Journal of Experimental Psychology:
Learning, Memory, and Cognition
1994, Vol. 20, No. 5, 1172–1182

0278-7393/94/$3.00

Interpretation of Differential Deficits: The Case of Aging and Mental Arithmetic

Timothy A. Salthouse and Vicky E. Coon

A fundamental issue in research on individual differences is the type of evidence sufficient to justify an inference of selective or distinct deficits in relevant theoretical processes. It is proposed that an important consideration is the extent to which the individual differences in 1 variable are independent of those in another variable. Specifically, the suggestion presented here is that a strong conclusion of selective impairment requires evidence that there is significant group-related variance in 1 variable after the variance in the other relevant variable is controlled. Furthermore, an inference that the groups are equivalent on a particular theoretical process requires evidence that the variable presumed to reflect that process has sufficient unique variance to justify the claim that a distinct process is being assessed. The proposed methods are illustrated with two studies comparing adults of different ages in mental arithmetic tasks.

A key issue in attempting to interpret results from studies involving comparisons of people from different groups (which could be formed on the basis of age, gender, neurological status, or any other classification of interest) concerns the type of evidence sufficient to warrant a conclusion of selective or differential group-related influences. Most researchers tend to rely on tests of interactions in an analysis of variance (ANOVA) framework to infer that the effects associated with group membership are greater in one variable than in another. (See Kausler, 1982, pg. 200–219, for a detailed discussion of the rationale underlying the interpretation of interactions in age-comparative research.) The reasoning can be outlined with the aid of Figure 1A, with subject age as the classification variable. Assume that variable 1 primarily reflects the efficiency or effectiveness of Process Y, and that variable 2 primarily reflects the efficiency or effectiveness of Process Z. If the age-related effects are larger on variable 2 than on variable 1, then the typical inference would be that Process Z is particularly age-sensitive. A result of this type is frequently interpreted as evidence for the localization of age-related effects, or at minimum, as suggesting that the age-related effects are differential and selective. To illustrate, imagine that variable 2 represented the time required to run 50 meters, and that variable 1 represented hand grip strength as measured by a hand grip dynamometer. If the Age × Variable interaction were in the direction of larger age differences in the running time measure than in the hand strength measure, one might infer that processes related to aerobic fitness were more susceptible to age-related influences than were processes associated with peripheral muscle strength.

Timothy A. Salthouse and Vicky E. Coon, School of Psychology, Georgia Institute of Technology.

This research was supported by National Institute on Aging Grant R37 AG06826 to Timothy A. Salthouse. We would like to acknowledge the valuable comments of Julie Earles, David Geary, Alan Kersten, and John Nesselroade on an earlier version of this article.

Correspondence concerning this article should be addressed to Timothy A. Salthouse, School of Psychology, Georgia Institute of Technology, Atlanta, Georgia 30332. Electronic mail may be sent to *tim.salthouse@psych.gatech.edu*.

Numerous reservations have been expressed regarding the interpretation of interactions in this manner (e.g., Baron & Trieman, 1980; Bogartz, 1976; Chapman & Chapman, 1973, 1988; Loftus, 1978), particularly when the groups being compared differ in the baseline level of performance (perhaps as indicated by a difference in variable 1). Among the issues that have been raised are that interactions vary according to the relations between process and variable and as a function of the discriminating power of the variables. Discussions focusing on the problems of interpreting interactions in the context of adult developmental research are contained in Salthouse (1991) and Salthouse and Kausler (1985).

We propose that there are at least two additional complications associated with the reliance on statistical interactions as the basis for inferring the presence or absence of differential or selective age-related influences. Our arguments can be elaborated by reference to Figures 1B and 1C. The illustration in Figure 1B is intended to represent the possibility that interactions might emerge not because of selective age-related influences operating directly on the relevant processes, but because those processes differ in the demands they make on a common factor (or "processing resource") that is related to age. A possible example of this type of situation is when variable 2 corresponds to the time to run 50 meters, and variable 1 represents the time to swim 50 meters. In this case, it seems reasonable to speculate that both types of performance might be influenced by aerobic or cardiovascular fitness, and therefore that at least some of the age-related effects in each variable could be indirect and mediated through that common influence. Notice that because of variations in the dependence of each variable on the common influence, this conceptualization does not preclude the existence of significant interactions. That is, in the running versus swimming example, it is conceivable that swimming could be more dependent on overall cardiovascular fitness than running because all four limbs are directly involved in propulsion in swimming compared with only two for running. If this is the case, and if factors such as amount of experience with each activity were equal, then the magnitude of the age differences in swimming time might be larger than those in running time. The important point for the

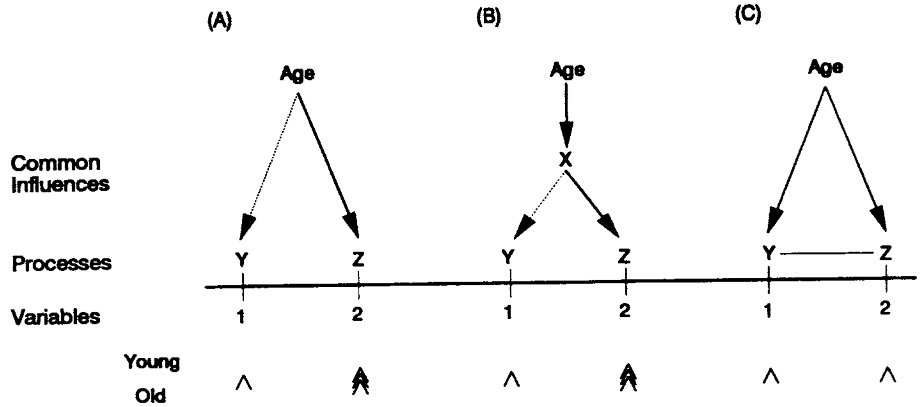

Figure 1. Schematic illustration of alternative interpretations of Age × Variable interactions. Note that variable 1 is assumed to reflect Process Y and variable 2 is assumed to reflect Process Z, and that the age differences in (A) and (B) are larger in variable 2 than in variable 1.

current discussion is that under the circumstances represented by Figure 1B, differential age-related effects would not necessarily imply that the age-related influences were distinct, unique, or independent.

Interpretations are also complicated when the relevant interactions are not statistically significant. Of course, one obvious problem is relying on acceptance of the null hypothesis to conclude that there are no true differences. This concern can be at least partially addressed by conducting power analyses (e.g., Cohen, 1988; Kraemer & Thiemann, 1987; Lipsey, 1990) to assess the probability that an effect of a given magnitude would have been detected if it were to exist.

A theoretically more interesting reason for a failure to detect a significant interaction is that the variables might not represent truly distinct processes. That is, if the variance shared across the two variables is very high relative to the respective reliabilities, then there may be little unique or independent variance in each variable. This situation is illustrated in Figure 1C where the line between Processes Y and Z is intended to signify that the processes are not independent because they share a large proportion of their systematic variance.

