Developmental Designs Revisited

K. Warner Schaie

The Pennsylvania State University

Paper Presented at the

1986 West Virginia

Life-span Developmental Psychology Conference

Morgantown, WV

April 24-26, 1986

Developmental Designs Revisited

K. Warner Schaie

The Pennsylvania State University

Introduction

I want to take the opportunity of this presentation to trace in some detail the evolution of my thinking on how to design studies that will characterize developmental progressions with minimal conceptual ambiguity. As part of this review it will become clear that my approach has obviously been affected by the impact of the contributions of numerous colleagues, whether as research collaborators or as critics. Equally important, however, has been the continuous effort to collect developmental data in such a way that these designs could be put to empirical test. Indeed here is a prime example of the dialectic interaction of puzzling data sets resulting in the examination of the appropriateness of standard research design, and the specification of alternate designs leading to the collection of new data sets that would fit the new paradigms.

Some Historical Comments

The work that I wish to review began with the realization that data on the adult development of mental abilities showed wide discrepancies between cross-sectional and longitudinal data collected on the same subject population over a wide age range. In particular, it became evident that for some dependent variables substantial age

differences obtained in cross-sectional studies could not be replicated in the longitudinal data while for other dependent variables, longitudinal age changes reflected more profound decrement than was shown in the comparable cross-sectional age difference patterns (Schaie & Strother, 1968).

I attempted to explain these discrepancies by constructing a general model for the study of developmental change that explicated the relationships between the cross-sectional and longitudinal methods. From this model it became possible to show that cross-sectional data involve the description of age differences at a single point in time; it is by definition a separate samples design _a la_ Campbell and Stanley (1963). Such a design suffers from the problem that maturational change (age) is confounded with cohort acting as a selection factor (see also Schaie, 1984). Longitudinal data, by contrast, involve a time series assessing the same individuals at two or more points in time. Here maturational change (age) is confounded with historical (secular) trends; a factor that I formally described as time-of-measurement effect. The general model also allowed the derivation of a third approach to the collection of developmental data for which I coined the term "time-lag." This latter approach involves the comparison of two samples at the same chronological age but at different calendar times, as would be the case, for example, in the comparison of SAT scores for successive classes of high school graduates. In this design cohort differences are confounded with time-of-measurement effects. More complex designs were then

described, termed "sequential methods," that proposed to permit estimates of the magnitude of specific components of developmental change by controlling for the confounds mentioned above (Schaie, 1965).

The sequential designs were applied to empirical data sets obtained as part of the Seattle Longitudinal Study (Schaie, 1979, 1983; Schaie & Hertzog, 1983). From these applications it soon became apparent that there were specific patterns of data acquisition that lend themselves most readily to optimal utilization of the sequential analysis strategies (Schaie & Geiwitz, 1982; Schaie & Willis, 1986b). It also became apparent that different sequential designs were appropriate for different developmental questions (Schaie, 1973), and that there was a need to specify design complications that allowed for the control of some of the validity threats specified by Campbell and Stanley (1963). Design variations were therefore explicated that permit controls for reactivity, practice and experimental mortality (Schaie, 1977, 1982; Schaie & Parham, 1974). Most of our early concerns were related to the estimation of development as a function of change in performance level. More recent work has extended these concerns to the comparison of structure (i.e., the regression of observables upon latent constructs) across different age groups and within cohorts over time (Hertzog & Schaie, 1986; Schaie, 1986; Schaie & Hertzog, 1982, 1985, 1986).

Scope of Presentation

In the body of this presentation I will formally examine some of the characteristics of the general developmental model, review how the strategies derived therefrom relate to non-developmentally oriented prescriptions for quasi-experiments, describe an optimal data collection approach that permits flexible application of sequential data analysis strategies, and discuss strategies for the estimation of specific developmental components and the presentation of developmental change gradients for a number of applications. Because others at this conference will deal extensively with design problems in studying structural change, my remarks will be confined primarily to the examination of design problems in the study of developmental changes in level of performance

The Basic Model

The original formulation of the General Developmental Model specified that any developmental change could potentially be decomposed as being influenced by one or more of three independent sources (Schaie, 1965). The general equation offered was of the form,

$$\underline{R} = \underline{f}(\underline{A},\underline{C},\underline{T}) \qquad (1)$$

where the observed change in response ($\underline{R}$) could be described as a linear combination unobserved components of age (maturational) change ($\underline{A}$), differences between cohorts in experiences that occurred prior to the first time-of-measurement ($\underline{C}$), and environmental impact (period

+ at_{ik} + ct_{jk} + act_{ijk} + e_{ijk} (2).

