Presenting and Analyzing Results in Aging Research: A Methodological Note

Moshe Naveh-Benjamin
Ben-Gurion University of the Negev, Israel

Fergus I. M. Craik
University of Toronto, Canada

Salthouse (1991) suggested two methods of presenting and analyzing the data of studies in which patterns of performance for young and old participants in different tasks are compared. The first method presents the original scores in terms of the standard deviations of the distribution of the young adults group, and in the second method, the reference distribution is performance of each age group in a control or baseline condition. The authors question the arbitrariness of the baseline chosen in both methods, which might lead to misleading results, and reanalyze the results of 2 studies and show that alternative, equally legitimate methods, using different baseline conditions, yield different results. This raises a question regarding the suitability of a derived-scores methodology in presenting and analyzing summary results.

Any scientific theory of cognitive aging relies on empirical observations made on groups of young and old adults. Two major questions are important in the context of cognitive aging, and aging in general: First, what is the magnitude of the age differences observed in a particular task? Second, what is the pattern that characterizes age differences in different tasks? The first question addresses the issue of age as a main effect, and the issues raised are whether there is any age trend in a particular task; if so, whether performance improves or worsens with advancing age, and finally, how large the trend is. The second question deals with the interaction of age with conditions in different tasks. This question concerns the pattern of performance of different age groups in at least two different conditions, and the issue of whether age differences are larger or smaller in one condition in comparison to others. Studies related to the first

Received 11 June 1996; accepted 12 April 1997.

This research was supported by the Canada–Israel Foundation for Academic Exchanges Fellowship to Moshe Naveh-Benjamin and by a grant from the Natural Sciences and Engineering Research Council of Canada to Fergus I. M. Craik.

Address correspondence to Moshe Naveh-Benjamin, Department of Behavioral Sciences, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel. E-mail: Naveh@bgumail.bgu.ac.il.

Experimental Aging Research, 24:83–98, 1998

0361-073X/98 $12.00 + .00

question were characteristic of early research on cognitive aging, whereas contemporary research is focusing increasingly on the second question (Salthouse, 1991).

One important question concerns the best way to present data to compare performance of young and old adults. The traditional way of presenting results, either in a table or in a figure format, is to provide means for the original (raw) scores. In such a case, performance of two or more age groups is tabulated for either one or several conditions. A trend for a main effect of age is inferred when there is enough separation, in units of standard error, between the performance levels of young and old adults, whereas a trend for an interaction effect is inferred when the differences in performance of young and old adults depend on the type of condition evaluated; that is, when one condition shows little or no difference between young and old, whereas the other shows large differences.

Salthouse (1991) discussed several reservations about the reliance on the aforementioned methodology of Age × Condition interactions in analytical approaches to cognitive aging. He stresses the fact that the interpretation of interactions as a way of inferring differential deficits is complex and cannot be used without consideration of several issues. Among these are the reliability and discriminating power of the dependent measures, the justification of assumptions concerning the particular form of the function relating theoretical processes and observed measures, and the meaningfulness of absolute versus distribution-referenced measures. Specifically, in his discussion of the problem of reducing the ambiguity inherent in the arbitrary scaling of the dependent variables, Salthouse made the point (p. 296) that only cross-over interactions, in which the direction of the age differences reverses from one performance variable to the other, are unambiguous (as long as linear transformations are considered). All other interactions presented, using absolute scores, are problematic and depend on specific assumptions about the process-variable relations. Moreover, such interactions can be eliminated by monotonic transformations of the measurement scale (e.g., Baron & Treiman, 1980; Bogartz, 1976; Loftus, 1978). Because few of the performance measures in cognitive psychological research are meaningful in an absolute sense, Salthouse noted the benefit of expressing performance in units of population variability when presenting the data, rather than in the original units of measurement (p. 297).