Continuing with the example of the time required to run 50 meters as variable 2, consider the situation if variable 1 corresponded to the time required to run 40 meters. With moderately sensitive measures there would almost certainly be a significant difference between variables 1 and 2. However, it is doubtful that distinct and independent processes contribute to the two variables. Instead it may be more plausible to assume that both variables are determined by the same processes, such that there is little unique variance in one of the variables that is independent of the variance in the other variable.

Distinguishing among the conceptualizations represented in Figure 1 requires methods other than conventional ANOVAs to evaluate the independence of the variance in each variable. We propose that correlation-based techniques can be used for this purpose. Specifically, hierarchical multiple regression procedures could be used to determine the amount of unique or independent age-related variance in variable 2 after the variance in variable 1 is controlled, and the proportion of shared variance between variables 1 and 2 can be contrasted with the reliability of each variable to determine the unique variance of each variable.

A possible sequence of statistical evaluations is therefore as follows. First, determine whether the Age × Variable interaction is significant with traditional ANOVA or multiple regression techniques. If the interaction is significant, examine the amount of unique or independent age-related variance in the variable with the greater age difference by determining the increment in variance associated with age after the variance in the other variable was controlled. If the residual age-related variance is significant, then a conclusion that the processes were selectively and independently influenced by age would be warranted. However, a conclusion of little or no independent influence would presumably be implied by a finding of no significant residual age-related variance.

If the interaction is not significant, and the power of the comparisons is at least moderate, then the independence of the two variables should be examined by inspection of the correlation between the variables and their respective reliabilities. The correlation by itself is inadequate for the current purposes because its magnitude is limited by the amount of systematic, or true score, variance in each variable, which is represented by the reliability of the variable. No absolute criteria will be specified for when two variables can be considered sufficiently independent to justify the postulation of separate processes, but the degree of independence obviously decreases with increases in the ratio of the proportion of shared variance (r^2) to systematic variance (reliability).

The proposed research strategy can be illustrated with data from a recent project by Salthouse (1993). The primary question of interest in that project was whether there were selective, and independent, age-related influences on the slowest and fastest responses produced by an individual in a choice reaction time task. That is, it could be hypothesized that a major cause of age-related slowing is an increase in the frequency of attentional blocks or lapses of attention, which are likely to have the greatest effects on the individual's slowest responses. For purposes of this investigation, fast and slow were defined in terms of the distribution of the individual's own response times, with fast corresponding to the 10th percentile of the individual's response time distribution, and slow corresponding to the 90th percentile of his or her distribution. Analyses contrasting the magnitude of the age differences for the fast and slow responses revealed that the age differences were significantly larger for the slow responses than for the fast responses. A traditional interaction interpretation might therefore suggest that increased age was associated with an alteration of processes (e.g., lapses of concentration, failure to inhibit irrelevant information) primarily responsible for producing very slow responses.

However, hierarchical regression analyses were also conducted in which the variance associated with the fast responses was controlled before examining the variance associated with age in the slow responses. These analyses revealed that there was no significant residual age-related variance in the slow responses after the variance in the fast responses was controlled. That is, although the absolute magnitude of the age differences was larger for the slowest responses, the age-related variance in these responses was not unique, in the sense of being independent of the age-related variance in the fastest responses. Because the same pattern was evident with response time measures from two different tasks across four independent studies, with between 100 and 258 subjects in each study, the basic phenomenon appears robust.

The results just described indicate that although the age-related effects varied in absolute magnitude, they were not independent or distinct. This pattern is therefore consistent with the interpretation represented in Figure 1B, in that a common influence may be contributing to both variables. The primary evidence in support of this view is that if the common influence is held constant by controlling the variance in one variable, then many of the age-related effects in the other variable are also eliminated.

Although the necessary data to allow the relevant calculations to be performed have not been published, there are numerous reports in the aging and cognition literature in which the interpretation represented in Figure 1C might apply. For example, the failure to find Age × Priming interactions (Burke, White, & Diaz, 1987; Howard, Shaw, & Heisey, 1986) or Age × Attentional Shift interactions (e.g., Hartley, Kieley, & Slabach, 1990; Madden, 1986) could be a consequence of too little independent variance in the relevant variables to allow differential relations with variables such as age to be detected. According to the guidelines proposed above, the power of the comparisons and the amount of unique systematic variance in each variable should be considered before accepting a conclusion that the lack of an interaction necessarily implies that distinct processes had equivalent age-related effects.

To summarize, we propose that additional types of analyses are needed to determine whether Group × Variable interactions should be interpreted in terms of a differential deficit or selective influence on the relevant theoretical processes. The analytical methods are applied in two studies conducted to examine age-related effects in mental arithmetic. To investigate the possibility that measures derived from separate tasks might be additional manifestations of the hypothesized common influence (i.e., variable *x* Figure 1B), measures of perceptual-motor processing speed obtained from other paper-and-pencil and computer-administered tests were also included. These particular variables were used because of an assumption that a reasonable candidate for the hypothesized common construct is the speed with which the individual can execute elementary processing operations. However, variables representing other constructs could obviously be used in these types of analyses.

Mental Arithmetic

Recent reports of no age differences, or in one case even better performance by older adults than by young adults, in measures of hypothesized components of mental arithmetic are of great interest because of the common finding of age differences favoring young adults in many cognitive tasks. Nearly all studies find that older adults are slower than young adults in arithmetic involving addition, subtraction, or multiplication (Allen, Ashcraft, & Weber, 1992; Birren, Allen, & Landau, 1954; Birren & Botwinick, 1951; Charness & Campbell, 1988; Geary & Wiley, 1991; Rogers & Fisk, 1991; Salthouse & Kersten, 1993). However, more complex patterns of results have been reported with respect to interactions between age and measures presumed to reflect components of mental arithmetic. The interaction between age and the similarity of the incorrect answer to the correct answer was significant in a study by Rogers and Fisk (1991), but it was not significant in a later study by Allen et al. (1992). Tests of the interaction between age and answer magnitude were not significant in studies by Allen et al. or Geary, Frensch, and Wiley (1993). The Age × Answer Magnitude interaction was also inferred not to be significant in a study by Geary and Wiley (1991), although these investigators did not test it directly and instead based their conclusion on analyses of relations among group means rather than an analysis of variance conducted on the measures from individual subjects. The absence of significant interactions is of considerable theoretical interest because findings of this type have been interpreted as indicating that some hypothesized processes, such as the rate of retrieval from long-term memory, may be spared from the ubiquitous age-related slowing.