E ion of developmentally oriented data bases then suggested hat ditional cross-sectional and longitudinal approaches as well newly defined time-lag alternative are simply special case model in which one of the terms has been set to zero, whil her two are confounded. In cross-sectional studies, succ samples must, by definition, be drawn from different birth coh single cross-sectional study, of course, does not reflect any effects, but individual differences on the dependent var st be composed of some combination of variances associated wi rences in age and cohort. Traditional single-cohort lo al studies, by contrast, do not reflect any cohort di s, but confound the effects of age changes in the dependent va ith period effects occurring over the calendar time during wh ge is monitored. In the third alternative, a time-lag s ccessive samples are obtained from different birth cohorts at e t ages. This strategy consequently controls for the m nal component, but confounds cohort and period effects. lizing the differential composition of cross-sectional age ces and longitudinal age changes, it became clear immediately effects could be equivalent only under the excessively strong on that both cohort and period effects would be zero over the time/cohort frame from which the data to be compared had been sampled. To estimate the unobserved pure "age" effects it would therefore be necessary to decompose the sources of developmental variance. It was for the purpose of generating data that would permit suitable decomposition that I proposed to collect sequential data, in essence advocating replication of either cross-sectional and/or longitudinal studies (also cf. Schaie & Baltes, 1975). Such data were hoped to provide a sufficient set of equations that would allow decomposing individual differences into the three unobserved sources of developmental change specified above.

Limitations of the Model

Our proud hopes were soon humbled somewhat when it became apparent that unambiguous decomposition of the three components was possible only by imposing the side condition that,

either $a_i = 0$, or $c_j = 0$, or $t_k = 0$ (3),

and that the triple interaction as well as the simple interactions involving the main effect set to 0, were also set to 0 (cf. Costa & McCrae, 1982; Schaie, 1977; Schaie & Hertzog, 1982).

Given the validity of any one of these side conditions, it can be shown that the classical ANOVA approach suffices to partition the variance associated with any set of observed measures of change into the remaining two components and their interaction. Three strategies were proposed that serve to partition the variance associated with any combination of two of the three postulated components. Figure 1

provides an illustration of minimum data sets required for each of the three paradigms (cf. also Schaie, 1983b).

Insert Figure 1 about here

The first strategy was called cohort-sequential (CS) and requires that at least two cohorts be assessed for at least two age levels (requiring a minimum of three times of measurement). Decomposition can then occur in the form,

$$Xcs_{ij} = a_i + c_j + ac_{ij} + e_{ij} \quad (4).$$

Data for the age by cohort matrix can be obtained by repeatedly measuring the same cohorts, in which case the cohort component represents individual differences while the age component and the interaction represent intra-individual change. Alternately the data matrix may be formed by the repeated assessment of successive independent random samples from the same population cohorts. In the latter case, the age component and interaction represent shifts in age-indexed population parameters.

The second strategy was termed time-sequential (TS) and requires that at least two age levels be assessed at a minimum of two times of measurement. To fill this data matrix, one would ordinarily require a minimum of four samples extending across three cohort levels. Decomposition of this matrix will be of the form,

$$Xts_{ik} = a_i + t_k + at_{ik} + e_{ik} \quad (5).$$

All components of this age by period model involve population estimates, and no estimates of intra-individual change are provided.

The third strategy, called cross-sequential (XS), requires a minimum of two cohorts assessed at two times of measurement (in this instances involving at least three age levels). This matrix can be decomposed as follows,

$$Xxs_{jk} = c_j + t_k + ct_{jk} + e_{jk} \quad (6).$$

As for the cohort-sequential strategy, it is possible to assess samples repeatedly to yield individual difference estimates between cohorts, and a period component and cohort by period interaction that estimate intra-individual change. Alternately independent random samples from successive cohorts would yield estimates that reflect population differences between cohorts and across time for all parameters.