Along this line, Salthouse (1991) suggested two alternative ways to present and analyze the data, both of which involve derived scaling using linear transformation. His suggestion follows research utilizing the traditional psychometric approach to aging (e.g., Jones, 1959). One way is to replot and present the original scores in terms of standard deviations

from the distribution of scores of the young adults group. Salthouse claimed that

> this particular method of representing the scores allows direct comparisons of the performance at any given age relative to the performance of young adults because the units are immediately interpretable in terms of the distribution of scores in this standard, or comparison group. (p. 41)

In the second method that he suggested, the reference distribution is performance of each age group in a control or baseline condition: performance in each experimental condition is then expressed in terms of the standard deviations of the control-group performance and derived scores are compared (this method can be used when a control condition is used and it requires caution when comparisons are made between studies that use different control conditions). Salthouse then used both methods (particularly the first one), to replot the data of several studies, and presented these replotted data in figure form. Under such presentation, the mean of each age group is treated as a score in the reference-group distribution, as it is plotted in standard deviation units of the reference group. He then used these figures to draw conclusions regarding the relative degree of age-sensitivity of different tasks.

We believe that Salthouse's (1991) suggestions regarding transformations are problematic and might, on occasion, be misleading. Although the practice of using standardized scores is well accepted and allows the comparison of individuals on different tasks, the variant suggested here may be inappropriate. The major problem that we see with these methods is their arbitrariness in choosing one group or condition (e.g., the young adults) as the reference distribution. Though such a procedure fits Salthouse's theoretical views regarding cognitive slowing as the underlying agent in old people's performance, and hence the use of the young adults as the reference group, such a presentation might affect the interpretation of the results—both the main effect of age and the interaction of age and other variables.

The problem with such a presentation is that it does not take into consideration the full variability in each age population, only that obtained in the young group. There is no a priori reason why young adults should provide the reference group rather than the old adults. Likewise, for the second method suggested by Salthouse (1991), the presentation and the analysis of the data for each age group, in terms of the variability of the control or baseline condition, might be problematic for the same reason: It uses only part of the variability in the data to plot the results. We would like to suggest that other different derived scores can be extracted that could lead to different results, and because there is no a priori reason to prefer one derived-scores methodology over the other, researchers should be careful in basing their conclusions on a specific method and on derived-scores methodology.

To make our points concrete, we present results from two studies used by Salthouse (1991), in which he applied his suggested methods. We then present these same results in alternative ways that are equally legitimate. The intention of these examples is to show that these different modes of presentation could lead to different, sometimes opposite, interpretations of the results, both in terms of evaluating the main effect of age and the interaction of age with other variables. Our analysis addresses the question of the way to present the data in a table or figure form and its resulting statistical inferences.

STUDY 1: DIVIDED ATTENTION AT STUDY AND TEST (PARK, SMITH, DUDLEY, & LAFRONZA, 1989)

This study, involving two experiments, evaluates the degree to which memory performance of verbal material is affected by a secondary task (digit monitoring) either at study or at test. Both young and old groups of participants were used. Summary results for each of the two experiments, where memory performance for each group is tabulated as a function of the presence or absence of a secondary task at study or test, appear in Table 1. The original data that appear in Line 1 of each experiment indicate effects of secondary task at study (presentation), which are on average larger in the old group (average differences between young and old in the Yes/No and Yes/Yes conditions vs. average differences between young and old in the No/No and No/Yes conditions). No such differential effects of secondary task on age were observed at test (average differences between young and old in the No/Yes and Yes/Yes conditions vs. average differences between young and old in the No/No and Yes/No conditions).

From these results, Park et al. (1989) concluded that the effects of a secondary task are more demanding for old than for young people in the encoding (study) phase but not at retrieval (test). They suggested that these data do not support the notion that older adults suffer from a general deficiency in processing information.

Salthouse (1991, pp. 298–299) reanalyzed the results using his two suggested measures. In Line 2 of each experiment (Table 1), the data are presented for old adults in terms of young adults' standard deviation units; the values shown are (Old − Young)/Young standard deviations for each condition. Presented this way, it can be seen, especially in Experiment 2, that there were no differential effects of the secondary task in the different age groups on memory for the information presented, either at encoding or at retrieval.

