On the Analysis of Sequential Data In Life-Span Developmental Research

Christopher K. Hertzog and K. Warner Schaie

The Pennsylvania State University

A paper presented at the 90th Annual Convention of the American Psychological Association, August, 1982.

Note: Comments welcome; please address correspondence to the first author at Department of Individual and Family Studies, College of Human Development, S-110 Human Development Building, University Park, PA 16802.

On the Analysis of Sequential Data in

Life-Span Developmental Research

One of the common goals in research on the life-span development of psychological constructs is description of the changes that are correlated with chronological age. Descriptive research on age-associated changes is often seen as a necessary first step toward explanatory analysis - after all, it is only reasonable to argue that one should demonstrate the existence of an age-related phenomenon prior to any concerted effort to account for its origins.

As is by now well known, there are major potential problems with descriptive research based upon simple "developmental" designs, such as the cross-sectional or single cohort longitudinal designs, in the form of rival interpretations to chronological age as the primary correlate (Schaie, 1965, 1977). Indeed, a vast literature on life-span developmental methodology has grown out of the realization that the methodological problems in descriptive inference in life-span development are often made difficult by a host of potential internal validity threats (e.g., Baltes & Nesselroade, 1979; Nesselroade & Reese, 1973). Schaie's (1965) general developmental model, which advocated the use of expanded sequential sampling, promoted the expansion of simple cross-sectional and longitudinal designs into sequential sampling designs (see also Baltes, Reese, & Nesselroade, 1977; Schaie & Hertzog, 1982)as a way of unconfounding the primary rival factors, cohort and period effects, from age effects.

Currently there is a great deal of controversy about the validity and utility of the design and analysis procedures advocated by Schaie, and about the validity of empirical research reports which have used these methods to study adult intellectual development (e.g., Adam, 1978; Baltes & Schaie, 1976; Botwinick & Arenberg, 1976; Donaldson, 1979; Horn & Donaldson, 1976; Schaie, 1979). Most of the methodological criticism has not questioned the usefulness of sequential sampling, i.e., the collection of data in cross-sectional or longitudinal sequences (Baltes et al., 1977). The papers have instead criticized Schaie's (1965) three bifactorial sequential designs and his methods of data analysis. Some of the strongest criticism may be found in the papers by Donaldson (1979) and Horn and McArdle (1980). These authors reject the validity of the parametric assumptions of Schaie's bifactorial designs and strongly criticize the use of traditional ANOVA to test hypotheses about mean differences in sequential data. They advocate instead the use of methods developed initially by Mason, Mason, Winsborough, and Poole (1973), which simultaneously estimate the effects of all three factors -- age, period, and cohort -- using the general linear model to estimate fitted constants representing mean differences among the levels of each factor (see also George, Siegler, & Okun, 1981).

In our view, the approach advocated by these authors represent a major contribution to the literature on sequential methodology, particularly in statistical treatment of sequential data. In particular, Horn and McArdle's modeling mean and covariance structures to estimate the age, cohort, and period effects from longitudinal sequences represents a significant methodological advance. However, it is also our view that these and other papers advocating the Mason et al. approach overstate the case against Schaie's bifactorial models and for the alternative designs. In doing so, they have inadvertently obscured the major issues. The purpose of this paper, then, is to present our somewhat different perspective on the problem hoping, thereby, to contribute to a dialogue on the appropriate methods for the analysis of sequential data.

To discuss the problem of sequential analysis, it is necessary to distinguish between the parametric model of the design and the statistical method of estimating the parameters. The parametric model specifies whether there is an effect associated with a given level of the age, cohort, or period factor. Given a parametric model in which each of the parameters is uniquely identified, one may then proceed to obtain statistical estimates of the parameters.

To facilitate the comparison of the different parametric models, consider first a univariate linear model for a cross-sectional sequence in which j ages, k cohorts, and m periods are measured (we will have more to say about the exact form of the sequential data matrix below):

$$Y_{ijkm} = \mu+\alpha_j + \beta_k + \gamma_m + (\alpha\beta)_{jk} + (\alpha\gamma)_{jm} + (\beta\gamma)_{km} + (\alpha\beta\gamma)_{jkm} + \varepsilon_{ijkm} \quad (1)$$

Equation (1) indicates that in some population of individuals, the individual scores, Y_{ijkm}, are potentially a function of the main effects and interactions of age, cohort, and period factors, where α_j are the effects of age, β_k are the effects of cohort, γ_m are the effects of time, interactions are given using parenthetical notation (e.g., $(\alpha\beta)_{jk}$ are interactions of age and cohort effects), and ε_{ijkm} are individual error scores. The well-known problem, of course, is the linear dependency among chronological age (A), cohort birth year (C), and time period (T), such that T = A + C (e.g., if you were born in 1900 and it is now 1981, you must be 81 years of age). This linear dependency limits our ability to estimate the effects given in Equation (1). That is, even though all the effects given are theoretically and conceptually distinct, and even though in principle all effects could be present for some dependent variables in certain populations, the linear dependency removes the possibility of estimating all effects in the analysis. All three-way interaction effects are non-estimable. Other effects are estimable, but only if certain other effects associated with the same cells in the sequential data matrix are assumed null. In other words, we must exclude certain types of effects from the model on theoretical grounds before the remaining effects are estimable.