As suggested in our earlier discussion, violation of any of the side conditions stated above would cast doubt upon the interpretation of estimates for any of the components, and violation of the equal interval assumption would further compromise interpretations regarding the relative importance of any of the components (cf. Botwinick & Arenberg, 1976). We will later on return to possible ways in which planned design mis-specifications might be used for intuitive tests of the presence and relative magnitude of confound effects (cf. Costa & McCrae, 1982; Schaie, 1982; Schaie & Hertzog, 1983; Schaie & Labouvie-Vief, 1974; Schaie & Parham, 1974, 1976). First we must deal though with a number of other matters

Sequential Designs as Quasi-experiments

Developmentalists tend to think of their designs as rather special inventions that distinguish their work from that of their more statically oriented colleagues. Nevertheless it should be recognized that all of the approaches that were deduced from what we called the general developmental model can be readily described within the Campbell and Stanley (1963) classification of quasi-experiments. One important caution is in order, however. In non-developmental research paradigms maturation is considered a threat to the validity of experiments; hence, design recommendations argue for holding maturational level constant across experimental conditions. For developmentalists, by contrast, the validity threat consists of controlling for maturation and thus being unable to assess its effects! What then is the status of our designs? Table 1 shows some previously specified equivalencies for both simple and sequential designs (Schaie, 1977) Here I will simply call attention to certain implications for design choices.

Insert Table 1 about here

Single Cohort Longitudinal Studies

We begin with the most straightforward developmental design, the single-cohort longitudinal study. This design, of course, is simply a time series where the the aging of the organism is assumed to be the "treatment." Random assignments to different levels of this treatment

cannot be made and the effects of history and testing cannot be controlled or estimated. As a consequence the utility of this design in the developmental sciences remains confined to the assessment of maturational processes that are already known to be species universals that are conducted over brief periods of time in well-controlled environments. To the best of my knowledge such requirements are reasonably met only in the study of physiological growth occurring in infancy. Longitudinal designs are ideal, however, for assessing the effects of time-dependent change over those portions of the adult life span where little or nonsystematic maturational changes is observed for many behavioral variables.

Cross-sectional Studies

In certain ways the cross-sectional approach, a separate samples design is actually somewhat less restrictive. It does control testing, instrumentation and experimental mortality. However, unless we know that cohort differences are indeed trivial for our dependent variable, we are always faced with the presence of a selection-maturation interaction that cannot be controlled or estimated. When comparisons are made across large age spans, it is likely that the variance due to cohort far exceeds that due to age. In such instances the cross-sectional design is the method of choice to identify magnitudes of generational differences. There are some instances, however, where it is reasonable to assume cohort differences to be zero, as in the comparison of closely spaced age levels in early

childhood or advanced old age. In such instances, cross-sectional designs might actually provide more credible estimates of maturational change, than do the more volatile repeated measure change scores obtained from short-term longitudinal data. Needless to say the reduction of error due to the effects of history, reactivity and practice are obtained at the price of not being able to investigate individual differences in intra-individual change.

Non-equivalent control group designs are frequently required in longitudinal (or short-term intervention) studies when it is important to compare prospective developmental effects in individuals who are members of the same birth cohort and thus of the same chronological age but who have had differential developmental histories. For example, it would make no sense to assign randomly early and late maturing children in a study that would investigate growth during late adolescence, or to assign randomly individuals who had declined on different abilities in a study that would investigate the possibility of reversing such decline on these abilities (cf. Schaie & Willis, 1986b; Willis & Schaie, 1986).

Sequential Strategies

The sequential designs may be conceptualized explicitly as quasi-experiments that control for threats to the internal validity of studies that are not directly addressed by the single cohort longitudinal or single cross-section studies. Such designs may be complex separate samples plan that control for history

(cohort-sequential with independent samples) or secular trends (cross-sequential with repeated measurements), may involve multiple time series (cohort-sequential with repeated measurements), or may represent institutional cycle designs that control for age trends (time-sequential). Certain design complications of the sequential methods, furthermore allow control and/or estimation of the validity threats due to reactivity (practice effects) and experimental mortality.

Threats to the internal and external validity of quasi-experiments, of course, are substantively driven. As stated earlier, maturation is never a threat to a study that is interested in change over time. Likewise, history or selection effects may be minimal when narrow age ranges are investigated. Practice effects may be of great concern in performance assessment but may me of minimal importance when data are obtained via behavior observations. Realizing that it is not possible to control for all validity threats in a quasi-experiment, the experienced researcher will always need to engage in a cost/benefit analysis to differentiate the crucial from the trivial validity threats. Let me stress, therefore, that the sequential strategies, are methods of choice only when it would be unwise to dismiss the validity threats they are designed to control or estimate, but not otherwise.