Geary et al. (1993) have recently reported that older adults were significantly faster than young adults at performing borrow operations in subtraction, as inferred from the difference in the time to perform subtraction problems that did or did not require borrowing. This is surprising both because older adults are seldom faster than young adults (Salthouse, 1985) and because borrowing problems presumably involve working memory that is often assumed to decline in efficiency with increased age. There are, however, reasons to be cautious about the Geary et al. finding. For example, the older adults in that study were highly educated, with 14 out of 36 having advanced graduate degrees, and the young and old adults did not differ in either the performance of paper-and-pencil or computer-administered tests with problems involving subtraction of a single digit from a two-digit number. The results of their study may therefore be specific to particular samples of highly educated older adults and possibly moderate-to-low-

Table 1
Demographic Characteristics of Research Participants

	Study 1								Study 2			
	Students (n = 64)		19–39 (years) (n = 183)		40–59 (years) (n = 85)		60–82 (years) (n = 72)		Students (n = 40)		60–80 (years) (n = 40)	
Characteristic	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Age (years)	19.6	1.6	30.0	5.4	49.8	6.0	67.0	5.6	19.4	1.3	71.3	4.4
Education (years)	13.2	1.9	14.1	2.7	14.0	2.4	13.3	2.4	13.5	1.3	15.8	2.5
Health[a]	2.0	1.8	1.8	0.9	2.4	1.1	2.4	1.1	1.6	0.7	1.6	0.7
Boxes[b]	58.3	14.5	58.4	11.3	50.9	10.5	44.1	12.8	55.7	15.4	50.5	13.5
Digit copy[b]	59.1	8.1	59.0	10.1	52.8	9.7	46.2	8.6	58.7	7.0	48.5	7.9
Letter comparison[b]	12.6	2.5	11.3	2.9	9.5	2.8	8.0	2.5	13.3	2.3	8.9	3.3
Pattern comparison[b]	20.6	3.7	18.2	3.7	15.3	2.9	12.5	2.8	20.8	3.8	13.7	2.7
Digit digit												
Accuracy (%)	95.6	2.8	96.5	3.6	97.5	2.1	96.7	3.7	96.5	3.1	95.5	8.7
Time (ms)	586	70	720	119	841	200	974	226	556	62	836	264
Digit symbol												
Accuracy (%)	95.6	2.8	96.0	3.3	96.9	3.2	96.0	3.9	96.2	2.7	96.4	3.1
Time (ms)	1,063	189	1,367	287	1,642	343	1,909	393	1,006	213	1,754	325
Computation span	—		—		—		—		5.6	1.8	4.0	1.6
Reading span	—		—		—		—		3.4	1.4	2.3	1.2

[a]Ranked on a scale ranging from *excellent* (1) to *poor* (5). [b]Number per 30 s.

ability young adults who perform equivalently in tests of arithmetic speed.

Study 1

The data from this study were obtained from a verification subtraction task performed by participants in a larger project (Salthouse, in press). One half of the problems required a borrowing operation (e.g., 72 − 9 = 63, TRUE), and one half did not (e.g., 46 − 2 = 43, FALSE). If the results of Geary et al. (1993) are to be replicated, then there should be a significant Age × Problem Type interaction in the direction of a smaller difference between the two types of problems with increased age.

Method

Subjects. Demographic characteristics of the 240 participants in this study are summarized in Table 1. The table also contains characteristics of an additional sample of 64 college students whose data were used in a supplementary analysis.

Procedure. The subtraction arithmetic task was performed on a microcomputer after the subject read a brief description of the task. The arithmetic problems were presented in a verification format, with subjects instructed to press the Z key on the computer keyboard as rapidly as possible for FALSE (or INCORRECT) problems and to press the slash key as rapidly as possible for TRUE (or CORRECT) problems.

The arithmetic problems were presented on a single line of the display monitor and always contained a two-digit number in the first position and a single digit in the second position. On one half of the trials the solution to the problem required borrowing from the tens column, and on one half of the trials it did not. Incorrect problems, which compared one half of all trials, had an answer that differed from the correct solution by plus or minus one digit in either the units or the tens column of the answer.

After reading the instructions, subjects performed a block of 10 practice trials, followed by two blocks of 48 experimental trials each. The distribution of correct and incorrect problems, and problems with and without the borrowing requirement, was random within each block of trials. Instructions emphasized that both time and accuracy were important in the task.

Other measures. All participants in this study also performed several additional tasks, including six that were intended to measure various types of processing speed. Four of the speed tests were in a paper-and-pencil format, with a 30-s time limit allowed to complete as many items as possible. Two of these, boxes and digit copy, were postulated to assess sensory and motor speed because the subject merely had to draw a line on a three-sided figure with a missing side to form a box (boxes) or to copy digits in the space immediately below them (digit copy). The letter comparison and pattern comparison tasks were hypothesized to assess perceptual speed because the subject had to make judgments about whether a pair of letter strings (letter comparison) or line patterns (pattern comparison) were the same or different.

Two computer-administered speed tasks were also performed. The digit symbol test involved the presentation of a code table containing digits paired with symbols and probes of a digit paired with a symbol. The subject was instructed to decide as rapidly as possible whether the digit and symbol were associated according to the code table. If the digit and symbol were associated in the code table then the slash key was to be pressed, and if they were not paired in the code table then the Z key was to be pressed. The digit digit version of the task was identical except that the symbols were replaced with digits, and thus the yes–no decision was based on physical identity rather than associational equivalence. In both tasks subjects were instructed to respond as rapidly and accurately as possible.

Results

Mean values for the measures of performance from the nonarithmetic speed tasks are summarized in Table 1. As expected, the age relations were significant ($rs > .44, p < .01$) in all four paper-and-pencil measures and in the time measures from the digit symbol and digit digit tasks. For purposes of later analyses, the paper-and-pencil speed measures were converted to z scores, a motor speed index was created by averaging the boxes and digit copy ($r = .67$) scores, and a

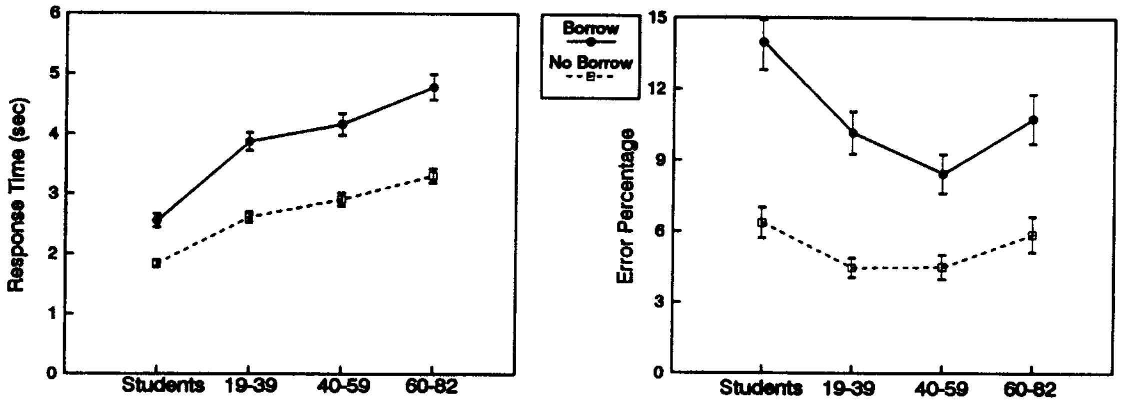

Figure 2. Mean response time and error percentage for subtraction problems with and without borrowing requirements as a function of age, Study 1. Bars above and below each data point correspond to one standard error.

perceptual speed index was created by averaging the letter comparison and pattern comparison ($r = .62$) scores.