Schaie's (1965) solution to the problem was to assume that all effects associated with one factor in the model were null. The cross-sequential design hypothesizes cohort and period effects, assuming all age main effects and interactions are nonexistent; the cohort-sequential design hypothesizes age and cohort effects, assuming all period effects to be null, etc. A model for the cohort-sequential design, then, is:

$$Y_{ijkm} = \mu+\alpha_j + \beta_k + (\alpha\beta)_{jk} + \varepsilon_{ijkm} \qquad (2)$$

This approach has been criticized by Donaldson (1979) and Horn and McArdle (1980), who argue that it is rarely, if ever, the case that all effects associated with one of the factors will be null. They offer instead an approach we call the Additive Effects model. It assumes (i) all interaction effects are null so that the general model is of the form

$$Y_{ijkm} = \mu+\alpha_j + \beta_k + \gamma_m + \varepsilon_{ijkm} \qquad (3)$$

and (ii) at least one additional effect is null, such that an assumption of the form $\alpha_1 = \alpha_2$ may be placed on the main effects. This latter assumption breaks the linear dependency in the main effects, and all remaining effects for age, cohort, and period are then estimable (except for scaling constraints; see below). The parameter estimates from the Additive Effects model are, therefore, valid only if both types of assumptions are satisfied.

Our objection to the Additive Effects model is only that, as a parametric model, it is in principle no better than Schaie's bifactorial models: if the assumptions of any of those models are incorrect, then the resultant parameter estimates are invalid. Its proponents are quite specific about the need for an accurate assumption about nullity of one or more main effects, and the potential problems if those assumptions are violated: models with differing assumptions of the second type produce estimates for the remaining parameters that often vary considerably between different models (Glenn, 1981; Horn & McArdle, 1980). However, they ignore the problematic nature of the additivity assumption, which has been criticized by others (e.g., Glenn, 1976). Simply put, one could turn the tables and argue that the assumption of additivity will rarely, if ever, hold for developmental phenomena. One reason for this argument is that age-correlated trends do not necessarily imply organismic development in its strongest sense; we might well expect age trends in many psychological constructs to vary over time, given secular trends and generational differences in nutrition, health maintenance behavior, education, life style, etc.

Put in perspective, then, the Additive Effects model is a contribution to the area because it provides another model to consider in a theoretical evaluation of psychological phenomena. If we believe, a priori, that the effects of age, period, and cohort exist but are completely additive, then the Additive Effects model is obviously most appropriate. If, on the other hand, we believe a priori that there are no effects associated with time period, then a cohort-sequential model is most appropriate. The essential problem is, of course, that we often do not know in advance which, if any, model is "true" -- usually because not enough is known about the constructs of interest, but especially because previously collected data is never model-free. An obvious goal for any developmental theory is a specific model for the nature and influences of the age, period, and cohort variables. However, assumptions based upon a well known hypothesis or theory often represent a parametric interpretation of previous data. This certainly appears to be the case for the fluid/crystallized theory, which developed as an age effects interpretation of differential patterns in cross-sectional age differences.

We cannot validate a theory about age, cohort, and period effects for a given set of developmental constructs merely by showing that descriptive effects estimated for a parametric model may be replicated, since all that is shown is that the phenomena are consistent. To be blunt, our model may simply be, consistently invalid. Similarly, we cannot validate a theory by analyzing a different sequential data matrix by building in assumptions which are a simple restatement of the old interpretation. In this sense, the Mason type approach cannot hope to discover the "true" model. We agree with Glenn (1981), who argues that the age/period/cohort analysis can only be evaluated on the basis of information outside the sequential data matrix. Thus, we cannot agree with the position of George et al. (1981) that the Additive Effects model with additional assumptions can discover the effects that are present in the data. Their analysis is dependent upon the validity of the additional assumptions, and the special characteristics of their simulated data. Certainly, one of the major contributions of the critiques of Schaie's model may be that we are forced to assess more precisely the theoretical validity of our parametric assumptions, and to inquire after ways in which our models may be falsified at explanatory levels other than the descriptive age, cohort, and period analyses discussed here.

While we take exception to the claims made on behalf of the parametric model for the Additive Effects approach, we believe that the methods of analysis used by proponents of that approach represent a significant advance over the ANOVA methods originally advocated by Schaie. That is, the use of the general linear model to estimate the actual effects under given model assumptions is undoubtedly superior to using omnibus ANOVA F-tests for a given class of effects (e.g., testing the null hypothesis that all age effects are equal to zero). Estimating the effects themselves is superior, becasue (a) one does not need to do post-hoc analysis on the untransformed marginal means to isolate the source of significant differences, and (b) one should not inspect marginal means, but rather effect estimates that are conditional on the parameteric assumptions used to produce estimability. Donaldson (1979) was quite clear and, we believe, quite accurate on this point.

Note, however, that these analysis methods may also be applied to models using the parametric assumptions of Schaie's bifactorial designs. Thus, one should not use advances in statistical estimation procedures as a criterion for evaluating the validity of parametric assumptions. For example, Schaie and Hertzog (Note 1) used multivariate regression with orthogonal polynomials for age trends to do exploratory analysis of sequential data using the cohort-sequential design (analyzing both cross-sectional and longitudinal sequences). Polynomial effect estimates were tested as planned comparisons, and the results were reported and evaluated in terms of the estimated polynomial effects, rather than a visual inspection of the observed means. If the parametric assumptions of the cohort-sequential design were in fact true for these data, then this approach to the data analysis, while not isomorphic with the single cell effect estimates obtained from a Mason type dummy regression analysis, would be perfectly appropriate and interpretable.