The side conditions for each of the sequential strategies explicated earlier also have clear implications for economy in design choice. Thus, for the estimation of population parameters, it would

be most advantageous in terms of experimenter time and efforts if it is possible to set cohort effects to zero. In that case it would be possible to use cross-sectional data to estimate age changes, and to use a time-sequential data set to differentiate age changes from secular trends. In this case, intra-individual differences in age change could be studied by a single-cohort longitudinal design, and secular trends estimated by adding an additional sample at each measurement point that would allow time-lag estimates.

If maturational effects (age) can be set to zero, then generational differences can be estimated from a single cross-sectional data set, and secular trends can be estimated from a single longitudinal data set followed over the period(s) of interest. A two-point cross-sequential study with independent samples would suffice in this case to estimate both cohort differences and period effects.

Finally, if period effects (time of measurement) are assumed to be zero, it is then possible to estimate maturational effects unambiguously within the context of a single cohort longitudinal study. Moreover, replication within the context of a longitudinal sequence would permit separation of age and cohort effects. Such a design has long thought to be of primary interest to developmentalists because it essentially permits separating intra-individual change from inter-individual differences across generations (Schaie & Baltes, 1975). The assumption of zero period effects, however, can rarely be justified with respect to behavioral variables; other conceptual

difficulties with this design have been discussed elsewhere in greater detail (Schaie, 1986a).

An Optimal Data Collection Approach

Given all the above caveats the question still remains as to whether it might be possible to recommend an optimal data collection approach that permits flexible application of sequential data analysis strategies. Noting the fact that all longitudinal studies must begin somewhere with a single first measurement occasion, I have long been convinced that it is always prudent to commence with an age-comparative cross-sectional design. However, in those instances where such design cannot answer the questions of interest it would then prove reasonable to collect additional data across time. A hypothetical data collection of this kind which I have previously designated as the "most efficient design" is depicted in Figure 2 (Schaie, 1965; Schaie & Geiwitz, 1982; Schaie & Willis, 1986a).

Insert Figure 2 about here

The "most efficient design" requires the identification of a population frame that provides a reasonable representation of the full range of the dependent variables to be studied. Optimally, the population frame should be a natural one, such as a school system, health plan, broadly based membership organization, or the like. If the population frame is reasonably large, it is then possible to

assume that members leaving the population will on average be replaced by other members with similar characteristics (sampling with replacement). An age range of interest is defined at Time 1, and is sampled randomly at intervals that are optimally identical with the time chosen to elapse between successive times of measurement. At Time 2, previous participants are retrieved and restudied, providing short-term longitudinal studies of as many cohorts as there were age intervals at Time 1. At the same time, a new random sample is drawn from the population frame over the same age intervals as in the Time 1 sampling, with one additional sample at the age level currently attained by the oldest sub-sample assessed at Time 1. The whole process can be repeated again and again with retesting of old subjects (adding to the longitudinal data) and initial testing of new samples (adding to the cross-sectional data). As will be seen, three assessment points will provide maximize design benefits, although one or two additional measurement points will allow additional design refinement. What are the analyses that become feasible as additional data is gathered over time?

Time 2 Analyses

The second data collection converts the original longitudinal study into a series of C (number of cohorts studies at T1) longitudinal studies, and the additional cross-section provides a cross-sectional sequence over two occasions. A series of C time-lag comparisons is also available. In addition, it is now possible to

examine the repeated measurement data as a cohort x time matrix allowing cross-sequential analyses. The cross-sectional sequence can be further examined as an age x time matrix (time-sequential strategy) or as an independent measures cohort x time matrix (cross-sequential strategy). A cross-sectional experimental mortality analysis can be done by comparing T1 data for those individual who were successfully reexamined and those who were not retrieved at T2.

Time 3 Analyses

The third data collection, in addition to replicating longitudinal findings upon the samples first tested at T2, permits the analysis of age x cohort data matrices for C - 1 data sets for both repeated measurements on the same subjects or independent samples from the cross-sectional sequence. In addition age x time and cohort x time matrices can be analyzed for a 3 x C data matrix. Experimental mortality analyses, classifying dropout to occur after both the first and second occasion, can now be conducted using either an age x time x dropout or a cohort x time x dropout model. Alternatively it is possible to estimate effects of practice for either an age x time x practice level or a cohort x time x practice level design (cf. Schaie & Parham, 1974).

