PART I

Developmental Designs

CHAPTER ONE

Issues in the Use of Longitudinal and Cross-Sectional Designs

Kelly Robinson Todd Schmidt and Douglas M. Teti

Baltes, Reese, and Nesselroade (1977) defined the task of developmental science as "*the description, explanation, and modification (optimization) of intraindividual change in behavior and interindividual differences in such change across the life span*" (p. 84, italics in original). This task was embraced by many over the last 100 years, and indeed the discipline has yielded a wealth of knowledge about the physical, cognitive, emotional, and social development of individuals across the life span.

Developmental research has traditionally been conducted using one of two methodologies. One involves the repeated measurement of a sample of individuals, usually at the same age at the start of the study, over a period of time, termed a *longitudinal* study. The "task" in longitudinal studies is to find meaningful associations between age changes and changes in specific outcome behaviors or abilities of interest. The second involves the measurement of several samples of differing ages simultaneously, termed a *cross-sectional* study, in which the task is to discover age group differences in particular behaviors or abilities.

This chapter reviews these two approaches from the vantage point of the general developmental model, discusses the advantages and pitfalls of each, and highlights exemplars of each from the developmental literature.

The Developmental Function and the General Developmental Model

Wohlwill (1970a) defines a variable as "developmental" when it changes with age in a generally uniform and consistent way across individuals and environments. He asserts that our interest should not be in looking for significant age-related differences but in discovering the nature of the age function – its shape and form. The developmental function is defined as "the form or mode of the relationship between the chronological age of the individual and the changes observed to occur in his responses on some specified dimension of behavior over the course of his development to maturity" (Wohlwill, 1973, p. 32). Wohlwill (1973) asserted that extending the concept of the developmental function to the whole life span was not useful because of the challenge of studying the life span longitudinally, the lack of measurable change in some aspects of behavior at maturity, and the difficulties in ascertaining the onset of aging. There are many life-span developmental researchers, however, who take issue with this premise and have conducted interesting, valuable research on "mature" individuals (e.g., Schaie, 1996; Schaie & Caskie, this volume; Siegler & Botwinick, 1979).

The parameters that make up the developmental function were explicated in a seminal paper by Schaie (1965), who is widely credited with laying out the paradigm of developmental research that has shaped research for over three decades and continues to do so. The three parameters that define developmental change according to Schaie's (1965) general developmental model are *age*, *cohort*, and *time of measurement* (also called *period*). *Age* is commonly defined as chronological age; this definition is not without some controversy, as will be discussed later. *Cohort* is defined as a group of individuals experiencing an event or set of events associated particularly with that cohort (a cohort-defining event) (Mayer & Huinink, 1990). The most frequently used cohort-defining event is the birth of an individual. Time of measurement is most typically defined according to calendar time, although this definition too has been questioned by some (e.g., Schaie, 1986).

According to the model, different developmental research designs can be seen as combinations of the three variables. Simple longitudinal and cross-sectional designs are defined by the ages of interest to the researcher, the cohort(s) from which the sample is drawn, and the time or times of measurement. More complex developmental designs are proposed under the model, but these are discussed by Schaie and Caskie in Chapter 2. When used in cross-sectional research, the age variable really taps *interindividual differences*. When used in longitudinal research, it taps *intraindividual change* (Schaie, 1983, 1984, 1986). The cohort variable is an *individual differences* variable, while time of measurement or period is an *intraindividual change* variable.

These three variables by definition are not independent; that is, once two of these three parameters are determined, the third is automatically defined (Baltes, 1968; Schaie, 1965, 1986). This means that age – the variable most often of interest to developmental researchers – is always inevitably confounded with either cohort or time of measurement. Schaie (1986) pointed out that this is because we tend to define each variable in terms of calendar time, and proposed designs for unlinking calendar time from the variables in order to get at the independent effects of them (e.g., defining cohort more broadly than time of birth) (see Chapter 2, this volume). These designs, such as cross-sequential and cohort-sequential designs, appear infrequently in the developmental literature.

Schaie's general developmental model has been subjected to much criticism over the years (e.g., Baltes, 1968; Baltes, Reese, & Nesselroade, 1977). Strict adherence to it requires the researcher to make some perhaps untenable assumptions, such as assuming that one variable in the model has no effect on the dependent variable (Schaie, 1986). Another limitation of the model is that it assumes that change occurs incrementally over time with age in a linear fashion (Kosloski, 1986). This in fact may not be true for many developmental functions such as personality traits. Finally, some have argued that the model is really only useful for describing change, not for explaining it (Baltes, Reese, & Nesselroade, 1977).