We estimated reliabilities of the measures of arithmetic performance by boosting the correlations between the measures from the first and second trial blocks with the Spearman–Brown formula. The estimated reliabilities were .93 for response time on no-borrow problems, .93 for response time on borrow problems, .63 for error percentage on no-borrow problems, and .73 for error percentage on borrow problems. The correlations between response time and error percentage were −.10 for no-borrow problems and .01 for borrow problems, suggesting that any between-subjects speed–accuracy trade-offs were relatively minor. A borrow–no-borrow difference score was also used in some analyses, and its estimated reliability was .76.

Means of the median response times and percentage of errors for the no-borrow and borrow problems are illustrated as a function of age in Figure 2. Age (19–39, 40–59, 60–82, in years) × Problem Type (no borrow vs. borrow) ANOVAs conducted on these data revealed the following significant effects. For the response time measure: Age, $F(2, 237) = 8.60$, $p < .01$, $MS_e = 2.93$; and problem type, $F(1, 237) = 474.77$, $p < .01$, $MS_e = 0.044$; but not Age × Problem Type, $F(2, 237) = 1.37$, $p > .25$. For the error percentage measure: Problem type, $F(1, 237) = 129.50$, $p < .01$, $MS_e = 21.70$; but not age, $F(2, 237) = 1.92$, $p > .14$, $MS_e = 69.85$; or Age × Problem Type, $F(2, 237) = 1.49$, $p > .20$.

We also conducted a one-way ANOVA on the difference between the borrow and no-borrow response times. Consistent with the lack of a significant Age × Problem Type interaction, the age effect in this analysis was not significant, $F(1, 237) = 1.37$, $p > .25$, $MS_e = 0.88$. The power of this analysis to detect an effect of moderate size ($f = 0.25$) with an alpha of .05 was .94.

Because Geary et al. (1993) compared young adult students with older adults, additional analyses were conducted in which the data from the 64 students were contrasted with the data from the 72 adults above 60 years of age. Neither the age main effect nor the Age × Problem Type interaction was significant for the error percentage measure, but both were significant with the response time measure. That is, age, $F(1, 134) = 104.53$, $p < .01$, $MS_e = 2.22$, and Age × Problem Type, $F(1, 134) = 25.24$, $p < .01$, $MS_e = 0.39$. The age effect was also significant in an analysis on the borrow minus no-borrow difference score, $F(1, 134) = 25.24$, $p < .01$, $MS_e = 0.79$, with a mean of 0.71 s for students and 1.47 s for older adults. We repeated the analyses after eliminating subjects with less than 13 years of education to provide a comparison of high-education subjects. Although this resulted in a reduction in sample size to only 23 students and 26 older adults, the pattern of results remained unchanged. That is, the age, $F(1, 47) = 29.76$, $p < .01$, $MS_e = 1.91$, and Age × Problem Type, $F(1, 47) = 8.98$, $p < .01$, $MS_e = 0.35$, effects were still significant, as was the age effect on the difference score, $F(1, 47) = 8.98$, $p < .01$, $MS_e = 0.84$ (students = 0.64 s and older adults = 1.36 s). In the comparison that appears to be most comparable to that of Geary et al. (1993), therefore, the results are exactly the opposite of what they report.

We conducted a hierarchical regression analysis on the data from the college students and the older adults in which a significant Age × Problem Type interaction was found in the ANOVA. Age was a significant predictor of borrow response time when it was the only predictor in the regression equation, that is, $R^2 = .344$, $F(1, 47) = 21.09$, but not when it was entered in the regression equation after the no-borrow response time measure, that is, increment $R^2 = .000$, $F(1, 46) = 0.06$. With the student–older adult data, therefore, the situation most closely resembles that portrayed in Figure 1B in that the absolute magnitude of the age differences was larger in one measure than in another, but the two measures appear to have little or no unique or independent age-related variance.

Results of hierarchical regression analyses conducted on the no-borrow, borrow, and difference response time measures from the sample of nonstudent adults are summarized in Table 2. Two points should be noted from this table. The first is that there was no significant age-related variance in either the difference score measure or in the borrow response time measure after the variance in the no-borrow response time

Table 2
Proportion of Age-Related Variance in Mental Arithmetic Response Time, Study 1 (N = 240)

Analysis	No borrow	Borrow	Difference
Age alone	.087*	.051*	.008
Age after no borrow	—	.001	.003
Age after health and education	.058*	.032*	.004
Age after health, education, and error	.056*	.032*	.005
Age after health, education, error, and P & P MSpd	.034*	.013	.000
Age after health, education, error, and P & P PSpd	.002	.000	.005
Age after health, education, error, and digit digit	.001	.000	.001
Age after health, education, error, and digit symbol	.004	.003	.002

Note. Health and education information is based on self-reports (Table 1). Error is percentage of errors in the arithmetic task. P & P MSpd is the paper-and-pencil motor speed index (i.e., boxes and digit copy), and P & P PSpd is the paper-and-pencil perceptual speed index (i.e., letter comparison and pattern comparison). Dash indicates data not applicable to analysis.
$^*p < .01$.

measure was controlled. These results are thus consistent with one another and with the ANOVA results previously reported.

The second important point to note from Table 2 is that the age-related variance in both of the arithmetic measures was eliminated after controlling the variance in the paper-and-pencil perceptual speed measure, the digit digit measure, or the digit symbol measure. As mentioned in the introduction, this is the pattern one would expect if a factor common to all of the measures is influenced by age, and its effects on all of the measures are reduced when the variance in any of the measures is controlled.

The absence of a significant Age × Problem Type interaction in the data from the adult sample raises the possibility that, at least for these subjects, there may not have been a separate and distinct process responsible for the borrowing operation. This issue can be addressed by considering the ratio of the variance shared between the no-borrow and borrow measures to the amount of systematic (reliable) variance in the borrow measure. The correlation between the no-borrow and borrow response time measures was .86, indicating that 74% of the total variance in each variable was shared with the other variable. After correction for attenuation due to unreliability, the correlation was .92, indicating that 85% of the systematic variance in each variable was shared with the other variable. Because the proportion of systematic independent variance in the borrow response time measure is not very large, it is not clear how the absence of an interaction in this case should be interpreted. The relationship between the two variables is not perfect but is clearly substantial, and hence it is not obvious whether the amount of independent variance is sufficient to allow relations with other variables to be detected.