Time 4 and 5 Analyses

Although all of the sequential paradigms can be estimated by three measurement occasions, there are some additional options

available if further extensions of the data collection are possible. A fourth measurement occasion would permit three replications of longitudinal data over one time segments, two replications over two time segments, as well as an estimate of longitudinal change over three time segments. Experimental mortality data can now be studied also for an age x cohort x dropout model, and it is possible to cross experimental mortality and practice effects within an age x time x practice x dropout or a cohort x time x practice x dropout paradigm. A fifth measurement occasion, finally, would allow estimating an age x cohort x practice x dropout model, in addition to allowing four replications of longitudinal data over one time segment, three replications over two time segments, two replications over three time segments, and an estimate of change over four time segments (cf. Schaie, 1977).

Planned Design Misspecifications

We have noted elsewhere that estimation of all three developmental components becomes feasible only if one reconceptualizes one of the component in non-calendar terms (Schaie, 1984, 1986a). When such reconceptualization is not feasible, and setting one of the components to zero may require unsupportable assumptions, it may still be possible to obtain some approximate estimates via planned design misspecifications (Costa & McCrae, 1982; Schaie & Hertzog, 1982). For example, given the availability of longitudinal data (whether repeated measurement or separate samples) over two occasions, it may be useful

to estimate a cohort x time matrix, even though the assumption of zero maturational effects is not thought to be tenable. Given equal cohort and time intervals, it can then be assumed that both cohort and time effects and their interaction will all be inflated in equal amount by the variance due to age. The difference in proportion of variance for the two components will then inform us on the relative magnitude of age differences (measured as cohort effects) and age changes (measured as period effects). Moreover, if we are willing to assume zero interaction between cohort and period, then it becomes possible to interpret the proportion of variance contributed by the interaction as a direct estimate of the age confound (Schaie, 1965).

Further conclusions may be drawn if change on one of the components fails to be significant. Such an outcome could first of all be interpreted to mean the absence of change on that component as well as on the age component thought to be confounded with it. Alternatively, it is possible that the specifically examined component and age exhibit opposite trends such that the change observed in either would be suppressed when summed. The plausibility of the second alternative could be tested further by examining an appropriate age x time matrix to discover whether age and time do indeed show significant change in opposite directions, or an age x cohort matrix to determine whether age and cohort effects suppress one another. As a logical decision rule I prefer to argue that the absence of a main effect in any of the sequential strategies would legitimize use of the design for which that effect is required to be zero. On the other

hand, finding a significant interaction would clearly cast doubt on the validity of the zero confound assumption (cf. Schaie & Parham, 1974, 1976; Schaie, Orshowsky, & Parham, 1982; Schaie & Hertzog, 1983, for empirical examples of planned design misspecifications).

Lower-bound Estimates of Developmental Components

Planned design misspecifications are useful also for the estimation of lower-bound estimates of developmental components. That is, for equal interval data sets it can be concluded that for an age x time matrix (each component being equally confounded with cohort), the excess of the age component over the age x time interaction must be a lower-bound estimate of the unconfounded age component. Similarly in a cohort x time matrix, the excess of the cohort over the time x cohort interaction would provide a lower-bound estimate of the unconfounded cohort component. The excess of the time component over its associated interaction in either design would provide the lower-bound estimated of that component in unconfounded form. Note, however, that this approach has primarily theoretical interest in that it addresses the relative contribution of developmental components as proportions of variance accounted for.

An alternate approach to the estimation of lower-bound component estimates that may have direct applicability to the construction of generalizable age gradients (to be discussed later), takes advantage of the component imbalance of typical data sets that have a large number of age/cohort levels as compared to the time of measurement

levels. As Botwinick and Arenberg (1976) have pointed out this will lead to the cumulation of the confounded variance in the effect having the larger number of levels. On the other hand the confound becomes asymptotic for the effect having only a single interval when averaged across many levels of the larger effect. This leads to the possibility of estimating relatively unbiased period effects, generalized over the entire age/cohort range under consideration, that can then be used to construct corrected age gradients that are more likely to be generalizable beyond cohorts and historical period of the data from which they were generated. If the same cohorts have been assessed at multiple time points, it is then also possible to obtain similar corrections for average cohort levels.