In spite of these criticisms, the general developmental model has spurred much thought about how development should be studied. Because age, cohort, and time of measurement serve as proxies for other causal variables (Hartmann & George, 1999), the model has forced researchers to think more creatively and complexly about developmental processes. The goal for many developmental researchers is to understand the contribution of age to the developmental function, but it should be clear that researchers need to investigate the contributions of age, cohort, and time of measurement to the developmental function because they are inextricably linked.

Simple Cross-Sectional Designs

The simple cross-sectional study consists of at least two samples of different ages drawn from different cohorts and measured simultaneously. For example, a researcher might want to examine the social strategies used to enter a group of children by 6-, 8-, and 10-year-olds. This research approach stems from the assumption that when an older age group is drawn from the same population as a younger age group, the eventual behavior of the younger group can be predicted from the behavior of the older group (Achenbach, 1978). Thus, a researcher can examine the relationship between earlier and later behavior without actually waiting for development to occur (Achenbach, 1978). Longitudinal conclusions are typically drawn from cross-sectional data, but the validity of this is questionable (Achenbach, 1978; Kraemer et al., 2000).

Cross-sectional studies are relatively inexpensive, quick and easy to do, are useful for generating and clarifying hypotheses, piloting new measures or technology, and can lay the groundwork for decisions about future follow-up studies (Kraemer, 1994). They provide information about *age group differences* or *interindividual differences* (Miller, 1998). They do not, however, provide information about *age changes* or *interindividual differences in intraindividual change* (Miller, 1998; Wohlwill, 1973). That is, the results of the above-mentioned study on socialization might reveal differences among 6-year-olds, 8-year-olds, and 10-year-olds, but they would not inform us of how and when these differences emerge and how the behaviors evolve over time.

Cross-sectional studies are subject to many methodological concerns and limitations. They cannot answer questions about the stability of a characteristic or process over time (Miller, 1998), and information is lost because of the use of averages to create group means (Wohlwill, 1973). A researcher planning to conduct a cross-sectional study needs to ensure at the outset that the measurement instruments (e.g., personality tests, intellectual assessments, etc.) he/she plans to use measure similar things at each age and are valid for each age under investigation (Miller, 1998). Another criticism of cross-sectional studies is that their external validity (i.e., generalizability) is possibly affected by historical/cultural differences between cohorts (Achenbach, 1978). For example, if one were studying the development of some reading behaviors, the comparability of a first grade class and a third grade class within the same school would be compromised if the first graders were exposed to a new reading curriculum that the third graders never experienced. This would represent a historical event that renders the cohorts non-equivalent. This problem, termed the age by cohort confound, is perhaps the most serious limitation of the cross-sectional design; that is, one cannot easily separate the effects of age from the effects of belonging to a particular cohort, especially if that cohort is defined by birth. Miller (1998) argues that the seriousness of this problem relates to the dependent variable: the more "basic" or "biological" the variable (e.g., heart rate, visual acuity), the less likely it is that the cohort effect will be present. It also depends on the age span of the sample: the wider the spread, the more likely a cohort effect could be operating. This is particularly problematic in aging or life span research.

Another major risk of the cross-sectional method is that the researcher will unwittingly create bias in the samples through flawed selection procedures, especially if random assignment to age groups is not possible (Baltes, Reese, & Nesselroade, 1977; Flick, 1988; Hertzog, 1996; Kosloski, 1986; Miller, 1998; Wohlwill, 1973). Traditional experimental research methods (e.g., Cook & Campbell, 1979) mandate the formation of groups that are identical except for the variable of interest, which in this case is age. Matching on variables other than age could result in a non-representative sample; for example, if one were comparing 25-year-olds and 75-year-olds, matching on educational level (e.g., college graduate) would yield a positively biased sample of 75-year-olds. That is, 75-year-old college graduates would be less representative of their age cohort in terms of education, than would the 25-year-olds of their age cohort. Furthermore, if the entrances and exits of individuals from the sampling population are not random, then the researcher is at risk for making incorrect inferences about the developmental process under investigation (Kraemer et al., 2000). For example, if one is interested in the relation between age and the move toward assisted living, one must account for the fact that entering or exiting an assisted living facility is not a random occurrence but is most likely related to factors associated with age. Adopting a cross-sectional approach to studying this developmental process would not permit the identification of predictors associated with moving into assisted living, whereas adopting a longitudinal approach would allow such analyses.