Use of the general linear model has another important advantage, however (regardless of whether single cell effects or larger contrasts such as the orthogonal polynomials are estimated). Perhaps the major advantage for life-span developmental research is that it leads us away from thinking in terms of balanced ANOVA designs when designing studies, and towards thinking in terms of unbalanced, incomplete factorial designs in which one is concerned with the estimability of effects, rather than whether or not one has completely crossed factorials. Breaking the "mind set" of traditional ANOVA designs is useful, for one reason because it allows for a solution to the problem of estimating age effects in sequential sampling matrices which would normally be considered only amenable to a cross-sequential (cohort by period) design.[1] We certainly agree with the previous complaints (e.g., Adam, 1978; Botwinick & Arenberg, 1976) that use of cross-sequential designs to make inferences about age effects is a hazardous and probably invalid endeavor. It also avoids one of the major problems of the cohort-sequential method, which is that collection of data for a completely crossed cohort-sequential design requires a large number of time periods -- unacceptably long for much work on adult development. Schaie and Hertzog (Note 1), desiring to apply a cohort-sequential approach, addressed this problem by taking data from longitudinal and cross-sectional sequences, partitioning them into multiple cohort-sequential data sets, and then performing separate multivariate regression analyses on each data set. The advantage of an alternative strategy of using the general linear model approach and including the entire data matrix (as an incomplete design in ANOVA logic) is that the parameter estimates are calculated simultaneously in a single analysis. The critical point, however, is that such an analysis _need not_ be performed under the parametric assumptions of the Additive Effects model. One could use cohort-sequential assumptions, or some hybrid set of assumptions that may be justified _a priori_ so long as these assumptions provide theoretically meaningful and statistically estimable parameters. In our view, then, the major contribution of the Donaldson and other critiques may not be the promotion of the Additive Effects model (particularly if this leads to blind application of such an approach by unsophisticated users), but rather the movement away from the logic of fully crossed ANOVA designs toward theory-based specifications of parametric assumptions.

Full consideration of the issues raised above is not possible in the present forum. What we hope to do in the remainder of this paper is to demonstrate an empirical application of the Additive Effects model to the data collected by Schaie and colleagues and analyzed previously by Schaie and Hertzog (Note 1) using the cohort-sequential approach. We have two specific goals: (a) a test of the hypothesis that a violation of the cohort-sequential assumptions, in the form of additive effects for period, may have biased the results. In particular, we are interested in whether an Additive Effects model would lead to substantively different conclusions regarding age changes and cohort differences than the cohort-sequential analysis; and (b) we wish to demonstrate empirically the problem of indeterminacy in the use of Additive Effects models under different parametric assumptions of the second type (e.g., two age effects equal). To anticipate, when the Additive Effects approach is used without strong theoretical justification for a given assumption, then multiple analysis using widely different assumptions is probably appropriate, only in the sense that one can then determine the degree of change in the parameter estimates under different enabling assumptions. As we shall see, this is often less than satisfactory, however, because the resulting parameter estimates can vary considerably.

Method

The data analyzed in this report consist of psychometric test scores on Thurstone's 1948 Primary Mental Abilities (PMA)test, collected on a sample of adult members of a health maintenance organization in the greater Seattle area (Schaie, 1979). The samples consisted of two separate longitudinal sequences, tested over three seven-year periods (1956, 1963, 1970; and 1973, 1970, & 1977). The data from eight seven-year birth cohorts, each measured at three occasions, enabled us to examine effects for eight birth cohorts (mean birth years 1889 to 1938), four time periods (1956, 1963, 1970, and 1977) and nine age levels (mean ages 20 to 81). Table 1 reports the cell frequencies of the data for these longitudinal sequences.

For the purposes of this report, we used only the composite measure of general intelligence (IQ) from the PMA, calculated as a weighted composite of the five PMS subtests (see Schaie, 1979, for a detailed description of the individual subtests and for the formula for calculating the composite IQ variable). IQ was selected to T-scores (mean = 50, s.d. = 10).

In order to estimate the coefficients for the Additive Effects models, we used the LISREL IV program to specify a series of models fitting the mean structure of the observations for the 14 age/cohort groups reported in Table 1 (collapsing over Sex). This method, described by Horn and McArdle (1980) in their RAM notation, involves fitting constants to the mean structure, while leaving the covariance structure of the variables unconstrained. The approach, therefore, has the advantage of not making the usual assumptions of traditional MANOVA applications that the covariance matrices of the groups are equal in the population. In order to use the Horn and McArdle approach, it is necessary to establish a matrix of dummy coefficients in LISREL's Λ_y matrix, and then model the age, period, and cohort effects as regression coefficients in LISREL's $\underset{\sim}{\Gamma}$ matrix.[2] Identification is achieved by constraining the appropriate coefficients to equivalence across different groups. Thus, groups 1, 2, 3, and 4 all contain subjects who have a mean age of 32 at some point in time. Identification of one of the regression coefficients as being the "age 32" parameter is achieved by constraining one of the coefficients for each of these groups to be equal over all the groups (constrained parameters in LISREL terminology). The coefficients for one age, one period, and one cohort must be fixed at 0 to resolve the scaling indeterminacy. This procedure defines the remaining effects as deviation contrasts from the age, period, or cohort with a fixed zero effect. The logic is essentially the same as that for the assignment of 1's and 0's to dummy coding vectors in usual multiple regression applications of the Mason et al. approach.