Table 1. Alternative analyses of the Park, Smith, Dudley, and Lafronza (1989) experiments

Condition	Digit monitoring at presentation/test			
	No/No	No/Yes	Yes/No	Yes/Yes
	Experiment 1			
1. Young	22.06 (3.61)	15.50 (4.84)	13.69 (4.09)	10.81 (3.47)
Old	17.81 (6.61)	15.13 (2.92)	8.06 (5.09)	5.73 (3.77)
2. Old in young *SD* units	−1.18	−0.08	−1.38	−1.46
3. Young in old *SD* units	+0.64	+0.13	+1.11	+1.35
4. Old in pooled *SD* units of young and old	−0.79 (5.33)	−0.09 (4.00)	−1.22 (4.62)	−1.40 (3.62)
5. Young in young no/no *SD* units	0.00	−1.82	−2.32	−3.12
Old in old no/no *SD* units	0.00	−0.41	−1.48	−1.83
6. Young in young pooled no/no and relevant condition *SD* units	0.00	−1.54 (4.27)	−2.17 (3.86)	−3.17 (3.54)
Old in old pooled no/no and relevant condition *SD* units	0.00	−0.52 (5.11)	−1.65 (5.90)	−2.24 (5.38)
7. Young in young pooled *SD* units for all conditions	0.00	−1.61 (4.04)		
Old in old pooled *SD* units for all conditions	0.00	−1.99 (4.80)		

However, this conclusion is based only on part of the variability in the data—the young adults' standard deviations. If the results of the young are tabulated in standard deviations of the old adults (which is methodologically as legitimate), a different pattern of results emerges. In this case the values are calculated as (Young − Old)/Old standard deviations. The results, which can be seen in Line 3 of Table 1 for each experiment, indicate differential effects for young and old in secondary task effects at study (encoding). The results, which have a plus sign indicating the increase in performance of the young adults' mean performance relative to the old adults' mean performance in each condition, show clearly that young adults have an advantage over the old adults when a secondary task is introduced at study (+1.35 and +1.11 when a secondary task is introduced vs. +0.64 and +0.13 when it is not). The impact of the age group variability used as a baseline is especially noticeable in Experi-

Table 1. Alternative analyses of the Park, Smith, Dudley, and Lafronza (1989) experiments (*Continued*)

Condition	Digit monitoring at presentation/test			
	No/No	No/Yes	Yes/No	Yes/Yes
	Experiment 2			
1. Young	20.25 (2.73)	16.44 (3.10)	13.44 (4.66)	12.94 (4.91)
Old	15.62 (3.90)	12.31 (4.57)	7.00 (4.06)	5.81 (3.88)
2. Old in young *SD* units	−1.70	−1.33	−1.38	−1.45
3. Young in old *SD* units	+1.18	+0.90	+1.59	+1.83
4. Old in pooled *SD* units of young and old	−1.37 (3.37)	−1.06 (3.90)	−1.49 (4.31)	−1.61 (4.42)
5. Young in young no/no *SD* units	0.00	−1.40	−2.49	−2.68
Old in old no/no *SD* units	0.00	−0.85	−2.21	−2.52
6. Young in young pooled no/no and relevant condition *SD* units	0.00	−1.30 (2.92)	−1.78 (3.82)	−1.84 (3.97)
Old in old pooled no/no and relevant condition *SD* units	0.00	−0.78 (4.25)	−2.16 (3.98)	−2.52 (3.89)
7. Young in young pooled *SD* units for all conditions	0.00	−1.30 (3.96)		
Old in old pooled *SD* units for all conditions	0.00	−1.84 (4.11)		

Note. No/No = no secondary task either at study or at test; No/Yes = no secondary task at encoding and a secondary task at retrieval; Yes/No = secondary task at encoding and no secondary task at retrieval; Yes/Yes = secondary task both at encoding and at retrieval. The data in Experiment 1 in each of the four attention conditions is based on either 15 or 16 young or old participants. In Experiment 2, the data in each of the four attention conditions is based on 16 young and old participants. The numbers in parenthesis indicate standard deviations.

ment 2. When young adults' standard deviations are used, as Salthouse (1991) did, there are clearly no differential effects of secondary task on the young and the old. When old adults' standard deviations are used, however, there are clear differential effects of secondary task at study for young and old.