The basic premise for using the cross-sectional approach is that we can draw conclusions about intraindividual age-related changes from observing interindividual differences. This requires the strong assumption that participants in all comparison groups are equivalent in all respects save chronological age. Indeed, it is commonly held that the longitudinal inferences drawn from cross-sectional research are not seriously misleading, when in fact this might not be valid (Hertzog, 1996; Kraemer et al., 2000). One's ability to draw inferences from cross-sectional research is affected by factors such as how time is measured, the type of developmental trajectory of the developmental process (i.e., fixed trait, parallel trajectories, or nonparallel trajectories), the reliability of measurement, and the time of measurement (i.e., fixed or random for all subjects) (Kraemer et al., 2000). Furthermore, Kraemer and colleagues (2000) suggest that cross-sectional research done as pilot studies for subsequent longitudinal studies in fact might actually serve to discourage longitudinal research because they intimate that the answers are already known.

Examples of cross-sectional studies

Flavell, Beach, and Chinsky (1966) employed the cross-sectional design in a study which examined the use of verbal rehearsal strategies for a memorization task among children at three ages: kindergarten, second grade, and fifth grade. Ten boys and girls of each age were matched on grade and sex, and were instructed to remember the order of pictures presented. The children wore a "space helmet" with a visor that allowed the experimenters to watch the children's mouths. The study revealed that most kindergartners did not use verbal rehearsal strategies, while most fifth graders did. This study therefore generated intriguing hypotheses about the development of memory strategies during middle childhood.

Gopnik and Astington (1988) examined the apparent developmental changes in representational thought in 3-, 4-, and 5-year-old children. In one experiment they used deceptive objects such as a candy box containing pencils and asked the children to guess the contents of the box before opening it. Once the surprising contents were revealed, the children were asked what they thought was in the box before it was opened. The youngest children tended to maintain that they knew pencils were in the box, even though they guessed "candy" earlier, while the older children demonstrated some awareness of the appearance/reality distinction. The experimenters also had the children complete a false belief task in which they asked the children, "X has not seen this box, what will s/he think is in the box?". Again, the younger children incorrectly stated that X would think pencils were in the box, while older children tended to correctly recognize that the appearance of the box would lead one to think candy was in the box.

Thus, cross-sectional research designs can be quite useful in their ability to demonstrate age group differences in developmental processes such as cognition and memory, but it is essential that one remember that inferences about *how* and *when* these changes emerge and evolve over time are impossible to make. Furthermore, the age by cohort confound makes untangling the independent effects of each variable difficult.

Simple Longitudinal Designs

An obvious solution to the shortcomings of the cross-sectional research strategy would appear to be a strategy in which a sample of participants of a given age and from a given cohort were observed over a period of time – that is, employing the longitudinal research design. As Campbell (1988, p. 43) noted, "There are few issues that evoke greater agreement among social scientists than the need for longitudinal as opposed to cross-sectional studies."

Miller (1998) defined longitudinal designs as "repeated tests that span an appreciable length of time" (p. 27). The notion of "repeated tests" is not well defined, and frequently seems to be conceived of as two occasions, which has been found questionable by some (e.g., Rogosa, 1995). The concept of "appreciable length of time" appears to vary with the developmental level of the sample. For example, one week between testing does not likely constitute a longitudinal study for a 5-year-old, but might for a newborn.

Longitudinal designs are useful and necessary in that they allow us to focus on intraindividual change, developmental sequences, and co-occurring social and environmental change that enable one to develop theoretical/explanatory accounts of whatever change occurs (McCall, 1977). They are perhaps most valued due to the fact that they permit a direct measure of age changes – intraindividual development over time (Farrington, 1991; Miller, 1998). Also, the researcher can examine interindividual differences in intraindividual change (Baltes & Nesselroade, 1979). Longitudinal designs permit the investigation of individual consistency or change and let the researcher look at early–later relationships (Farrington, 1991; Miller, 1998; Wohlwill, 1973). Longitudinal studies allow construction of the shape of the developmental function and let the researcher examine differences between individuals in terms of the entire developmental function, not just at a particular age (Wohlwill, 1970b, 1973).

