

Author's response to commentaries

## Responses to commentaries by Finch, Nilsson et al., Abrams, and Schaie

Timothy A. Salthouse\*

*Department of Psychology, University of Virginia, 102 Gilmer Hall, P.O. Box 400400, Charlottesville, VA 22904-4400, United States*

Received 29 December 2008; accepted 1 January 2009

Available online 23 February 2009

I appreciate the interest of the commentators in my article, and I welcome the opportunity to respond to their comments.

I have admired Caleb Finch's work on the neurobiology of aging for many years, and therefore I was pleased to see that he supports my general conclusion that some cognitive changes likely begin in early adulthood. As he mentions in his commentary, he has long advocated research focusing on the entire period of adulthood instead of merely on groups at the extremes, and he briefly summarizes some of his excellent research documenting age differences in various aspects of neurobiology in his commentary. I also agree with his point that the next phase of research should address causal linkages between cognitive aging changes and changes in brain structure and metabolism. This will likely require communication and interaction among researchers who primarily focus on age-related changes in cognitive abilities and those who primarily focus on neurobiological measures of brain structure and function, and his commentary may signal the beginning of this type of endeavor.

Abrams raised three particularly relevant points in her commentary. In her first point she noted the existence of other cognitive variables in which cross-sectional differences are apparent in middle age. I suspect that this is likely true for most variables in which extreme group comparisons have revealed age differences because it seems implausible that age-related differences occur abruptly immediately before the age of the older groups.

The second point concerns the existence of variation in the results across variables. As I noted in the target article, the estimates of the retest effects are somewhat noisy, but it is nevertheless important to recognize that there is also considerable consistency in that all 12 variables had negative

cross-sectional age trends whereas 8 had positive longitudinal age trends, all 12 had positive short-term retest estimates, 10 of 12 had positive retest estimates from the twice-versus-once tested contrast, and 10 of 12 had positive retest estimates from the method based on variable retests. Finally, in comparisons of the retest estimate with the longitudinal change, the retest estimate was larger in 11 of 12 comparisons with short-term retests, larger in all 12 comparisons with retest estimates based on the twice-versus-once test, and larger in 8 of 12 comparisons with the estimate derived from the mixed effects model. Regardless of the amount of variability in the patterns of results, however, it is clear that better understanding is needed of the nature of cognitive change, including the relative contributions of retest and other influences.

Abrams' third point was that researchers need to be cautious in interpreting results based on the use of participant pools in which the same individuals serve in multiple studies. I particularly liked her phrase "inadvertent retesting", because I believe the influence of different types of testing experience on subsequent performance is an important unresolved question.

In their commentary Nilsson and colleagues note several disagreements with my target article, and in fact, based on the description of the article in their abstract and in other places in their commentary, I also disagree. For example, I am not comfortable with claims that longitudinal data should be "dismissed," and "should not be trusted because they are flawed," that "cross-sectional data are to be preferred," and that "cognitive function is a homogeneous and unitary entity."

Because these authors apparently found the article confusing, it may be useful to briefly restate the major arguments. The impetus for the article was the striking discrepancy between the age trends in some measures of cognitive functioning revealed from cross-sectional and longitudinal comparisons, particularly before age 60. As noted

\* Tel.: +1 434 982 6323.

E-mail address: [salthouse@virginia.edu](mailto:salthouse@virginia.edu).

in the article, the discrepancy is not apparent with measures of knowledge such as crystallized intellectual ability or semantic memory, and thus they were not considered relevant to the question of when cognitive decline begins. This does not mean that aspects of knowledge and information are not an important part of aging cognition, but the neglect merely reflects the fact that these aspects do not exhibit the phenomenon of primary interest in the target article.

Two major interpretations of the discrepancy have been proposed. One is that cross-sectional comparisons are confounded by cohort differences, such that people of different ages also differ in other respects that might contribute to different levels of performance. The second interpretation is that longitudinal comparisons are confounded by retest effects, such that performance on the second occasion reflects influences of age and of prior testing experience. Other possibilities could obviously be operating as well, but the focus in the target article was on these two alternatives, and whether they could be distinguished by determining if the age trends were altered by eliminating the critical confounding.

The argument with respect to the cohort interpretation was that if cohort is a meaningful concept, then it should be possible to quantify its characteristics and examine their influence. Specifically, people of different ages should vary in the relevant cohort-defining characteristics, and these characteristics should be related to measures of cognitive performance. Furthermore, a primary implication of the cohort interpretation is that if the variation in these characteristics were statistically controlled, the age differences in relevant measures of cognitive performance should be greatly reduced. Although seemingly straightforward, the problem with this strategy is that it has been difficult to identify and quantify presumably relevant cohort characteristics such as quality of education, child-rearing practices, or impact of the media on the individual. One approach to the problem is to accept that critical dimensions of cohort are not yet measurable, and hence that the concept is not currently amenable to scientific investigation. However, a more productive alternative is to examine characteristics that are quantifiable, such as amount of education, aspects of physical and mental health status that reflect medical advances, sensory abilities, etc. These measures are unlikely to reflect all aspects of what is meant by the cohort concept, but they nevertheless serve as a beginning for investigation of the concept.