It can be claimed, and justifiably so, that our reanalysis and presentation of the results in terms of the older people's standard deviations might be as biased as Salthouse's presentation of the results in terms of the young's standard deviations, as both approaches do not take the full variability in the data into account. A more balanced approach is one in which the variability of both age groups is considered in the presentation of the results. This approach is based on the assumption underlining inferential test statistics—the use of pooled within-group diversity. Line 4 of Table 1 presents the performance of old adults relative to the young group in terms of pooled standard deviations of both groups for each condition (SD^2 young + SD^2 old) assuming independence of the two groups. In contrast to Salthouse's conclusions, the results indicate that old adults were affected more than the young adults when a secondary task was used at the study phase. Note that the trend obtained with the pooled standard deviations is very similar to the original raw data presented in Line 1.

The second method suggested by Salthouse (1991) to present and analyze the data is in terms of standard deviations of each group's performance in the control (No/No) condition. Results of the reanalyzed data using this method are presented for each age group and each experiment in Line 5 of Table 1. Here, as Salthouse pointed out, the effects of secondary task at study are larger for the young group in Experiment 1, and equal for the two groups in Experiment 2, in contrast to the original conclusion drawn by Park et al. (1989). As mentioned above, however, such a presentation could be misleading because it takes into consideration only part of the variability in the data; in this case, only the variability of the control condition (the No/No condition) in each age group.

A more balanced analysis is one in which the variability in performance under all conditions within each age group is considered in presenting the data. Line 6 for each experiment in Table 1 presents the performance of the old adults relative to the young adults when the pooled variances of each condition and the control condition for each age group separately are used. Such tabulation provides results that are in line with Salthouse's analysis for Experiment 1: The young adults show a greater decrease in performance when the secondary task is performed at study (the Yes/No and the Yes/Yes conditions). The picture changes in Experiment 2, where the old adults show a greater deficit when the secondary task is introduced at study. When all sources of variation are used for each age group separately (the pooled within age group variance of all conditions), not only control and a given group as in Line 6, the results again show a different trend than the one depicted by Salthouse's analysis: For both experiments, old adults show a larger overall decrease when a secondary task is introduced at study (Yes/No and Yes/Yes vs. No/No and No/Yes), relative to their performance in the control condition (Line 7, Table 1).

Note again, as with the first method, that the trend obtained with the pooled standard deviations is very similar to the original raw data presented in Line 1.

To summarize, presentation and the analysis of the data in the Park et al. (1989) study are affected by one's choice of the variability of a specific group as a baseline. Salthouse's use of either young adults' variability in his first suggested method, or control group variability in his second, represents one way of presenting the results. However, use of an equally legitimate baseline (the old adults' variability) or a pooled-variability option (of young and old in the first method, or of all conditions within each age group in the second), yielded different patterns of results for the differential effects of age. The pooled-variability alternative produces trends that are very similar to the original raw data.

STUDY 2: VISUAL PAIRED ASSOCIATES: IMMEDIATE VERSUS DELAYED (WECHSLER, 1987)

These data are part of the results reported in the standardization of the 1987 revision of the Wechsler Memory Scale. The data were collected for different age groups engaged in the visual paired-associates task, using either immediate or delayed testing. The original results (Wechsler, 1987, p. 52) are presented in Line 1 of Table 2. These results show a trend for faster forgetting in the young (20–24) than the old (70–74) participants. Salthouse (1991, p. 234) presented these results for the different age groups as a function of standard deviations of the young group (20–24 years old). His results, which are presented in Line 2 of Table 2, indicate equal forgetting from immediate to the delayed test for the young and the old participants. These trends are taken by Salthouse to support his conclusion that patterns of loss of information in storage do not differ for young and old adults.