Some Prescriptions for the Construction of Age Functions

As suggested above, the ultimate objective of most descriptive studies of age-related development is to arrive at some descriptive function that will provide indications of the level and slope of changes in the dependent variable across the age range of interest. Such functions are not only of theoretical interest, but are also useful in providing the kind of data that may be relevant for the determination of social policies in areas such as age discrimination in employment, mandatory retirement, and the determination of levels of competence required for independent function at developmental transition stages such as adolescence and the onset of advanced old age. There are actually three different topics that should be

addressed in this context. I will first discuss specific prescriptions for constructing generalizable age gradients, will next discuss what I think may be useful ways of legitimately contrasting cross-sectional age difference data with longitudinal age change data, and finally will deal with the need to responsibly deal with the issues of individual differences and overlap of samples that may show statistically significant mean differences.

Constructing Corrected Age Gradients

Estimates of average age change within individuals that are to be used as the basis of policy recommendations or for generating normative data that can be used to assess whether the change observed in a particular individual is within the average range of individual differences are best based by averaging change data over as many cohorts and times of measurements as possible. The age gradient is then constructed by adding the average amount of change over each age segment for which longitudinal data are available, and adding these amounts, successively to the base value for the youngest sample. The validity of within-cohort estimates of ontogenetic age changes can, however, be questioned by arguing that such data may be specific to the possibly atypical performance of cohorts from which the data were gained. A second validity threat refers to the possible maximization of idiosyncratic period effects. And thirdly, the unavoidable attrition in longitudinal panels might restrict generalizability of findings only to the favorably endowed subset remaining in a given

study. The first two validity threats primarily affect estimated magnitudes of change, the third has relevance for specifications of performance levels in relation to a base age (cf. Schaie 1983a).

Adjustment of attrition effects. Due to selective mortality and factors such as declines in health and motivation most longitudinal studies experience selective attrition leading to the loss of the less able and/or socially responsible panel manners (cf. Schaie, Labouvie, & Barrett, 1973). As a result retest survivors tend to score higher at first test than individual who drop out. Gradients based on retest survivors will therefore often lead to more favorable conclusions as to the magnitude of change as a function of base levels than may be warranted. Longitudinal estimates, therefore, should routinely be adjusted by assigning values to base levels that ware obtained from the full rather than the attrited sample.

Adjustment for period effects. Longitudinal estimates of change, as indicated earlier, confound age and period effects. The period effects may either attenuate or exacerbate the age changes, and should therefore be adjusted for better generalizability if relevant data are available. The least biased estimates of period effects are obtained by averaging time-lag differences between successive occasions across all available cohorts equated for chronological age. The estimates of age change are then adjusted by subtracting the relevant period effects weighted according to the size of sub-samples involved for each period.

Adjustment for cohort effects. To the extent that age change estimates were obtained by averaging across several cohorts, this correction may be excessive for some and insufficient for other cohorts, and a further adjustment is therefore required. This adjustment involves estimating average cohort differences across all available times of measurement, and to add an adjustment weighted according to the size of sub-samples belonging to each cohort.

It should be noted that a similar approach can be taken to the construction of cohort gradients that will alert us to the relative advantage or disadvantage of successive cohorts on dependent variables of interest. Such cohort gradients must also be corrected for the fact that the data entering our computations for different cohorts will be impacted both by idiosyncratic period effects as well as by the effect that cohort difference estimates perforce will have been obtained for different cohorts at different age levels.

Comparison of Age Change and Age Difference Data

The effect of the adjustments just described is to generate age gradient that will represents age change patterns that would prevail if there was no attrition, and equal contributions of period and cohort effects over all age segments. While this approach is advocated for the construction of age gradients that will have greatest possible generalizability, a somewhat different paradigm is needed for a direct comparison of age change and age difference data. The principal issue in this comparison, of course, is to compare the

pattern of age changes over the age range of interest within a particular age cohort, with the age differences observed at a particular point in historical time for groups defining the same age range. It should be noted that a direct comparison of cross-sectional and longitudinal data is practical only over short time spans. Retrospective data for a lengthy longitudinal study would certainly