The researcher conducting a longitudinal study can explore the causes of intraindividual change because this methodology meets one necessary, but not sufficient, criterion for making causal inferences: time ordering (Baltes & Nesselroade, 1979; Campbell, 1988; Farrington, 1991; Pellegrini, 1996; Wohlwill, 1973). That is, one can examine antecedents and consequences and make some reasonable speculations about causality. However, Schaie (1988) noted that, although observations in a longitudinal study are by definition time ordered unidirectionally, this does not mean that time-ordered change is unidirectional. Although many developmental processes may be unidirectional over certain time periods of the life span, others are likely to be cyclical or recursive.

Threats to Validity in Longitudinal Research

In spite of the apparent benefits of the longitudinal research strategy, it is expensive, time consuming, and labor-intensive. Furthermore, longitudinal research designs are quite vulnerable to many of the threats to validity commonly associated with quasi-experimental research, namely selection, attrition, instrumentation, and regression to the mean (Shadish, Cook, & Campbell, 2002). The process of assembling an appropriate sample for a longitudinal study is no easy task. Sampling depends upon whether the researcher is doing a prospective or a retrospective study. In a prospective study, the sample is constructed based on the independent variable (e.g., if one were interested in studying the long-term effects of prenatal exposure to alcohol, one would recruit newborns exposed to alcohol in utero), while in a retrospective study it is assembled based on the dependent variable (e.g., if one were interested in assessing the long-term effects of an early intervention program, one could recruit graduates of a Head Start program) (Jordan, 1994). Also the researcher must decide upon which type of population from which to sample: a normal representative population (e.g., birth cohort, school and adult cohorts, community cohorts) or nonrepresentative population (e.g., specialized cohorts such as twin, adoptees, identified patients, etc.) (Mednick, Griffith, & Mednick, 1981). The advantages of using a representative sample are the increased generalizability of the findings, the ability to study a variety of phenomena (e.g., social and medical variables), and the ability to obtain incidence and prevalence data on diseases and illnesses (Baltes, Reese, & Nesselroade, 1977; Goldstein, 1979; Mednick, Griffith, & Mednick, 1981; Schaie, 1977; van der Kamp & Bijleveld, 1998). However, a representative sample can become less representative over time; that is, a population may change over time so that a sample studied at Time 2 may no longer be representative of the population as it was at Time 1 (Baltes, Reese, & Nesselroade, 1977; Goldstein, 1979; Mednick, Griffith, & Mednick, 1981; Schaie, 1977; van der Kamp & Bijleveld, 1998).

It is desirable to obtain a sample that is readily available and cooperative over the length of the study, but this is quite challenging for most researchers, and doing so may in fact create sampling bias (Achenbach, 1978; Miller, 1998). One can do screenings at the outset of a study to maximize the possibility of obtaining a sample that is high in cooperation and stability, but the sample might become biased in the process (Jordan, 1994). However, noncooperative and mobile families/individuals are likely to be quite different from cooperative stationary ones, so including them in a sample might bias the sample anyway (Jordan, 1994).

Thus, the researcher must take into account the problems associated with constructing and maintaining a sample over the course of a longitudinal study. The researcher planning to do a longitudinal study must decide between selecting a large sample for which less detailed information can be collected and to which less time and effort can be devoted to reducing attrition, or a smaller sample where external validity may be compromised but in which one can devote greater effort and time to obtaining more detailed, process-oriented data (Bergman & Magnusson, 1990).

Sample attrition is probably one of the most common and frustrating problems faced by longitudinal researchers. Attrition is problematic in that non-responders usually differ from responders in ways that might be related to the variables being studied (Bergman & Magnusson, 1990; Goldstein, 1979). It can also be problematic for the researcher studying several groups over the same period of time, such as in a treatment study, when participants in one group drop out at a higher rate than those in other groups (Miller, 1998). Goldstein (1979) advises researchers to plan for attrition and thus plan to trace subjects. Jordan (1994) recommends obtaining the name, address, and phone number of a relative most likely to know the participant's address in the future as a way to prevent attrition. Another suggestion for situations in which a participant has moved away is to enlist the help of a colleague in that area to conduct any testing or interviewing (Jordan, 1994). Thus the researcher can potentially control some types of attrition, such as that caused by lack of interest, relocation, or active refusal, but cannot control factors related to age such as physical decline which may impede participation (Schaie, 1977).