We estimated a series of models under Additive Effects assumptions that there were no interactions present in the population. There were essentially three types of models: (a) models which placed just one restriction on the age, period, and cohort effects, achieving a just-identified solution for the remaining parameters. We estimated a series of such models with widely varying assumptions (some that were plausible to us, others that were implausible) in order to examine variations in effect estimates; (b) models with more than a single restriction on the age, cohort, and period effects, based in large part upon modifications of models of the first type; and (c) models which hypothesized that a whole class of parameters (e.g., all age effects, all cohort effects, or all effects were equal to zero). These restricted models were used to calculate likelihood ratio χ^2 statistics testing the null hypothesis of the form: all cohort effects are zero, conditional on the existence of all age and period effects, etc. This approach was used by Horn and McArdle (1980) to evaluate whether reduced rank model for the additive effects could successfully account for the data. It is logically equivalent to the approach advocated by George et al. (1981) as a method for using the Additive Effects model for exploratory purposes, although they used more traditional multiple regression methods to solve the problem.

RESULTS AND DISCUSSION

In order to facilitate discussion of the Additive Effects models and the rationale for arguing for or against certain assumptions, we provide a graphic representation of the observed means for IQ for all fourteen groups in Figure 1. Table 2 also provides the cohort-sequential F-tests and coefficients for the seven fourteen-year age ranges spanned by independent groups of subjects. As can be seen from Table 2, the main results of the cohort-sequential analysis were as follows: there was significant age decrement over the 46-60 age interval and beyond; (ii) cohort differences were statistically reliable only between the 1938 and 1931 cohorts;[3] and (iii) there were reliable Cohort x Age interactions, involving interaction with the quadratic age trend, in several data sets. We should note that these data, which include information from a 1977 testing not previously published, tend to indicate statistically reliable decrement earlier in the life-span than did earlier analyses from the Seattle study (Schaie and Hertzog, Note 1).

Table 3 reports the results from a series of Additive Effects Models that are just-identified in the age, cohort, and period parameters (Models I through IX). The goodness of fit test is not significant ($x^2 = 27.16$ with 24 df, $p = .30$), indicating that the assumption of additivity cannot be rejected.[4] The results from these just-identified models are relatively discrepant. Models I-III place different restrictions on pairs of cohort groups. Model I constrains the two oldest cohorts, C_1 and C_2, to be equal. This assumption does not appear to be problematic if one merely inspects the date in Figure 1, which shows the two cohorts to be nearly overlapping. However, this assumption results in benign estimates of age change (note the increment until age 60, followed by modest decrement) relative to the observed slopes in Figure 1. It appears that much of the putative decline has been absorbed into negative (monotonically decreasing) period effects. On the other hand, equality constraints on Cohort 5 and Cohort 6, which appear to be quite different in the observed means, has an opposite effect: large, early declines with age and cohort and period effects in the opposite direction!

The age constraints were selected to represent a variety of guesses as to the age of initial onset of decline in IQ performance. A proponent of the "early decrement" hypothesis might assume that the best guess is the most conservative -- assume no decline between the youngest two ages and set age 25 and 32 equal. Model IV, which does this, finds large scale decrement beginning with the parameter for age 39, and achieving a fully two standard deviation decline by age 81. The problem with this assumption, however, is that the observed data show the opposite pattern -- increment from age 25 to 32 that is consistent across the two cohorts (see Figure 1). Thus the assumption that there is no age-related change from age 25 to age 32 forces the model to account for the increment in terms of large positive period effects.

Note, that the assumption that age 32 = age 39, represented in Model V, does not produce the large positive period effects, nor does it produce the same magnitude of decline. Similar results are obtained in Model VI, which assumes age 39 = age 46, and a still more benign picture of age changes emerges (as might be expected) when the equality constraint is placed upon ages 46 and 53. As was the case for the cohort parameter constraints, Models VIII and IX, which place different equality constraints on period effects, arrive at dramatically different results.

Quite obviously, one has a wide variety of assumptions and outcomes to choose from in these models. Let us reiterate Glenn's (1976) point: there is no purely statistical solution to the problem. Given the relative(!) consistency of the results under the age assumptions, we might elect to adopt the assumption that $a_{32} = a_{39}$, given the implausibility of the $a_{25} = a_{32}$ assumption when compared against the observed means in Figure 1. Note however, that Model V actually shows an increase between ages 39 and 46. Given a decremental conceptualization, we might require that this difference be no greater than zero. Model X shows the results obtained when assuming stability in age-correlated trends from age 32 through age 46. This additional constraint does not affect the parameter estimates, nor does it cause an appreciable reduction in the fit of the model (X^2 = 27.22 with 25 df; change in X^2 <1).

Model XI imposes an additional change; it hypothesizes that all time period effects are null in the population. This model modification derives from the fact that the period effects are not reliably different from zero in Model X, given their standard errors. Model XI does produce a significant reduction in fit (X^2 = 51.49 with 28 df, $p < .01$; change in X^2 = 14.27 with 3 df, $p < .001$). Even though the period effects are not reliably different from zero, they are needed to achieve a level of fit equal to the just-identified models.

At this point we might end the process and decide to accept Model X. We note, somewhat disingeniously, that Model X just so happens to be highly consistent with the cohort-sequential results reported in Table 2. Are we arguing parsimoniously in rejecting the other solutions, or are we simply "validating" our own perceptions of reality? Would the logical status of our hypothetical acceptance of Model X truly be any better if we had previously pronounced and adopted some face valid theoretical argument justifying the assumption that age 32 = age 39?