n

Treatment of Individual Differences in Developmental Studies

A final concern to be discussed concerns the common misuse of descriptive findings of statistically significant age differences to imply that such difference are of universal prevalence. I would like to call attention to the fact that past adolescence and until advanced old age is reached, statistically significant age differences typical represent rather modest effect sizes. What this means then, is that there will be substantial overlap in the score distributions for successive age groupings, with many individuals in the group showing lower performance averages scoring above the mean of the group with the higher average. This issue is particularly salient when reporting average age decrements. Here it would be useful to specify the proportion of overlap between groups when cross-sectional data are reported, or in longitudinal studies indicate the proportion of study participants whose performance has remained stable, has declined or has improved. Such differentiated reporting would be most useful for the application of developmental data to policy issues. Of course, suitable confidence bands must be specified for such analysis. The criteria for such confidence bands that I would propose to use, would at the group level require that there be less than 50 per cent overlap between groups; that is, the mean of the group to be compared would be below the value of the first quartile of the reference group. At the individual level a reasonable confidence band for reliable longitudinal change would seem to be provided by a standard set at 1 SEM about the base score (also see Schaie, 1984).

Concluding Remarks

I have attempted in this presentation to sketch a brief history of my attempts to systematize and improve the methodological rigor with which developmentalists describe changes in behavioral phenomena that occur over age and time. I have consciously restricted my remarks primarily to issues that effect changes and differences in level of function, since several other speakers at this conference will address the issue of structural change. I have reiterated some of the fundamental aspects of a basic model that describes components of developmental change that have retained value over time. I have also tried to show how the role of developmental studies as quasi-experiments demands certain design complications. As part of such complications I have discussed some of the decisions one has to make with respect to which threats to the validity of a study must be controlled or estimated and which others can be dealt with by means of specifying relevant and reasonable assumptions. Finally, I have shared with you a number of approaches that I have used to present the results of developmentally oriented studies in ways that may make their results relevant to policy issues in a comprehensible and responsible manner. Those of you who attended the 1972 life-span methodology conference will probably have noticed that many of the issues I have discussed today are not all that different from those that I tried to address at the earlier conference. At that time, however, many of my concerns arose from arm chair speculation while today's remarks are at least seasoned by having been tested by

empirical applications. I hope you will find some of the resultant experiences that I shared with you useful for your own work.

References

Botwinick, J., & Arenberg, D. (1976). Disparate time spans in sequential studies of aging. Experimental Aging Research, 2, 55-61.

Campbell, D.T., & Stanley, J.C. (1963). Experimental and quasi-experimental designs for research in teaching. In N. L. Gage (Ed.), Handbook of research on teaching (pp. 171-246). Chicago: Rand McNally.

Costa, P. T., Jr., & McCrae, R. R. (1982). An approach to the attribution of aging, period, and cohort effects. Psychological Bulletin, 92, 238-250.

Hertzog, C., & Schaie, K. W. (1986). Stability and change in adult intelligence: I. Longitudinal covariance structures. Psychology and Aging, 1, in press.

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

Schaie, K. W. (1973). Methodological problems in descriptive developmental research on adulthood and aging. In J. R. Nesselroade and H. W. Reese (Eds.), Life-span developmental psychology: Developmental issues (pp. 253-280). New York: Academic Press.

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

Schaie, K. W. (1978). External validity in the assessment of intellectual development in adulthood. Journal of Gerontology, 33, 695-701.

Schaie, K. W. (1979). 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 (pp.67-115). New York: Academic Press.

Schaie, K. W. (1982). Longitudinal data sets: Evidence for ontogenetic development or chronicles of cultural change? Journal of Social Issues, 38, 65-72

Schaie, K. W. (1983a). The Seattle Longitudinal Study: A twenty-one year exploration of psychometric intelligence in adulthood. In K. W. Schaie (Ed.), Longitudinal studies of adult psychological development (pp. 64-135). New York: Guilford Press.

Schaie, K. W. (1983b). What can we learn from the longitudinal study of adult psychological development? In K. W. Schaie (Ed.), Longitudinal studies of adult psychological development (pp. 1-19). New York: Guilford Press.

Schaie, K. W. (1984a). Historical time and cohort effects. In K. A. McCluskey and H. W. Reese (Eds.), Life-span developmental psychology: Historical and generational effects (pp. 1-15). New York: Academic Press.

Schaie, K. W. (1984b). Midlife influences upon intellectual functioning in old age. International Journal of Behavioral Development, 7, 463-478.

Schaie, K. W. (1986). Beyond calendar definitions of age, time and cohort: The general developmental model revisited. Developmental Review, in press.

Schaie, K. W., & Baltes, P. B. (1975). On sequential strategies in developmental research. Human Development, 18, 384-390.