The study reported in the target article statistically controlled amount of education as well as various indicators of physical and mental health status, and found large cross-sectional age trends in several different measures of cognitive functioning between 18 and 60 years of age. The conclusion was that cohort influences, at least as measured by those particular characteristics, cannot account for all of the cross-sectional age trends in cognitive functioning. It is important to emphasize that this does not mean that cohort differences do not exist and might not be contributing to cross-sectional age differences in cognitive functioning, but rather that con-

trol of the indicators that have been investigated thus far has not resulted in the elimination of the age differences.

Two other observations were also mentioned that appear inconsistent with the cohort interpretation. One is that the discrepancy between longitudinal and cross-sectional age trends is apparent over intervals as short as 7 years (Schaie), 5 years (Ronnlund et al.) or 2.5 years (current project). These results imply that if cross-sectional declines are attributable to cohort differences, then the cohort influences must operate over very short intervals, and not merely over periods of generations as sometimes assumed. The second observation is that cross-sectional age trends in measures of cognitive functioning have been reported in non-human animals raised in constant environments in which little or no cohort influences are likely to be operating.

In order to investigate retest confounds in longitudinal comparisons the hypothesized retest component in longitudinal change must be distinguished from other determinants of change. In my opinion no consensus is currently available regarding the best method for estimating retest effects. Instead of relying on a single method, therefore, I examined multiple methods, involving short-term retest influences, performance of individuals tested twice versus those tested once, and a statistical model capitalizing on variability of the retest intervals. Furthermore, rather than assuming that cognitive functioning was unitary, 12 different variables selected to represent 4 different cognitive abilities were examined with each method. As noted above, there was considerable variability in the absolute magnitudes of the retest estimates. However, most of the estimates of the retest effects were positive, which is consistent with the view that measures of longitudinal change reflect a mixture of influences, and that positive retest effects may be obscuring maturational decline occurring prior to about age 60.

Because the results revealed that the cross-sectional declines were not eliminated after adjusting for variations in characteristics that might be assumed to reflect cohort differences, but that the longitudinal age trends were likely distorted by the presence of positive retest effects, it was concluded that at least some of the discrepancy between cross-sectional and longitudinal age trends is probably attributable to the presence of retest effects masking declines in longitudinal comparisons.

Schaie raised a number of issues in his commentary, and with the exception of those that seem to reflect his personal preferences (e.g., the nature of the citations and the format of data presentation), each will be addressed in the following paragraphs.

A major objection appears to be that I am “reifying the cross-sectional fallacy”. I believe that this is a distorted characterization because my primary proposal is that researchers need to be careful in the interpretation of *both* cross-sectional and longitudinal data. Indeed, one of the final statements in the target article was that “strengths and weaknesses of both cross-sectional and longitudinal data . . . need to be considered when reaching conclusions about age trends in

cognitive functioning”. From my perspective, therefore, the most relevant fallacy may be the one I am trying to challenge, namely, the “single cause fallacy” of interpreting either cross-sectional or longitudinal data as a reflection of only a single type of influence. My guiding assumption was that both cross-sectional and longitudinal comparisons are likely influenced by multiple factors, and that researchers should try to identify and quantify those factors to investigate their relative contributions to the observed differences and changes. As Schaie points out, some version of the “cohort hypothesis” has been discussed for many years, but surprisingly little research has attempted to investigate the hypothesis by determining the specific variables that differ across birth cohorts, determining whether those variables are related to the measures of cognitive functioning which exhibit discrepancies between cross-sectional and longitudinal age trends, and determining if the cross-sectional age differences in those cognitive measures are reduced after statistically controlling the variation in those measures. Rather than accepting cohort and retest confounds as intrinsic and inevitable, and thereby reifying any particular assertion, my proposal was that the various determinants should be identified, quantified, and directly investigated.

It is indisputable that there have been generational increases in the average level of performance on some cognitive tests (i.e., the “Flynn Effect”), but the reasons for this phenomenon, and its implications for the interpretation of age trends in cognitive functioning, are still not well understood. In a recent monograph, I (Salthouse, *in press*) suggested that these historical increases in average level of performance should not necessarily be equated with cohort influences because the increases could reflect period influences, and might actually have greater distorting influences on longitudinal comparisons than on cross-sectional comparisons. That is, in a manner similar to how the existence of time-related inflation changes create greater complications in longitudinal contrasts of relations between age and salary than in cross-sectional contrasts, historical increases in average level of cognitive functioning could distort longitudinal comparisons more than cross-sectional comparisons. Until there is better understanding of the causes and consequences of generational differences in cognitive performance, therefore, it may be misleading to equate them with cohort differences, and to assume that they necessarily lead to confounds only in cross-sectional comparisons.