The missing data that is a result of attrition is problematic for the longitudinal researcher. One can use data collected on earlier occasions to make inferences about nonresponders (Goldstein, 1979), and Flick (1988) reviews a number of statistical solutions to the problem of missing data. Jordan (1994) asserts that it is not necessarily true that a missing subject is missing forever, since he or she may be recovered at a later point in time, if the researcher plans ahead to allow for such situations. It might be well worth the effort to try to recover subjects lost at one point in time, as can be seen in the results of a longitudinal study of adult intellectual development conducted by Siegler and Botwinick (1979). Adults between 60 and 94 years of age participated in 11 test sessions over 20 years, beginning in 1955 and ending in 1976. Significant attrition occurred over the course of the study as expected due to illness and death, but study procedural requirements also contributed to attrition because subjects who did not complete the entire test battery at a session were eliminated, and furthermore a subject's score at any one point in time could only be counted if he or she had been tested at each previous time. Siegler and Botwinick (1979) graphed IQ scores at time 1 by test session and found that IQ scores were dramatically higher for those who completed more sessions than for those who completed less, so that those who completed the 11 sessions appeared to have higher intellectual ability than those who dropped out. Thus the intellectually superior participants made it through the procedural requirements of test completion and attendance at all sessions. The researchers concluded that inferences about adult intellectual development made on results such as these could be quite misleading.

Another potential threat to the validity of a longitudinal study is *testing*, where performance by participants is enhanced due to practice or familiarity with the measurement tools and/or procedures. It has been noted that participating in a longitudinal study may actually change the course of growth and development because of a heightened awareness of the phenomena under investigation (Goldstein, 1979). The presence of a testing effect means that a sample has become less representative of the underlying population (Schaie, 1977). When examining the data for practice effects, one must first take into account attrition (Schaie, 1996). Practice effects can be lessened by the use of alternate forms and nonreactive measures (Wohlwill, 1973), and may be less of a problem for developmentally less mature participants. Miller (1998), for example, points out that testing is not likely to be a significant problem in infant research.

Instrumentation represents yet another validity threat. Often it is the case that researchers need to use different instruments at different ages, and one cannot assume that the same phenomena is being measured at each time (Baltes, Reese, & Nesselroade, 1977; Goldstein, 1979; Schaie, 1977). Even if the same instrument is used, interpretations of the results at each time of measurement could be different (Goldstein, 1979). Alternatively, an assessment tool might be appropriate across the age range under investigation but be cohort-specific, which limits generalizability (Schaie, 1977). Schaie (1988) suggests that the problem of measurement equivalence over time could be due to developmental discontinuities of the behavior in question (see Hartmann, this volume). Even if a variable can be measured identically at two points in time, however, it is still likely that the distribution of scores will change over time, which means that the *meaning* of a score may change (Achenbach, 1978).

Other instrumentation-related problems include the fact that tests and instruments are susceptible to "aging" and might even become obsolete, or at least undergo changes in validity and/or reliability (Baltes, Reese, & Nesselroade, 1977; Jordan, 1994; Mednick, Griffith, & Mednick, 1981; Miller, 1998; Schaie, 1977). Study personnel might also change significantly over the course of a longitudinal study, affecting the overall procedures of the study and the administration of the assessment tools (Jordan, 1994). The researcher should keep in mind that personnel will likely grow in psychometric skills over the course of a study (Jordan, 1994). In addition, definitions and measurement of the independent and dependent variables will likely change over time (Achenbach, 1978; van der Kamp & Bijleveld, 1998). Moreover, theories and hypotheses might become outdated as a study progresses and thus requires reformulation in light of findings along the way or other data from outside sources (Farrington, 1991; Jordan, 1994; Mednick, Griffith, & Mednick, 1981).

Regression to the mean is yet another potential problem in longitudinal research. Buss (1979) points out that repeated measures on the same variable introduce the possibility of regression toward the mean, especially when sampling extreme scores in a population, and researchers should seek to separate out true score changes from measurement error. Regression to the mean is also observed when error variance decreases over time (i.e., as reliability increases), so that researchers should also examine variance over time (Buss, 1979). According to Baltes and colleagues (1977), regression to the mean is mostly a problem when a sample is observed on only two occasions and when the sample is divided into subgroups along a continuum. One way to minimize regression to the mean as well as testing and instrumentation threats to validity is to draw independent samples at each time of testing (Kosloski, 1986).