Schaie, K. W., & Geiwitz, J. (1982). Adult development and aging. Boston: Little, Brown & Co.

Schaie, K. W., & Hertzog, C. (1982). Longitudinal methods. In B. B. Wolman (Ed.), Handbook of developmental psychology (pp 91-115). Englewood Cliffs, NJ: Prentice-Hall.

Schaie, K. W., & Hertzog, C. (1983). Fourteen-year cohort-sequential studies of adult intelligence. Developmental Psychology, 19, 531-543.

Schaie, K. W., & Hertzog, C. (1985). Measurement in the psychology of aging. In J. E. Birren & K. W. Schaie (Eds.), Handbook of the psychology of aging (2nd ed., pp. 61-92). New York: Van Nostrand Reinhold.

Schaie, K. W., & Hertzog, C. (1986). Toward a comprehensive model of adult intellectual development: Contributions of the Seattle Longitudinal Study. In R. J. Sternberg (Ed.), Advances in human intelligence (Vol. 3). New York: Academic Press.

Schaie, K. W., & Labouvie-Vief, G. (1974). Generational versus ontogenetic components of change in adult cognitive behavior: A fourteen-year cross-sequential study. Developmental Psychology, 10, 305-320.

Schaie, K. W., Labouvie, G. V., & Barrett, T. J. (1973). Selective attrition effects in a fourteen-year study of adult intelligence. Journal of Gerontology, 28, 328-334.

Schaie, K. W., Orshowsky, S. J., & Parham, I. A. (1982). Measuring age and socio-cultural change: The case of race and life satisfaction. In R. C. Manuel(Ed.), Minority aging: Sociological and social psychological issues (pp. 223-230). Westport, CT: Greenwood Press.

Schaie, K. W., & Parham, I. A. (1974). Social responsibility in adulthood: Ontogenetic and sociocultural changes. Journal of Personality & Social Psychology, 30, 483-492.

Schaie, K. W., & Parham, I. A. (1976). Stability of adult personality traits: Fact or fable? Journal of Personality & Social Psychology, 34, 146-158.

Schaie, K. W., & Strother, C. R. (1968). A cross-sequential study of age changes in cognitive behavior. Psychological Bulletin, 70, 671-680.

Schaie, K. W., & Willis, S. L. (1986a). Adult development and aging. 2nd ed. Boston: Little, Brown & Co.

Schaie, K. W., & Willis, S. L. (1986b). Can decline in adult cognitive functioning be reversed? Developmental Psychology, 22, in press.

Willis, S. L., & Schaie, K. W. (1986). Training the cognitive abilities of spatial orientation and inductive reasoning in the elderly. Psychology and Aging, 1, in press.

Author Note

Many colleagues and graduate students have had marked influence on my contributions to developmental research methodology. I would like to recognize in particular the influence of Paul Baltes, Christopher Hertzog, John Nesselroade, Charles Strother, and Sherry Willis. Current work on these issues has been supported by research grants AG03544 and AG04770 from the National Institute on Aging.

TABLE 1. Classification of Quasi-Experimental Designs in Terms of Familiar Developmental Paradigms

Quasi-experimental Design	Equivalent Developmental Paradigm	Applicability
1. Time series	Single cohort longitudinal	Limited to species universals
2. Equivalent time samples	None	Not applicable
3. Equivalent materials samples	Single cohort longitudinal with alternate forms for each measurement	Same as 1
4. Non-equivalent control group design	Time-sequential	Differentiates age and secular trends
5. Counterbalanced designs	None	Not applicable
6. Separate-sample	Cross-sectional	Most suitable for identifying generation differences
a) Controlled for history	Cohort-sequential with independent samples	Differentiates age and cohort differences, controlling for history, testing and reactivity
b) Controlled for secular trends	Cross-sequential with repeated measurement	Differentiates cohort and secular trends
7. Separate sample pretest-posttest control group design	Cross-sequential with independent samples	Differentiates cohort and secular trends controlled for testing and reactivity
8. Multiple time series design	Cohort-sequential with repeated measurements	Differentiates age and cohort effects controlled for history
9. Institutional cycle design	Time-lag or time-sequential	Assesses secular trends and/or differentiates age trends
10. Regression discontinuity	Functional age analysis	Developmental criteria other than age

Figure 1 Schematic showing cross-sectional and longitudinal sequences and the modes of analysis deduced from the general developmental model. Table entries represent times of measurment (periods). From Schaie (1983b).