The conundrum is brought even more clearly into (or, possibly, out of) focus when we attempt to use the restricted modeling approach as advocated by George, Siegler, and Okun (1981). Models XII, XIII, and XIV each assume the absence of all effects associated with one factor; age, period, or cohort. We are then in a position to evaluate the goodness of fit to determine whether cohort, age, or period effects should be truly considered null. Table 4 reports the parameter estimates and the goodness of fit statistics for these models. Although we might expect that all of the effects would be significant (recall that there were Cohort x Age interactions in the cohort sequential analysis), in fact, a parsimonious result emerges that is completely inconsistent with Model X; Model XIV, which constrains all cohort effects to equal zero, fits the data well! Moreover, the pattern of effects coincides nicely with discussions of these types of data by Botwinick (1977); that is, there are large age declines (because of the "mistaken" assumption of non-zero cohort effects) and large, positive period effects. Reasoning in a post-hoc fashion, we could argue that these are not true period effects (which we probably could not explain) but rather confounded practice effects in a longitudinal data set. We might argue, then (especially if we were predisposed to accept a decremental view of life-span intellectual development), that this most parsimonious model, fitting the data equivalently with fewest parameters.)

and consistent with the large body of cross-sectional data, is the true model.

The appeal to parsimony, however, is problematic if we examine closely the nature of the relationship between age and cohort effects in this range of birth years. Since cohort effects are essentially monotonically decreasing, and since period effects are (at least under Model XIV) monotonically increasing over time, it is all-too easy to absorb the cohort effects into the other effects with a simple linear tradeoff. Glenn (1981) has a recent and compelling demonstration of this problem, and he points out that when the effects are all linear, the most parsimonious model may not in fact be correct (see also Adam, 1978).

Thus we have come face to face with the fundamental indeterminacy of exploratory analysis using Additive Effects or other models on sequential data. Scientists with different points of view could analyze the same data and come to radically different conclusions. As noted by Glenn (1981), the only way to determine the validity of an age/cohort/period model is to attempt to go outside the data matrix, either by collecting new data with which the model is inconsistent, or by (preferably) theoretical consideration of the meaning of the assumptions involved, determining consequent premises, and then proceeding to test these premises. The latter approach rather quickly leads us away from descriptive research toward explanatory research. As Mason et al. (1973) noted, much of the problem derives from the logical status of the age, period, and cohort variables. If it is possible to replace them with process-oriented variables which measure the process by which the cohort effects (etc.) came to be manifested, then the linear dependency broken. More important, our understanding of the phenomenon is likely to increase. A prime candidate in this regard would be replacement of arbitrary birth cohorts qualitatively with theoretically-meaningful cohort distinctions

(Rosow, 1978; Schaie, 1983).

With respect to descriptive examination of data, Glenn (1981) has pointed out that the best procedure is to collect additional data from subsequent time periods to observe whether a model's parametric assumption seem to hold up under closer and continued scrutiny. A critical point reducing the uncertainty would be the presence of non-linear trends in the effects of one or more of the variables. For example, a reversal of the direction of cohort effects in the population (recent cohorts performing more poorly than earlier cohorts) would be useful because the effects could not be so easily "reabsorbed" from one set of effects to the other. In fact, given recent patterns in reading scores and other school tests, we might well expect such a reversal. Indeed, cross-sectional sequence data from Schaie's projects shows that Cohort 9 (mean birth year 1945) performed significantly poorer than Cohort 8. This suggests that analysis of the cross-sectional sequences to determine consistency with the age pattern of the divergent models would be useful. The problem, however, is that any reanalysis would also be subject to the indeterminancy problem; furthermore, we would expect _a priori_ that cross-sectional sequences would be subject to different types of additional internal validity confounds than the longitudinal sequences. Thus there is sufficient room for intellectually honest investigators to come to radically different, yet defensible conclusions from the same data -- a problem that has plagued age-comparative factor analysis for years (e.g., Reinert, 1970).

We therefore must inevitably return to the point made earlier: only if there exists a strong theory to be used as the basis for the parametric assumptions needed for age/period/cohort analysis can we have any confidence in the resulting outcomes. Perhaps the problems inherent in valid descriptive inference from sequential data will lead us to reconsideration of the theoretical status of concepts in life-span developmental psychology, and a more careful consideration of the criteria by which strong inference about competing predictions from developmental theories may be evaluated. If we are in the process led away from descriptive research on mean differences in cross-sectional or sequential data matrices, so much the better.

Reference Notes

1. Schaie, K. W. and Hertzog, C. Fourteen-year cohort-sequential analysis of adult intellectual development. Unpublished manuscript, 1982.

References

Adam, J. Sequential strategies and the separation of age, cohort, and time-of-measurement contributions to developmental data. _Psychological Bulletin_, 1978, _85_, 1309-1316.

Baltes, P. B. & Nesselroade, J. R. History and rationale of longitudinal research. In J. R. Nesselroade and P. B. Baltes (Eds.), _Longitudinal research in the study of behavior and development_. New York: Academic Press, 1979-1-39.

Baltes, P. B., Reese, H. W., & Nesselroade, J. R. _Life-span developmental psychology: Introduction to research methods_. Monterey, CA: Brooks-Cole, 1977.

Baltes, P. B. & Schaie, K. W. On the plasticity of intelligence in adulthood and old age: Where Horn and Donaldson fail. _American Psychologist_, 1976, _31_, 720-725.