Discussion

One interesting finding from Study 1 is that different results were obtained in the sample of nonstudent adults from a wide age range and in the contrast of students and older adults. That is, the Age × Problem Type interaction was significant in the student–older adult contrast but not in the sample of nonstudent adults. This discrepancy raises questions about the practice of using college students as the young adult sample in age-comparative research, particularly if the tasks of interest may involve abilities related to academic success.

Although the interaction was significant in one comparison and not significant in the other, in neither case were the results of Geary et al. (1993) replicated because those researchers found that older adults had smaller borrow–no-borrow difference scores than young adults. There are a number of procedural differences between this study and that of Geary et al. that might be responsible for the different outcomes. For example, vocal production responses were used in the Geary et al. study, whereas manual verification responses were used in this study. Sample differences may also be contributing to the discrepancy because the older adults in the Geary et al. study were very select, and possibly unrepresentative, whereas the students in this study attended a university with rigorous science and mathematics entrance requirements. However, inspection of Table 1 reveals that the groups of nonstudent adults were fairly similar with respect to amount of education and yet there is no trend in Figure 2 for the borrow–no borrow difference scores to be smaller for older adults, as reported by Geary et al. (1993).

In terms of Figure 1, the results of this study appear more consistent with the interpretations represented in Figures 1B and 1C than the traditional interpretation represented in Figure 1A. That is, regardless of whether the Age × Problem Type interaction was significant in a particular comparison, the hierarchical regression analyses revealed that there was no independent or unique age-related variance in the borrow response time measure after the variance in the no-borrow response time measure was controlled. Moreover, the age-related variance in both measures was greatly reduced after the variance in the other speed measures was controlled. All of these results are consistent with the hypothesis that a construct related to processing speed functions as variable x in the representation portrayed in Figure 1B.

The interpretation represented in Figure 1C cannot be rejected because of the high correlations (i.e., .86 in the adult sample and .89 in the student–older adult comparison) between the no-borrow and borrow response time measures. Some ambiguity therefore exists with respect to whether borrowing represents a separate process that is not slowed with age, or whether the no-borrow and borrow response time measures are affected by the same process(es) in a nearly equivalent manner.

Study 2

We designed the second study to examine age-related effects on mental arithmetic with more powerful manipulations of the processing requirements and to determine the relation of working memory to the age differences in mental arithmetic. Samples of young (college students) and older adults performed between zero and seven arithmetic operations in sequence or according to the grouping specified by parentheses and brackets. Problems of the second type were labeled *hierarchical* because when parentheses or brackets are present, intermediate products need to be temporarily preserved during the solution of the problem, and hence the solution process is hierarchical rather than strictly linear. We hypothesized that increased age would be associated with slower operation speed (i.e., the age effects would be larger when more operations were required) and with slower responses when the demands on working memory increased as a result of the parsing requirement (i.e., the age effects would be larger with hierarchical than with sequential problems). All research participants also performed two tasks designed to measure working memory to allow an assessment of the influence of working memory on mental arithmetic performance.

Method

Subjects. A total of 40 older adults and 40 young adults participated in a single session, which lasted between 1.5 to 3 hr. Demographic characteristics of the two groups are summarized in Table 1.

Procedure. We tested all research participants in groups of one to four with paper-and-pencil tests administered first, followed by tests administered on computers. The order of the paper-and-pencil tests was boxes, pattern comparison, letter comparison, and digit copy. The order of the tasks administered on the computers was digit digit, digit symbol, reading span, computation span, and mental arithmetic. The paper-and-pencil and computer speed tasks were identical to those described in Study 1.

Mental arithmetic tasks. Both sequential and hierarchical mental arithmetic problems were presented on a single line of the display monitor in a verification format. Examples of each type of problem with five operations are as follows:

Sequential $5 + 3 - 1 - 3 + 4 - 1 = 6$ (FALSE)

and

Hierarchical $[(5 + 3) - 1] - [3 + (4 - 1)] = 1$ (TRUE).

The following criteria were used in the construction of the arithmetic problems: The problems contained only the digits 1–9; there were no negative products for any operation; the value of the outcome for any operation never exceeded 9; and when the solution was incorrect, it differed from the correct solution by plus or minus 1.

A practice block of 16 trials and two blocks of 64 experimental trials each were presented for both types of problems in the same counterbalanced order (i.e., sequential, hierarchical, hierarchical, and sequential) for all subjects. Eight trials were presented with each number of operations from zero to seven in each block. (Problems with zero arithmetic operations simply consisted of a display of a digit on each side of an equals sign.) One half of the trials with each number of operations in each block were TRUE (or CORRECT), and one half were FALSE (or INCORRECT).

In the sequential task, subjects were instructed to perform the computations in the order in which they appeared, that is, operations were to be performed in left-to-right sequence. For the hierarchical arithmetic task, subjects received detailed instructions of the conventions of performing operations on arithmetic problems containing parentheses and brackets. That is, work from left to right, perform the operations inside the parentheses before adding or subtracting terms outside the parentheses, and perform the operations inside the brackets before adding or subtracting terms outside the brackets.[1] With both types of problems, subjects were instructed to respond as rapidly and accurately as possible.

Working memory. The working memory tasks used in this study were very similar to those described by Salthouse (1992). The reading span working memory task involved the presentation of simple sentences with the participant instructed to select an answer to a question about the sentence, from a set of three alternatives, while also remembering the last word in each sentence. Answers to the questions were selected by using the up or down arrow keys on the keyboard to move a pointer to the correct alternative. On completion of the designated number of sentences, the word RECALL appeared on the screen. Subjects then typed the last word of each sentence in the order in which the sentences appeared. The number of sentences presented on each trial increased from one to nine, with three trials at each series length. The program continued as long as the subject was correct on both processing (answering the questions) and recall (reporting the last words) on at least two of the three trials at each series length.

The computation span working memory task involved a series of arithmetic problems with participants instructed to solve the problem while also remembering the last digit from each problem. After the presentation of each problem, the participant selected the correct answer from a set of three alternatives. On completion of the designated number of problems, the word RECALL appeared on the screen. At this point, the participant typed the last number in each problem in the order in which the problems appeared. The number of arithmetic problems presented on each trial increased from one to nine, with three trials at each series length. The program continued as long as the subject was correct on both processing (answering the arithmetic problems) and recall (reporting the last digits) on at least two of the three trials at each series length. In both tasks, the subject's working memory span was defined as the largest number of items in which he or she was correct on both processing and recall in at least two of the three trials.

Results

Performance on the nonarithmetic speed tasks and the working memory tasks is summarized in Table 1. As expected, the young adults were significantly ($ts > 6.0$, $p < .01$) faster than the older adults on all speed measures except for boxes and also had significantly higher scores on the working memory tests ($ts > 3.60$, $p < .01$). The measures were converted to z scores and averaged to form composite motor speed (boxes and digit copy, $r = .63$), perceptual speed (letter comparison and pattern comparison, $r = .61$), and working memory (reading span and computation span, $r = .44$) indexes.

