measure.

The second point to note about Figure 4 is that only the cube assembly variable had a significant relation from age after taking into consideration the relation between age and what all of the variables had in common. In this particular case the unique age relation occurs because the age-related influences mediated through the common factor, estimated by the products of the age-common and common-variable coefficients, (i.e., \(-.61 \times .38 = -.23\)) was larger than the observed age correlation (i.e., \(-.09\)). The substan-
tive interpretation of this independent age-related influence is not yet clear.

**DISCUSSION**

The results of this study replicated previous findings of small to nonexistent age relations on measures of implicit learning (e.g., Cherry & Stadler, 1995; Frensch & Miner, 1994, Expt. 3; Howard & Howard, 1989, 1992; McGeorge et al., 1997), and generally weak relations between implicit learning measures and other cognitive variables (e.g., McGeorge et al., 1997; Reber et al., 1991). However, these results are not easily interpretable without considering the reliability of the measures. Reliability estimates were low for two of the three implicit learning measures, and this was particularly true among participants who reported that they were unaware of the stimulus structure during the implicit learning phase. (There are no absolute criteria for the interpretation of reliability estimates, but most textbooks on psychometrics suggest that internal consistency estimates of reliability should be greater than .7 in order for the assessment to be meaningful, e.g., Kline, 1993; Murphy & Davidshufer, 1994.) Because there is little evidence that these variables were consistent at the level of individuals, the lack of age relations on the variables could simply be an artifact of unreliable measurement, and may not have any substantive meaning.

The measure of implicit learning in the sequential reaction time task did have adequate reliability as assessed by the internal consistency (coefficient alpha) method. However, this measure had small negative relations with age
and small positive relations with other cognitive variables, and thus it resembled the pattern with many other cognitive variables.

One possible reason for the apparent variation in results across studies is that the age-related effects on implicit learning in the current study were rather small, and many of the earlier studies had relatively low power to detect small or moderate effect sizes. For example, with the sample sizes of 20 per group used in the Howard and Howard (1989, 1992) studies, the probability of detecting what Cohen (1988) refers to as a small effect ($d = .2$) was only .09, and even a moderate effect ($d = .5$) would only have a .33 probability of detection with these sample sizes.

Another factor that may have contributed to some of the variation across studies is the extent to which the research participants were aware of the stimulus structure during the initial learning phase of the tasks. The methods used to assess awareness and the percentages of participants classified as being aware of the stimulus structure have varied across studies, and consequently the studies may have differed in the degree of contamination of the implicit learning measures by influences of explicit learning. The issue of how to ensure that measures of what are purported to reflect implicit learning do not also reflect aspects of explicit learning is difficult to resolve, and at least some of the variation in results across studies may reflect the success with which researchers achieved this goal. However, it is important to note that in the present study the coefficient alpha estimates of reliability for the implicit sequence learning measure were similar for aware and unaware individuals, and if anything, the age relations were larger for the unaware individuals than for those individuals who reported that they were aware of the stimulus structure. These findings raise the possibility either that the awareness reports for measures of implicit learning were of limited validity, or that awareness status may have minimal effects on the psychometric properties of implicit learning variables, and on their patterns of relations to other variables.

The shared influence analysis based on a single common factor was designed to examine age-related influences on individual variables in the broader context of age-related influences on other variables. If measures of implicit learning represented a qualitatively different pattern of processing than that involved with other cognitive variables then it might have been predicted that the measures of implicit learning would be reliable, but unrelated to other variables. The lack of a significant relation with other variables, evident by the absence of a loading on the common factor, would have been consistent with the view that those variables reflect something different than the other variables. A second interesting outcome from the single common factor analysis could have been that the implicit learning variables had significant relations to other variables, indicating that they shared some aspects, but they might also have had independent or unique effects related to age, indicating that something else was operating to counteract the expected negative age-related influences.

As indicated in Figure 4, neither of these patterns was observed. Instead the patterns for the implicit learning variables from the sequential reaction time and artificial grammar tasks were generally similar to those for the other variables, although with weaker relations from the common factor. The implicit learning variable from the associative task was not related to the common factor, but its low reliability indicates that there was little systematic variance available to be associated with other variables. These results therefore imply that the weak relations between age and implicit learning variables were not because they represented a qualitatively distinct form of cognition or because they had additional specific age-related effects, but because they either had less of what is common to all variables or had too little systematic variance (i.e., reliability) to allow relations with age to be detected.

These conclusions are somewhat different than those reached by Hultsch, Masson and Small (1991) and Small, Hultsch and Masson (1995) in studies of implicit memory. In their studies measures of implicit memory (stem completion in both studies and fact completion in the 1995 study) were found to have moderate levels of estimated reliability, but very small correla-
tions with other cognitive variables. Because the correlations involving the implicit memory variables were close to zero, and much smaller than those among other types of cognitive variables, these researchers concluded that implicit memory tasks involve different cognitive processes than other types of (explicit) cognitive tasks. It is not yet clear what is responsible for the discrepancies in the two sets of results, but it is possible that measures of implicit memory are more likely to reflect a qualitatively different form of processing than measures of implicit learning. For example, it is conceivable that implicit memory may be dependent on different neuroanatomical structures than those involved in explicit memory, implicit learning, and other cognitive variables.

In summary, the results of this study indicate that when evaluating the effects of age on a variable it is important to consider not just the absolute magnitude of the relation of a variable to age, but also the reliability of the variable and the pattern of relations that the variable has with other variables. In this study it appears that implicit learning as typically assessed in the sequential reaction time task can be measured reliably at the level of individuals, and that it exhibits relatively small age-related effects. However, it does not appear to reflect a qualitatively different type of processing than that involved in other variables because the smaller age relations appear to be attributable to weaker relations of the implicit variables to whatever it is that many cognitive variables have in common. More generally, the strategy of first examining reliability of the critical variable followed by assessing its pattern within a shared influences analysis appears promising as a means of understanding the specific manner in which some cognitive variables appear to be spared from age-related influences.

REFERENCES