Botwinick, J. & Arenberg, D. Disparate time spans in sequential studies of aging. _Experimental Aging Research_, 1976, _2_, 55-66.

Donaldson, G. On the formulation, estimation, and testing of a developmental model specifying age, cohort, and time parameters. Unpublished doctoral dissertation, University of Denver, 1979.

George, L. K., Siegler, I. C., & Okun, M. A. Separating age, cohort, and time of measurement: Analysis of variance and multiple regression. _Experimental Aging Research_, 1981, _7_, 297-314.

Glenn, N. D. Cohort analysts' futile quest: Statistical atempts to separate age, period, and cohort effects. _American Sociological Review_, 1976, _41_, 900-904.

Glenn, N. D. Age, birth cohort, and drinking: An illustration of the hazards on inferring effects from cohort data. _Journal of Gerontology_, 1981, _36_, 362-369.

Horn, J. L. & Donaldson, G. On the myth of intellectual decline in adulthood. American Psychologist, 1976, 31, 701-719.

Horn, J. L. & McArdle, J. J. Perspectives on mathematical/statistical model building (MASMOB) in research on aging. In L. W. Poon (Ed.), Aging in the 1980's: Selected contemporary issues in the psychology of aging. Washington D.C.: American Psychological Association, 1980, 503-541.

Mason, K. O., Mason, W., Winsborough, H. H., & Poole, W. K. Some methodological issues in the cohort analysis of archival data. American Sociological Review, 1973, 38, 242-258.

Nesselroade, J. R., & Reese, H. W. (Eds.) Life-span developmental psychology: Methodological issues. New York: Academic Press, 1973.

Reinert, G. Comparative factor analytic studies of intelligence throughout the human life-span. In L. R. Goulet and P. B. Baltes (Eds.), Life-span development: Research and theory. New York: Academic Press, 1970, 115-145.

Roscow, I. What is a cohort and why? Human Development, 1978, 21, 65-75.

Schaie, K. W. The primary mental abilities in adulthood: An exploration in the development of psychometric intelligence. In P. B. Baltes and O. G. Brim, Jr. (Eds.), Life-span development and behavior, Vol. 2. New York: Academic Press, 1979, 67-115.

Schaie, K. W. A general model for the study of developmental problems. Psychological Bulletin, 1965, 64, 92-107.

Schaie, K. W. Can the longitudinal method be applied to psychological studies of human development? In F. J. Moenks, W. W. Hartup, and J. DeWitt (Eds.), Determinants of behavioral development. New York: Academic Press, 1972.

Schaie, K. W. Quasi-experimental research designs in the psychology of aging. In J. E. Birren and K. W. Schaie (Eds.), _Handbook of the Psychology of Aging_. New York: Van Nostrand Reinhold, 1977, 39-58.

Schaie, K. W. Historical time and cohort effects. In K. A. McCluskey and H. W. Reese (Eds.), _Life-span developmental psychology: Historical effects_. New York: Academic Press, 1983 (in press).

Schaie, K. W. & Hertzog, C. Longitudinal methods. In B. B. Wolman (Ed.) _Handbook of Developmental Psychology_, Englewood Cliffs, NJ: Prentice-Hall, 1982, 91-115.

Footnotes

[1]There are, however, conceptual problems with age-oriented analyses based upon decomposition of replicated cross-sectional samples over just two time periods. The data matrix will be dominated, so to speak, by the between subjects age differences, and an Additive Effects model may be particularly prone to mistaken inference given a misspecified model.

[2]This specification is highly complex, and cannot adequately be described here. We thank Jack McArdle for his help in translating the Horn/McArdle RAM model into a LISREL IV specification. Individuals interested in seeing the LISREL specification should contact C. Hertzog personally.

[3]There were, however, trends for differences between other pairs of adjacent birth cohorts. Analyses of the cross-sectional sequences, which had larger _N_, detected a greater number significant cohort differences than were found in the longitudinal sequences.

[4]The collection of extended sequential data over more than three times of measurement creates surplus degrees of freedom for certain cells in the matrix. Absolute χ^2 can then be used to see if any model just identified in the age, period, and cohort effects fits the means. If not, then the hypothesis of additivity is rejected. However, if it cannot be rejected, it should not be assumed true, since in theory an equivalent reduced rank model with interaction effects must exist. One could, however, investigate whether any theoretically plausible model with interaction fits the data as well. In general, however, likelihood ratio χ^2 testing of restricted models is not an "objective" statistical method for evaluating the "truth" of parametric models, given the conditional nature of the hypothesis tests.

[5] Note that previous analyses using ANOVA on these data have generally found little evidence for practice effects (e.g., Schaie, 1972). However, since these analyses did not use the Additive Effects approach, one could still argue that practice effects did exist but were observed by other countervailing influences.