[1] Although we cannot be certain that the subjects always performed the operations in the prescribed order and did not simply ignore the brackets and parentheses, we have two reasons for believing that most of them followed the instructions. First, many subjects complained about the difficulty of this task, and performance was substantially slower and less accurate than in the version of the task without brackets and parentheses (cf. Figure 3). Second, the pattern of results for the hierarchical problems in which an incorrect answer would have been produced if the brackets and parentheses were ignored (as in the sample problem in the text) was very similar to that for all hierarchical problems.

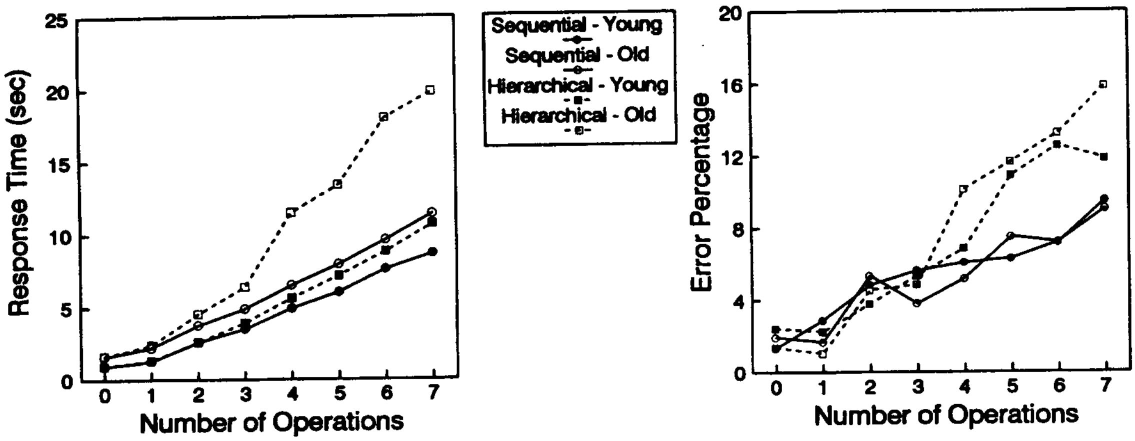

Figure 3. Mean response time and error percentage for young and old adults for zero to seven operations in sequential and hierarchical arithmetic problems, Study 2.

Means of the median response times and error percentages as a function of number of operations and problem type are illustrated in Figure 3. Results of the Age (young vs. old) × Problem Type (sequential vs. hierarchical) × Number of Operations (one to seven) ANOVAs are summarized in Table 3. It can be seen that neither the age main effect nor any interactions involving age were significant on the error measure, but that all effects were significant with the response time measure. These results thus indicate that increased age was associated with a slower operation time (from the interaction with number of operations) and with more time to perform parsing operations (from the interaction with problem type).

The proportions of age-related variance in the response time measures after control of one of the other variables derived from the hierarchical regression analyses are illustrated in Figure 4. That is, the values in this figure represent either the R^2 associated with age or the increment in R^2 associated with age after the designated variable had been controlled. (Analyses were also conducted with error percentage as another controlled variable but because this led to an average change in the age-related variance of only 1.1%, the error variable is ignored in the analyses reported here.) It is apparent in Figure 4 that the overall pattern was very similar for both problem types. Specifically, the amount of age-related variance (a) is nearly constant, or actually decreases, with additional operations; (b) is reduced only slightly after control of the working memory or motor speed indexes; (c) is reduced substantially after control of the perceptual speed, digit digit, or digit symbol measures; and (d) is completely eliminated after control of the speed of making physical identity judgments involving no arithmetic operations. Because there was no significant age-related variance with any number of operations after the time in the physical identity (zero operation) condition is controlled, it can be inferred that the age-related influences in the measures with one to seven operations are not independent of, or distinct from, the age-related influences on the physical identity measure. In other words, there are no age-related effects in speed of performing one to seven arithmetic operations in either the sequential or the hierarchical task that are independent of the age-related effects in the speed of performing physical identity decisions.

In an attempt to derive a more sensitive measure of arithmetic operation speed, linear regression equations were computed for each subject relating response time in the sequential problems to the number of operations between one and seven. The results summarized in the top portion of Table 4 indicate that there was a good fit of the equations to the data of individual subjects (as revealed by the high mean r^2 values), and that older adults had significantly larger slopes and intercepts than young adults. The bottom portion of Table 4 contains the results of the hierarchical regression analyses in which other speed measures were controlled in separate regression equations. It can be seen that the pattern is very similar to that with the other response time measures because the age-related variance was either greatly reduced or completely eliminated when other speed measures were controlled. In particular, there was no significant residual age-related variance in either the intercept measure or the slope measure when the variance in the identity measure was controlled. These results imply that there was no age-related influence on the speed of arithmetic operations (slope) or

Table 3
Analysis of Variance Results From Study 2

		Time		Errors	
Effect	*df*	*F*	MS_e	*F*	MS_e
Age	1, 78	80.98*	40.15	0.18	206.45
Problem Type	1, 78	116.27*	15.32	25.84*	59.34
Age × Problem Type	1, 78	54.16*	15.32	2.65	59.34
Number of Operations	6, 468	699.82*	3.67	42.63*	43.57
Age × Number Operations	6, 468	42.16*	3.67	1.21	43.57
Problem Type × Number Operations	6, 468	64.88*	2.68	8.36*	35.10
Age × Problem Type × Number Operations	6, 468	29.72*	2.68	1.15	35.10

*$p < .01$.

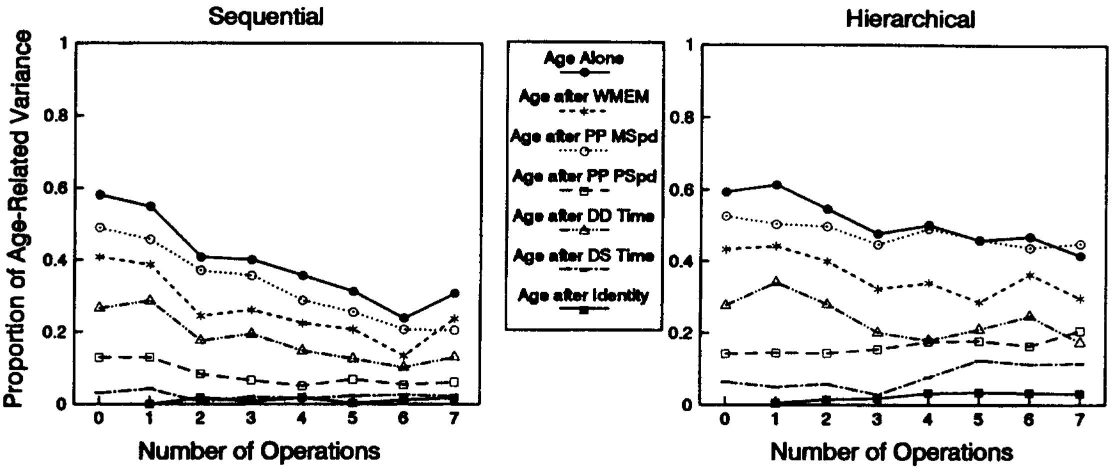

Figure 4. Proportion of age-related variance in response time with varying numbers of operations before and after control of other variables, Study 2. WMEM is a composite of the reading span and computation span measures, MSpd is a composite of the boxes and digit copying measures, and PSpd is a composite of the letter comparison and pattern comparison measures. DD Time and DS Time are median times in the digit digit and digit symbol tasks, respectively, and Identity refers to the median time in the physical identity (zero operation) arithmetic conditions. PP = paper and pencil.

other relevant processes (intercept) that was independent of the age-related variance in the speed of making physical identity judgments.