However, this pattern of results is obtained when only one source of variability is considered—that in the young group. Quite a different picture emerges when the old people's variability and performance are used as a baseline. Results portraying the data as the advantage of younger over older participants in terms of the old group's standard deviation appear in Line 3 of Table 2. Note two differences between these results and those calculated by Salthouse (1991). First, in absolute terms, the age-related differences are smaller here than in Salthouse's analyses. The reason for the difference in magnitude of the main effect of age is the larger variability in older people's performance relative to the young. Second, results in Line 3 indicate that the young show relatively better performance in the immediate versus the delayed condition (+1.60 vs. +0.89). This pattern of results, indicating an interaction of age with retention interval, is different from that obtained by Salthouse which showed no interaction trends. The reason for these differences is the smaller ratio of standard deviations of the old adults in the immediate and the delayed conditions (4.3/1.9) relative to this ratio in the young adults (3.0/0.7).

As described earlier, another possible measure is one that takes into consideration the variability of both age groups. Line 4 of Table 2 presents the results of changes in all age groups in comparison to the young (20–24) adults, using pooled variability of the young and the old (70–74) adults. In contrast to the picture obtained by Salthouse, the results again indicate both smaller overall age effects and an interaction of age and retention interval, where the old participants show less forgetting over time. In this specific study, where we have more than two groups of participants, it might be more representative to present the results in terms of the pooled variability of all age groups (except the 16–17-year-old group, which is not included by Salthouse either). These results, which are presented in Line 5 of Table 2, indicate the same trend obtained by using the pooled variances of the young and old groups (Line 4): a smaller main effect of age emerges accompanied by an interaction of age and retention interval. As in Study 1, the trends obtained with the pooled variances is similar to the one provided in the original raw data (Line 1).

Finally, although for this study Salthouse does not present his second method of analysis in which each age group results are shown in units of standard deviation of the control condition, we reanalyzed the data using this method and the results appear in Lines 6–8 of Table 2. Line 6 presents each group's delayed-condition results when the immediate condition is considered as the control. We can see that the trends are fairly consistent with our previous analyses in terms of the standard deviations of the old, or the combined variances of young and old (Lines 3 and 4 in Table 2): The young show relatively poorer performance in the delayed condition. When the delayed condition is considered control, the same trend emerges for the immediate condition (Line 7, Table 2), although the size of the effect is much larger, probably due to the very small variability of the young adults in the immediate condition. A similar pattern is obtained in the delayed condition performance when the pooled variances of the immediate and delayed conditions are used (Line 8, Table 2).

Overall, as in Study 1, presenting and analyzing the data in terms of standard deviations of the young adults may lead to misleading results. A different pattern of results emerges when the old group variability is used, or when the variability of both the young and old adults is used. In this study, results of the two methods using the combined variability (Lines 4 and 8, Table 2) are compatible; they also yield the same conclusion as the analysis based on the old group variability (Line 3, Table 2).

Table 2. Alternative analyses of the Wechsler (1987) data: Visual paired-associates (immediate vs. delayed)