TABLE 1

DESCRIPTION OF DATA
FROM LONGITUDINAL SEQUENCES

DATA SET	Cohort (mean birth year)	Age (mean age)	Males	Females	Total
I	1938	25,32,39	8	14	22
	1931	25,32,39	10	11	21
II	1931	32,39,46	14	27	41
	1924	32,39,46	11	15	26
III	1924	39,46,53	23	28	51
	1917	39,46,53	11	15	26
IV	1917	46,53,60	25	26	51
	1910	46,53,60	17	15	32
V	1910	53,60,67	17	31	48
	1903	53,60,67	13	15	28
VI	1903	60,67,74	8	10	18
	1896	60,67,74	3	12	15
VII	1896	67,74,81	8	12	20
	1889*	67,74,81	8	6	14

TABLE 2

COHORT-SEQUENTIAL RESULTS: EFFECT CONTRASTS AND F-RATIOS

Source	Static	Data Set (see Table 1)						
		I	II	III	IV	V	VI	VII
Age (Linear)	F	28.55***	1.65	1.69	6.73*	27.49***	29.23***	42.33***
	C	2.54	0.68	-0.56	-1.07	-2.59	-3.73	-7.71
Age (Quadratic)	F	6.81*	<1	8.43**	<1	12.13***	3.85	3.29
	C	-3.47	-0.50	-1.73	-0.08	-2.70	-1.67	-2.00
Cohort	F	6.11*	<1	<1	3.03	<1	2.02	<1
	C	-4.97	-0.20	-1.07	2.63	0.47	-4.07	1.17
Cohort x Age (Linear)	F	2.23	<1	<1	<1	<1	<1	<1
	C	-	-	-	-	-	-	-
Cohort x Age (Quadratic)	F	<1	6.06*	5.96*	1.51	7.84**	5.25*	<1
	C	-	-	-	-	-	-	-

Abbreviations: F-ANOVA F-Test; C-effect contrast.

Note: For data set descriptions, see Table 1.

TABLE 3

RESULTS FROM ADDITIVE EFFECTS MODELS

PARAMETER ESTIMATES

MODEL	ASSUMPTION	M_0	A_{25}	A_{32}	A_{39}	A_{46}	A_{53}	A_{60}	A_{67}	A_{74}	A_{81}	C_8	C_7	C_6	C_5	C_4	C_3	C_2	C_1	t_1	t_2	t_3	t_4
I	$C_1=C_2$	59.17	0*	2.59	3.84	5.20	5.37	6.01	5.18	3.60	-.41	0*	-4.09	-5.54	-8.78	-9.55	-11.28	-15.71	-15.71	0*	-1.72	-2.01	-3.63
II	$C_4=C_5$	58.40	0*	1.82	2.30	2.89	2.28	2.15	.55	-1.80	-6.57	0*	-3.32	-3.99	-6.46	-6.46	- 7.42	-11.09	-10.31	0*	- .95	- .47	-1.31
III	$C_5=C_6$	55.92	0*	- .66	-2.66	-4.55	-7.63	-10.23	-14.31	-19.31	-26.38	0*	- .84	.96	.96	3.45	4.97	3.77	7.03	0*	1.53	4.49	6.12
IV	$A_{25}=A_{32}$	56.58	0*	0*	-1.34	-2.57	-5.00	-6.94	-10.37	-14.54	-21.12	0*	-1.50	- .35	-1.01	.81	1.67	- .18	4.15	0*	.87	3.17	4.14
V	$A_{32}=A_{39}$	57.92	0*	1.34	1.34	1.45	.36	.24	- 2.33	- 5.16	-10.40*	0*	-2.84	-3.03	-5.03	-4.54	- 5.03	- 8.21	- 6.95	0*	- .47	.49	.13
VI	$A_{39}=A_{46}$	57.81	0*	1.22	1.11	1.11	- .08	- .80	- 3.00	- 5.94	-11.29	0*	-2.73	-2.81	-4.69	-4.10	- 4.47	- 7.54	-6.17	0*	- .36	.71	.46
VII	$A_{46}=A_{53}$	59.00	0*	2.43	3.51	4.71	4.71	5.19	4.19	2.44	- 1.71	0*	-3.93	-5.20	-8.28	-8.89	-10.45	-14.73	-14.55	0*	-1.55	-1.68	-3.13
VIII	$t_2=t_3$	58.88	0*	2.30	3.26	4.33	4.20	4.55	3.43	1.55	- 2.73	0*	-3.80	-4.95	-7.90	-8.30	- 9.82	-13.97	-13.66	0*	-1.43	-1.43	-2.75
IX	$t_3=t_4$	57.55	0*	.97	.60	.35	-1.11	-2.08	-4.53	-7.33	-13.34	0*	-2.48	-2.30	-3.93	-3.07	- 3.20	- 6.02	- 4.38	0*	- .10	1.22	1.22
X	$A_{32}=A_{39}=A_{46}$	57.89	0*	1.21	1.21	1.21	.06	- .61	-2.76	-5.66	-10.98	0*	-2.77	-2.91	-4.82	-4.27	- 4.68	- 7.80	- 6.47	0*	- .41	.62	. 32
XI	$A_{32}=A_{39}=A_{46}$; t=0	57.62	0*	1.91	1.91	1.91	1.13	.82	-1.09	-3.51	- 8.48	0*	-2.92	-3.42	-5.39	-5.03	- 6.02	- 9.46	- 8.72	0*	0*	0*	0*

Abbreviations: A_{32} (etc.) age 32; t_1-time period 1 (1956); C_1-cohort 1 (birth year 1938); M_0-intercept.

Note: 0* devotes a fixed zero parameter.