Because other studies (Salthouse, 1991, 1992; Salthouse & Babcock, 1991) have found that statistical control of various measures of speed greatly attenuated the age-related variance in working memory, hierarchical regression analyses were conducted with the composite working memory measure as the criterion variable. The proportion of age-related variance when age was the only predictor was .251, and the increments in variance associated with age were .241 after control of motor speed, .084 after control of perceptual speed, .103 after control of the digit digit measure, and .050 after control of the digit symbol measure. The proportion of variance associated with age after control of a composite index formed by averaging the *z* scores from the digit digit and digit symbol measures was .015. The values of .251 and .015 are similar to the values of .279 and .081 and .146 and .014 in the two studies with measures from nearly identical tasks reported by Salthouse (1992).

Table 4
Analyses of the Regression Parameters Relating Response Time to Number of Operations (1–7) in Sequential Arithmetic, Study 2

Measure	Intercept		Slope		r^2	
	Regression parameters					
	M	*SD*	*M*	*SD*	*M*	*SD*
Young	−.02	.56	1.24	.33	.95	.13
Old	.57	.72	1.51	.29	.97	.03
$t(78)$	−4.12*		−3.94*		−0.92	

Proportion of age-related variance		
Analysis	Intercept	Slope
Age alone	.193*	.169*
After working memory	.104*	.124*
After P & P MSpd	.223*	.108*
After P & P PSpd	.034	.036
After digit digit	.106*	.064
After digit symbol	.002	.021
After identity (zero operations)	.001	.015

Note. r^2 = square of correlation coefficient; P & P MSpd is the paper-and-pencil motor speed index (i.e., boxes and digit copy); P & P PSpd is the paper-and-pencil perceptual speed index (letter comparison and pattern comparison.).
*$p < .01$.

Discussion

The major results of this study were the significant interactions between age and number of operations and between age and problem type on the response time measures, together with the small to nonexistent age-related variance in these measures after statistical control of other measures of processing speed. For example, Figure 3 indicates that older adults were substantially slower than young adults on the hierarchical problems, and particularly as the number of operations increased above three. However, Figure 4 reveals that the age-related variance in the hierarchical response time measure was greatly reduced after statistical control of the digit symbol measure and completely eliminated after control of the physical identity measure. As in Study 1, this pattern of results is most consistent with the interpretation represented in Figure 1B.

Another interesting result from this study is the discovery that statistical control of the working memory measure led to relatively modest attenuation of the age-related variance in the arithmetic response time measures. The apparent implication is that working memory, at least as measured by the computation span and reading span tasks, contributes relatively little to the age-related influences in at least some measures of mental arithmetic performance.

General Discussion

A major goal of this article was to consider the type of evidence sufficient to warrant a conclusion that the group-related influences in one variable were separate and distinct from those in another variable. An important distinction was drawn between influences that vary in absolute magnitude and influences that are statistically independent of one another. We argued that an inference of truly differential or selective group-related effects should not simply be based on whether the group differences are greater in one variable than in another, but instead requires that at least some of the group-related influences in one variable are independent of those in another variable. Without evidence of distinct or unique influences associated with group membership, the two variables could both be affected by a common factor that is itself related to the group classification (as portrayed in Figure 1B).

Hierarchical regression analyses were proposed as one means of investigating the independence of group-related influences in two or more variables. The reasoning was that if there was little or no residual group-related variance in one variable after the variance in another variable was controlled, then one could infer that the group-related influences on the variables were not independent. Because other variables might also be influenced by the hypothesized common factor, the same type of statistical control procedures could be applied with variables derived from different types of tasks.[2] As with variables from the tasks of primary interest, a finding of little or no unique group-related variance in the critical variables after control of other variables would lead to a conclusion that the group-related influences on the variables were not independent.

The issue of independent variance is also relevant to the interpretation of nonsignificant Group × Variable interactions because differential influences cannot be expected if the variables have little or no independent variance. Although it is not clear exactly how much independent variance is sufficient to justify a conclusion that differential group-related influences could have been detected, confidence in the conclusion of equivalent influences should probably decrease as the ratio of shared to reliable variance increases. In other words, the smaller the proportion of unique systematic variance in a variable, the lower the confidence should be that the variable represents a construct distinct from those contributing to the other variables under examination.

It should be emphasized that the proposed procedures are designed to supplement rather than replace existing procedures. That is, traditional analysis of variance or regression methods are still needed to evaluate whether the differences between groups are larger in some variables than in others. Moreover, these procedures do not directly address the issue of the most meaningful scale (e.g., absolute differences or ratios) for interpreting interactions when groups differ in overall level of performance (Cerella, 1990; Salthouse, 1985). Instead the procedures are intended to provide a means of evaluating the independence of group-related influences and distinguishing between the alternative interpretations represented in Figure 1.

We applied the proposed analytical techniques to the data from two studies on mental arithmetic. Although Age × Variable interactions were significant in some comparisons in both studies, there was no significant age-related variance in the critical variables when the variance in other variables was controlled. That is, there was no significant age-related variance in the borrow response time measure in the student–older adult contrast in Study 1 after control of the no-borrow response time measure, and there was no significant age-related variance in any of the arithmetic response time measures in Study 2 after control of the physical identity response time measure. This pattern of results is therefore most consistent with an interpretation such as that represented in Figure 1B in which a common factor contributes to the age differences in each variable. Moreover, because statistical control of speed measures from other tasks also greatly reduced the age-related variance in the arithmetic measures, the data are compatible with the view that the hypothesized common factor represents a construct related to speed of processing.

The results of these studies also raise questions about the independence of the relevant variables, and by implication, of the distinctiveness of the hypothesized processes. Although the results portrayed in Figure 2 clearly indicate that there were substantial differences in the time and accuracy measures for problems with and without borrowing requirements, the correlation between the response time measures in the two conditions was .86, and .92 after adjustment for unreliability. This indicates that between about 74% and 85% of the total variance in each variable was shared with the other variable, leaving the remaining 15% to 26% of the variance to be partitioned into unsystematic and unique components. Even when the unsystematic portion is very low because of high reliability, one can question whether the amount of independent systematic variance was large enough to allow relations with other variables, such as age, to be detected. At the very least, therefore, one needs to be cautious in the interpretation of nonsignificant interactions, regardless of the apparent power of the comparisons.