Condition	20–24	35–44	55–64	65–69	70–74
1. Visual paired-associates: Immediate	15.30 (3.00)	14.80 (3.70)	11.20 (4.30)	10.20 (4.20)	8.40 (4.30)
Visual paired-associates: Delayed	5.70 (0.70)	5.40 (1.40)	4.60 (1.80)	4.50 (1.70)	4.00 (1.90)
2. All groups in young *SD* units					
Immediate	0.00	−0.16	−1.37	−1.70	−2.30
Delayed	0.00	−0.42	−1.57	−1.71	−2.42
3. All groups in old *SD* units					
Immediate	+1.60	+1.40	+0.65	+0.42	0.00
Delayed	+0.89	+0.74	+0.32	+0.26	0.00
4. All groups in pooled *SD* of young and old					
Immediate (*SD* = 3.71)	0.00	−0.13	−1.10	−1.37	−1.86
Delayed (*SD* = 1.43)	0.00	−0.21	−0.77	−0.84	−1.19
5. All groups in pooled *SD* of all groups					
Immediate (*SD* = 3.93)	0.00	−0.13	−1.04	−1.29	−1.76
Delayed (*SD* = 1.56)	0.00	−0.19	−0.71	−0.77	−1.09
6. Young in young immediate *SD* units. Delayed	−3.20 (3.00)				
Old in old immediate *SD* units. Delayed					−1.02 (4.30)
7. Young in young delayed *SD* units. Immediate	+13.70 (0.70)				
Old in old delayed *SD* units. Immediate					+2.32 (1.90)
8. Young in young pooled immediate and delayed *SD* units. Delayed	−4.40 (2.18)				
Old in old pooled immediate and delayed *SD* units. Delayed					−1.33 (3.32)

The data in the 20–24, 35–44, 55–64, 65–69, and 70–74 years old age groups is based on 50, 54, 54, 55, and 50 participants, respectively. The numbers in parentheses indicate standard deviations.

Salthouse's analysis in terms of young adults' variability yielded a different conclusion.

DISCUSSION

Let us summarize the major points demonstrated in the above examples regarding presentation and the analysis of results. First, expressing the results in terms of one particular age group may be problematic. Results could just as easily be presented in terms of the other age group's standard deviations. As we saw, such presentation might yield different trends. For example, in Study 1 (especially in Experiment 2) and in Study 2, the use of old adults' variability in the analysis produced trends for an interaction of age and secondary task effects at study, or an interaction of age and retention interval, when none existed when the young's variability was used. In addition, in all cases the main effect of age was smaller when the old people's variability was used. The studies analyzed are not unique in this respect. Table 7 of Wechsler's (1987, p. 52) data indicates that for most measures the standard deviations of the old are larger than for the young. This will tend to decrease the effect of age when the old group's variability is used as the baseline.

The same applies to the second method of presenting the results suggested by Salthouse (1991): tabulating the results of each age group in terms of the standard deviation of its control group. There are two problems with this suggestion. First, determining which is the control condition in an experiment is not always easy; some experiments do not use obvious control groups. Second, this might result in a biased presentation, because the results might depend on the variability of a particular group and not on that of other groups. As indicated in the examples presented above, different results may emerge depending on the control group used. The trend for an interaction of age and retention interval in Study 2 depended on the baseline condition used. As is the case with presentation in terms of young or old variability, the pattern of results will depend on the patterns of variability of performance in the different conditions.

More specifically, Salthouse's (1991) claim, that cross-over interactions are unambiguous, is correct in the sense that the derived-scores methodology will not change the trends in the results regardless of whether the variability of the old or the young is used (though it may change the statistical magnitude of the effects). When more common, ordinal (non-cross-over) interactions are observed, the derived-scores methodology could lead to different results depending on the reference group used. In such cases, only when certain relationships exist between the means and the variances of the young and old in the different conditions will the same patterns be observed whether we use one group or the other as the reference for computing derived scores (see Appendix).

One rule of thumb is that as long as the ratio of the standard deviations in the two age populations is fairly similar under the different conditions, results will be similar regardless of which age-group variability (or control condition) is used as a reference. In this case, Salthouse's (1991) suggestion of using the young adults' distribution as a reference would not distort the results. However, once the ratio of the variances in the two age groups is reversed in one condition relative to the other (e.g., old adults show larger variability than the young in one condition and smaller in the other), the results could be different, depending on which age-group variability is used as a reference (for formal properties, see Appendix). In this case, the use of pooled variability of both age-groups for each condition seems a better solution.

Note that although our discussion refers to descriptive statistics, the same conclusions apply to inferential statistics; significant differences in one direction may be obtained when one group is used as the reference for computing derived scores, and in the other direction when the other group is used as the reference.