TABLE 4

PARAMETER ESTIMATES AND GOODNESS OF FIT TESTS FOR RESTRICTED MODELS

MODEL	ASSUMPTIONS	M_0	A_{25}	A_{32}	A_{39}	A_{46}	A_{53}	A_{60}	A_{67}	A_{74}	A_{81}	C_8	C_7	C_6	C_5	C_4	C_3	C_2	C_1	t_1	t_2	t_3	t_4
XII	all a=0	59.10	0*	0*	0*	0*	0*	0*	0*	0*	0*	0*	.01	-2.19	-4.63	-4.78	-6.08	-15.25	-12.68	0*	-.87	-.45	-1.89
XIII	all t=0	57.61	0*	1.92	1.93	2.52	1.63	1.36	- .56	-2.98	-7.95	0*	-3.30	-3.66	-5.89	-5.55	-6.55	- 9.98	- 9.24	0*	0*	0*	0*
XIV	all c=0	55.72	0*	.29	-.67	-1.52	-3.54	-5.06	-8.12	-11.96	-18.41	0*	0*	0*	0*	0*	0*	0*	0*	0*	.50	.39	2.92

GOODNESS OF FIT TESTS

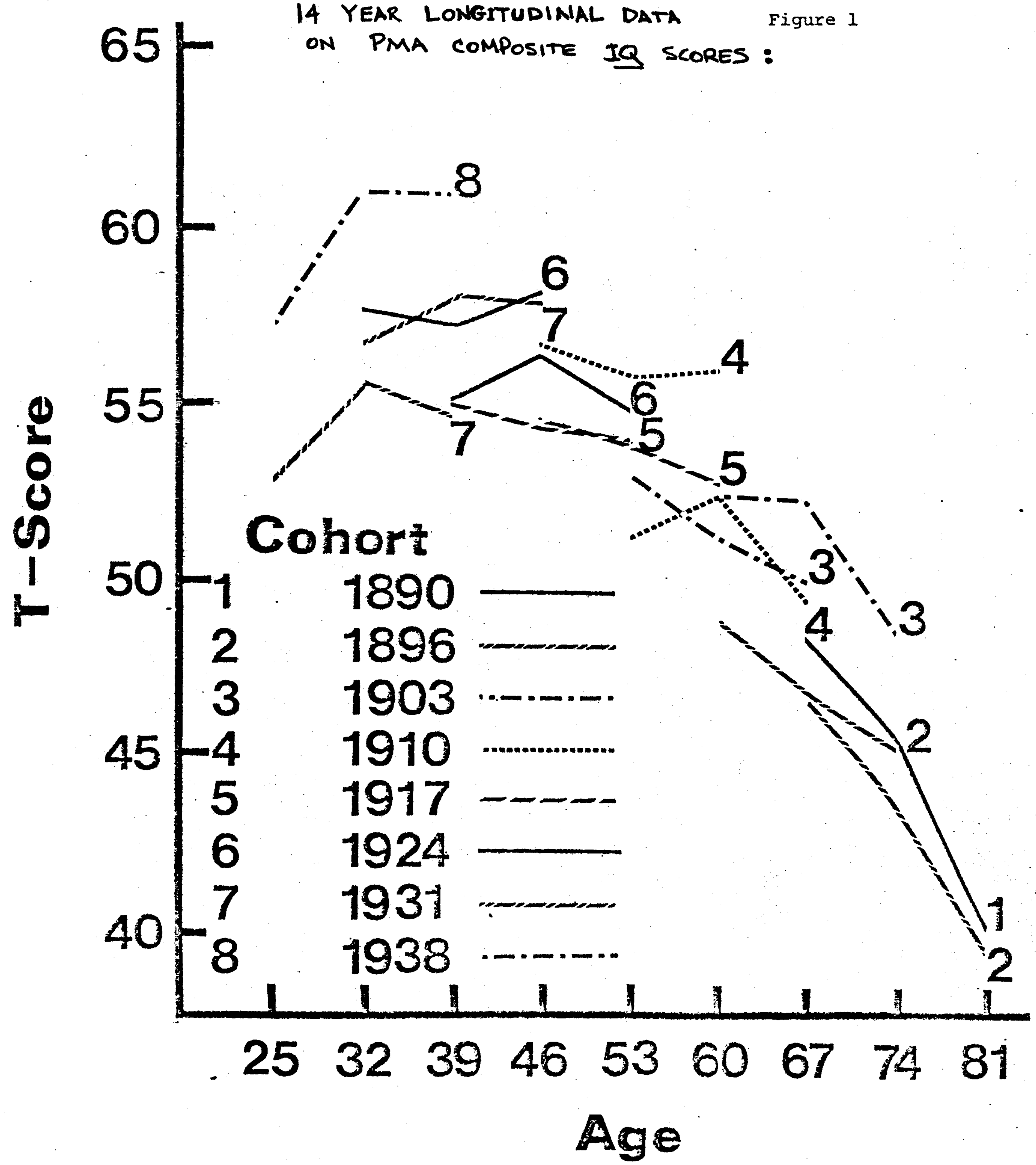
MODEL		χ^2	df	P	$\Delta\chi^2$	df	P
XII	all a=0	117.81	31	.00	90.65	8	<.001
XIII	all t=0	47.37	26	.01	20.21	3	<.001
XIV	all c=0	31.64	30	.38	4.48	7	N.S.

Abbreviations: A_{32} (etc.) age 32; t_1-time period 1 (1956); C_1-cohort 1 (birth year 1938); M_0-intercept. $\Delta\chi^2$-change in χ^2 from just identified solution.

Note: 0* denotes a fixed zero parameter.

14 YEAR LONGITUDINAL DATA ON PMA COMPOSITE IQ SCORES:

Figure 1



Raw data from 2 14-year longitudinal sequences used in Additive effects analysis.