On the basis of the results of these studies, then, it appears that increased age is associated with slower overall performance in mental arithmetic tasks, and with longer times for borrowing operations (i.e., the difference score measure in Study 1), and for basic arithmetic operations (i.e., the slope measure in Study 2). Measures of other hypothesized components, such as access to long-term memory, were not examined in these studies, and thus the present results are not directly relevant to the claims of little or no age differences in these components (e.g., Allen et al., 1992; Geary et al., 1993; Geary & Wiley, 1991). Nevertheless, consideration of the proposed guidelines suggests that some caution should be exerted in interpreting earlier results. One concern is that the statistical comparisons may not have been very powerful, either because of relatively small sample sizes or because of low reliabilities for the relevant measures. A second concern is that even if the measures were theoretically distinct, they may not have been

[2] In fact, the use of variables from different tasks has the advantage of providing converging information about the nature of the hypothesized common factor.

functionally independent. That is, possibly because of a shared influence of a common factor, the measures may have little or no unique variance. To the extent that this was the case, then it may be more plausible to view the measures as reflections of the same construct rather than of separate and distinct constructs.[3]

To summarize, there are both methodological and substantive conclusions of the research reported here. The primary methodological conclusion is that statistical interactions involving an individual difference variable such as age may be of less value than previously assumed for the purpose of identifying processes that have selective or differential effects. That is, an issue that needs to be considered is whether the age-related effects in one variable are merely larger than, or whether they are also independent of, the age-related effects in other variables. Because variables can differ with respect to the absolute magnitude of their relations with age and yet have all of their age-related influences mediated by a common factor (Figure 1B), we suggest that evidence of the independence of age-related influences is needed before a strong inference of truly selective or distinct age-related effects is justified. Furthermore, an inference of equivalent age-related effects may not be warranted unless there is evidence that the variables each have at least a moderate amount of distinct or independent variance.

The major substantive conclusions from this research are (a) that there are large age-related effects in several measures of mental arithmetic performance but (b) that those effects are apparently not independent of the age-related influences evident in other speeded tasks. We examined only time and accuracy measures in these studies, and thus it is possible that other aspects of performance, such as type or efficiency of strategy use, may have independent relations with age. On the basis of the available data, however, the age differences found in the current measures of mental arithmetic appear to be largely mediated by age-related reductions in a construct related to speed of processing.

[3] It is conceivable that the independence of two measures could differ across the groups being compared, perhaps as a result of similar experience or development. This possibility, which would raise questions about measurement equivalence and construct validity in the two groups, could be examined by determining whether there are significant group differences in the correlations between variables after adjusting for measurement unreliability.

References

Allen, P. A., Ashcraft, M. H., & Weber, T. A. (1992). On mental multiplication and age. *Psychology and Aging, 7,* 536–545.

Baron, J., & Treiman, R. (1980). Some problems in the study of differences in cognitive processes. *Memory & Cognition, 8,* 313–321.

Birren, J. E., Allen, B. S., & Landau, H. G. (1954). The relation of problem length in simple addition to time required, probability of success, and age. *Journal of Gerontology, 9,* 150–161.

Birren, J. E., & Botwinick, J. (1951). Rate of addition as a function of difficulty and age. *Psychometrica, 16,* 219–232.

Bogartz, R. S. (1976). On the meaning of statistical interactions. *Journal of Experimental Child Psychology, 22,* 178–183.

Burke, D. M., White, H., & Diaz, D. L. (1987). Semantic priming in young and older adults: Evidence for age constancy in automatic and attentional processes. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 79–88.

Cerella, J. (1990). Aging and information processing rate. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (3rd ed., pp. 201–221). San Diego, CA: Academic Press.

Chapman, L. J., & Chapman, J. P. (1973). Problems in the measurement of cognitive deficit. *Psychological Bulletin, 79,* 380–385.

Chapman, L. J., & Chapman, J. P. (1988). Artifactual and genuine relationships of lateral difference scores to overall accuracy in studies of laterality. *Psychological Bulletin, 104,* 127–136.

Charness, N., & Campbell, J. I. D. (1988). Acquiring skill at mental calculation in adulthood: A task decomposition. *Journal of Experimental Psychology: General, 117,* 115–129.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Geary, D. C., Frensch, P. A., & Wiley, J. G. (1993). Simple and complex subtraction: Strategy choice and speed-of-processing differences in young and elderly adults. *Psychology and Aging, 8,* 242–256.

Geary, D. C., & Wiley, J. G. (1991). Cognitive addition: Strategy choice and speed-of-processing differences in young and elderly adults. *Psychology and Aging, 6,* 474–483.

Hartley, A. A., Kieley, J. M., & Slabach, E. H. (1990). Age differences and similarities in the effects of cues and prompts. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 523–538.

Howard, D. V., Shaw, R. S., & Heisey, J. G. (1986). Aging and the time course of semantic activation. *Journal of Gerontology, 41,* 195–203.

Kausler, D. H. (1982). *Experimental psychology and human aging.* New York: Wiley.

Kraemer, H. C., & Thiemann, S. (1987). *How many subjects? Statistical power analysis in research.* Newbury Park, CA: Sage.

Lipsey, M. W. (1990). *Design sensitivity: Statistical power for experimental research.* Newbury Park, CA: Sage.

Loftus, G. R. (1978). On the interpretation of interactions. *Memory & Cognition, 6,* 312–319.

Madden, D. J. (1986). Adult age differences in the attentional capacity demands of visual search. *Cognitive Development, 2,* 201–216.

Rogers, W. A., & Fisk, A. D. (1991). Age-related differences in the maintenance and modification of automatic processes: Arithmetic Stroop interference. *Human Factors, 33,* 45–56.

Salthouse, T. A. (1985). Speed of behavior and its implications for cognition. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (2nd ed., pp. 400–426). New York: Van Nostrand Reinhold.

Salthouse, T. A. (1991). *Theoretical perspectives on cognitive aging.* Hillsdale, NJ: Erlbaum.

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

Salthouse, T. A. (1993). Attentional blocks are not responsible for age-related slowing. *Journals of Gerontology: Psychological Sciences, 48,* 245–255.

Salthouse, T. A. (in press). Aging associations: Influence of speed on adult age differences in associative learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition.*

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

Salthouse, T. A., & Kausler, D. H. (1985). Memory methodology in maturity. In C. J. Brainerd & M. Pressley (Eds.), *Basic processes in memory development* (pp. 279–311). New York: Springer-Verlag.

Salthouse, T. A., & Kersten, A. W. (1993). Decomposing adult age differences in symbol arithmetic. *Memory & Cognition, 21,* 699–710.

Received July 24, 1993
Revision received November 6, 1993
Accepted November 7, 1993 ■