Presentation and analysis of the results in terms of pooled variances of young and old adults has the advantage of using all sources of variability in the data, as is actually done in an analysis of variance. This may also increase the reliability of the data as the pooled variance is based on larger *N*s. Such a presentation may also have the advantage, as it did in the studies analyzed here, of producing converging trends for the two derived scores (in terms of a given age group and a control group). Interestingly, in all cases, the pooled-variability estimates showed trends that were very similar to those revealed by the original raw data. This raises the question of whether we should resort at all to derived-scores methodology. Because different derived scores can be extracted, and because they could lead to different trends, depending among other things on the ratio of the variances in the different conditions and different age groups, the alternative of presenting and analyzing the original means and standard deviations may often provide a parsimonious solution.

Though the conclusions reached above are not in line with Salthouse's (1991) suggestions, we do agree with him that the interpretation of statistical Age × Condition interactions as a way of inferring differential deficits is a complex issue. A full methodological analysis should address questions regarding the reliability and discriminating power of the dependent measures, the characteristics of the function relating theoretical processes to observed measures, and considerations of absolute versus distribution-referenced measures. Our discussion underlines the point that reaching a consensus on these issues will require further interchange among researchers.

REFERENCES

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

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

Jones, H. E. (1959). Intelligence and problem solving. In J. E. Birren (Ed.), *Handbook of aging and the individual* (pp. 700–783). Chicago: University of Chicago Press.

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

Park, D. C., Smith, A. D., Dudley, W. N., & Lafronza, V. N. (1989). Effects of age and a divided attention task presented during encoding and retrieval on memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 1185–1191.

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

Wechsler, D. (1987). *Manual for the Wechsler Memory Scale—Revised*. New York: The Psychological Corporation.

APPENDIX

As mentioned in the text, for ordinal (non-cross-over) interactions, the same trends will be obtained whether we use the young or old participants group as the reference for computing derived scores, only when certain relationships between the pattern for the means and for the variances in the different age groups are observed.

Using the following notations: Y = young adults, O = old adults, A = Condition A, B = Condition B, M = group mean, SD = group standard deviation, when the relationships between the means and the standard deviations are:

$$\frac{M_{Y(B)} - M_{O(B)}}{M_{Y(A)} - M_{O(A)}} > \left|\, |SD_{Y(B)} - SD_{O(B)}| - |SD_{Y(A)} - SD_{O(A)}| \,\right| \quad (1)$$

the results will not change as a function of the reference group used. However, when the ratio of the means (left component of the equation) is equal to or smaller than the absolute standard deviations differences (right component of the equation), different patterns will be observed with derived scores, depending on the reference group used.

As an example of the former case, where same patterns are obtained in the derived scores, consider the following raw results and derived

scores: for Condition A, young adults' M = 18 and SD = 1 and old adults' M = 14 and SD = 3; for Condition B, young adults' M = 12 and SD = 1 and old adults' M = 4 and SD = 3. Derived scores for old in young SD are −4.00 in Condition A and −8.00 in Condition B, and young in old SD are +1.33 in Condition A and +2.67 in Condition B.

Using these values in Equation 1, we get

$$\therefore \frac{12 - 4}{18 - 14} > \left|\, |1 - 3| - |1 - 3| \,\right| \rightarrow \frac{8}{4} > |0| \rightarrow 2 > 0.$$

In this case, when the left side of the equation is larger than the right, the same trends emerge in the derived scores, as can be seen above—a worse performance of the old group (or better performance of the young group) in Condition B, regardless of the reference group used.

As an example of the latter case, where we get different patterns in the derived scores, depending on the reference group, consider the following results: for Condition A, young adults' M = 18 and SD = 1 and old adults' M = 14 and SD = 3; for Condition B, young adults' M = 12 and SD = 3 and old adults' M = 4 and SD = 1. Derived scores for old in young SD are −4.00 in Condition A and −2.67 in Condition B, young in old SD are +1.33 in Condition A and +8.00 in Condition B.