Schaie also questions the relevance of animal and neurobiological results to this type of research. I am frankly puzzled by the suggestion that results with non-human animals and with neurobiological variables might not be relevant to the question of when age-related cognitive decline begins. In fact, I believe that the neglect of the vast literature on animal cognition, and until recently also the literature on neurobiological variables, within the area of cognitive aging has led to a narrow, and possibly distorted, view of the nature of age-related differences and changes in cognition. As mentioned in the target article, I believe there are several reasons this litera-

ture is relevant to the current topic. First, as Schaie notes, congruence of age differences and age changes might be expected under conditions of stable environments. The existence of many reports of cross-sectional age differences in measures of memory and cognition in species from primates to fruit flies raised in nearly constant environments is therefore clearly relevant to the interpretation of cross-sectional data because these results suggest that age differences can occur even in the absence of environmental changes. And second, if retest effects distort longitudinal comparisons, then a discrepancy between cross-sectional and longitudinal age relations would be expected in non-human animals with measures of behavior that are susceptible to practice improvements, but no discrepancy would be expected with neurobiological variables, such as measures of brain volume, that are not susceptible to practice effects. As reported in the target article, research has supported both of these expectations, and thus these results are more consistent with the retest interpretation of the longitudinal – cross-sectional discrepancy than the cohort interpretation.

Schaie also claims that I do not pay sufficient attention to domains that remain stable. As noted earlier in this response, and in the target article, domains that remain stable are not directly relevant to question of discrepancy between cross-sectional and longitudinal because they do not exhibit the discrepancy.

Another objection is that I dismiss “well-established effects of cohort differences in providing the major cause for discrepancies between cross-sectional age differences and longitudinal age changes.” Instead of dismissing the cohort hypothesis as a possible explanation of the discrepancy, I actually attempted to investigate it, along with a plausible alternative hypothesis that postulates that retest effects distort longitudinal changes. Because Schaie does not mention what variables should be controlled in order to adequately control for cohort differences, and because few explicit characteristics other than education have been mentioned in the research literature, it is difficult to determine the best method of controlling for cohort differences. However, it is worth reiterating that I was not claiming that cohort differences are not a potentially important determinant of the discrepancy between cross-sectional and longitudinal age trends, but rather was suggesting that until the relevant characteristics are identified and measured to allow their influences to be evaluated, the cohort concept runs the risk of not being scientifically meaningful.

Another point in the commentary is that in addition to practice effects, short-term retest effects reflect fluctuation of an individual’s observed score around his/her true score. I agree with this point, but I also assume that this type of fluctuation of an observed score around the true score operates at all measurement occasions, including those separated by long intervals in traditional longitudinal studies. It isn’t clear from his commentary whether Schaie assumes that fluctuations around the true score do not occur with longer retest

intervals, but only if this were the case would this objection affect the interpretation of the short-term retest effects.

As with Nilsson et al. (this issue), Schaie claims that one particular method of assessing practice effects is “generally accepted,” and that it was not done in the target article. Because several other methods have recently been used to investigate retest effects, and because the twice-versus-once-tested method has limitations such as the difficulty of evaluating statistical significance of the retest effects and the inability to examine relations of retest effects with other variables, I do not believe that there is currently a consensus regarding the best method of distinguishing retest and maturation effects. Furthermore, it should be noted that the target article reported results with three analytical methods, including the twice-versus-once tested method.

In his final comment, Schaie questions the power to detect significant correlations with the retest interval variable because of small  $N$ 's. It is certainly possible that some relations were not detected because of low power, but it should be pointed out that the correlations between the retest interval and the size of the longitudinal change score were signifi-

cant for some cognitive variables but not for others, and yet they all had the same sample sizes. Furthermore, an earlier study found similar estimates of long decay rates for retest effects with different variables and samples of participants (Salthouse et al., 2004).

In conclusion, I want to thank the reviewers for their thoughtful comments on the target article. Exchanges such as these can be very valuable in two respects. First, they serve to clarify points of disagreement and thereby might stimulate future research. And second, it can be argued that progress in science occurs by constant questioning, not only of new research findings, but also of long held, “widely established” and “generally accepted” assumptions.

## References

- Salthouse, T.A., in press. Major issues in cognitive aging. Oxford University Press, New York.
- Salthouse, T.A., Schroeder, D.H., Ferrer, E., 2004. Estimating retest effects in longitudinal assessments of cognitive functioning in adults between 18 and 60 years of age. *Developmental Psychology* 40, 813–822.