Inserting the values used in this example in Equation 1 produces the following:

$$\frac{12 - 4}{18 - 14} < \left|\, |3 - 1| - |1 - 3| \,\right| \therefore \frac{8}{4} < |4| \rightarrow 2 < 4.$$

In this case, when the left side of the equation is smaller than the right, it can be seen above that different patterns emerge in the derived scores, depending on the reference group. When young adults' SDs are used, old people show relatively poorer performance in Condition A. When old adults' SDs are used, old people show relatively poorer performance in Condition B. The same opposite trends occur in the performance of the young.

Finally, as mentioned in the text, when cross-over interactions are observed, using either the young or the old as the reference group to compute derived scores will not change the trends observed, though the specific values could be different. This is shown in the following example: for Condition A, young adults' M = 18 and SD = 1 and old adults' M = 20 and SD = 3; for Condition B, young adults' M = 12 and SD = 3 and old adults' M = 4 and SD = 1. Derived scores for old in young SD are +2.00 for Condition A and −2.67 for Condition B, and young in old *SD* are −0.66 for Condition A and +8.00 for Condition B.

In this case of cross-over interactions, the rule using Equation 1 does not hold. Using the values in this example will result in a smaller mean differences ratio than standard deviations differences (smaller left component) though the trends do not change as a function of the derived scores used. As can be seen above, these scores produce the same trend of performance, where old always perform worse in Condition B. The magnitude of this effect, though, is greater when old adults' *SD*s are used (young improve from −0.66 to +8.00, 8.66 units altogether), than when young adults' *SD*s are used (old worsen from +2.00 to −2.67, 4.67 units altogether).

Statement of ownership, management, and circulation (required by 39 U.S.C. 3685) of EXPERIMENTAL AGING RESEARCH published quarterly at Taylor & Francis Ltd., 1 Gunpowder Square, London EC4A 3DE, UK for September 24, 1997. General business offices of the publisher are located at Taylor & Francis, Inc., 1900 Frost Road, Suite 101, Bristol, PA 19007-1598. Name and address of publisher: Louis M. Fancher, 1101 Vermont Avenue, NW, Suite 200, Washington, D.C. 20005-3521. Name and address of editor: Jeffrey W. Elias, Department of Psychology, Texas Tech. University, P.O. Box 42051, Lubbock, TX 79409-2051. Name and address of managing editor: Louis Fancher, 1101 Vermont Avenue, NW, Suite 200, Washington, D.C. 20005-3521. The owner is Taylor & Francis Publishers, 1101 Vermont Avenue, NW, Suite 200, Washington, D.C. 20005-3521. Known bondholders, mortgagees, and other security holders owning or holding 1 percent or more of total amount of bonds, mortgages, or other securities: Taylor & Francis Group, Ltd., 1 Gunpowder Square, London, EC4A 3DE, UK. The average number of copies of each issue during the preceding 12 months is: (A) Total number of copies printed: 1000; (B) Paid circulation: (1) Sales through dealers and carriers, street vendors and counter sales: NONE; (2) Mail subscriptions: 300; (C) Total paid circulation: 300; (D) Free distribution by mail: NONE; (E) Free distribution outside the mail, carrier or other means: NONE; (F) Total free distribution: NONE; (G) Total distribution: 300; (H) Copies not distributed: (1) Office use, left over, spoiled: 700; (2) Return from News agents: NONE; (I) Total: 1000. The actual number of copies of single issue published nearest to filing date is: (A) Total number of copies printed: 1000; (B) Paid circulation: (1) Sales through dealers and carriers, street vendors, and counter sales: NONE; (2) Mail subscriptions: 305; (C) Total paid circulation: 305; (D) Free distribution by mail: NONE; (E) Free distribution outside the mail, carrier or other means: NONE; (F) Total free distribution: NONE; (G) Total distribution: 305; (H) Copies not distributed: (1) Office use, left over, spoiled: 695; (2) Return from News agents: NONE; (I) Total: 1000. I certify that the statements made by me above are correct and complete.

(signed) Louis M. Fancher
Executive Vice President, Washington, D.C.