The simple longitudinal research design is also susceptible to a *cohort effect*, which might impact both the internal and external validity of a study (Achenbach, 1978; Baltes, Reese, & Nesselroade, 1977; Bergman & Magnusson, 1990). That is, a particular cohort under investigation may have some unique characteristics or experience some unusual event that makes it unlike another cohort of the same age. This problem can be partially mitigated by obtaining some cross-sectional data on relevant variables (Bergman & Magnusson, 1990).

Finally, the validity of a simple longitudinal research study can be threatened by the age by time of measurement confound. In fact, Schaie (1977, 1983) suggests that this threat is most likely to impact such a study because it consists of only one cohort and thus makes separating out the independent effects of age and time of measurement impossible. For example, a researcher studying anxiety during adolescence would have to take into account time of measurement issues if he/she was collecting data both before and after the terrorist incidences of September 11, 2001; age alone could probably not explain any developmental changes seen among this cohort of adolescents. Following Miller (1998), the age by time of measurement confound is likely to be less serious when more "basic" or "biological" variables are under consideration, but perhaps more serious with outcome variables that are likely to be influenced by historical events that co-occur with age (e.g., attitudes about risk-taking).

Conceptual and Planning Considerations in Longitudinal Research

Friedman and colleagues (1994) argue that longitudinal research, especially follow-up research in which some sample is studied after completing a treatment or intervention, has tended to be atheoretical and driven largely by the availability of assessment tools such as IQ tests, rather than by theory or methodological considerations. They point out that IQ tests historically were intended to be predictor rather than criterion variables, which is not how they are frequently used. In addition, they assert that the choices made about sources of data seem to be determined by factors other than methodological considerations.

Indeed, the conceptualization, methodology, and data analyses of longitudinal studies need to be tightly linked, but this frequently does not happen in large-scale studies (Campbell, 1988). Several writers advise the researcher to be broad-minded and eclectic when developing theories and choosing measures and to plan for studies to be multipurpose and multidisciplinary (Bergman & Magnusson, 1990; Mednick, Griffith, & Mednick, 1981; Mednick, Mednick, & Griffith, 1981). Adopting such an approach will likely minimize the problem of fading relevancy (Bergman & Magnusson, 1990). The researcher must anticipate the possibility, however, that the methods of measurement, scales of measurement, and the meaning of scores will change over time during the course of the study when looking for patterns of stability and change (Achenbach, 1978).

Many researchers offer practical suggestions with regard to planning and implementing a longitudinal study. It behooves the researcher planning a study to be flexible, especially in light of the possibility of sleeper effects (i.e., effects that emerge a considerable time later) or simply the length of time it takes for some phenomena to manifest themselves. This is particularly applicable to longitudinal research involving infants (Mednick, Mednick, & Griffith, 1981). Longitudinal studies require more careful planning than cross-sectional research studies as well as consistent funding over time and a major time commitment from the head researcher and other personnel, which makes such undertakings demanding. Many note that the expense and time investment required of a well-done longitudinal study is commonly a deterrent to such an endeavor (Bergman, Eklund, & Magnusson, 1991; Miller, 1998; Wohlwill, 1973). With regard to expense, Mednick, Griffith, and Mednick (1981) point out that a longitudinal study may especially be expensive initially as staff training and the purchase of new equipment and materials are required. The researcher committed to conducting a longitudinal study must be willing to cope with a slow rate of return on the amount of work invested (Wohlwill, 1973) and also realize that no researcher can study across the life span (Baltes, Reese, & Nesselroade, 1977).

Planning involves theory, organization, and administration of the study (Bergman, Eklund, & Magnusson, 1991). Jordan (1994) notes that planning is essential, because an enormous amount of data will be collected that need to be processed. It is important for the researcher to pay very close attention to data collection and storage, and to take advantage of opportunities to collect additional data on subjects (Mednick, Griffith, & Mednick, 1981). Goldstein (1979) recommends that the researcher try to anticipate future data needs and try to be redundant in the early stages of the research project. Others suggest collecting data in a way that would allow them to be used in different ways and from different theoretical perspectives, as well as keeping them in their most basic form (i.e., raw data rather than composites or summary measures) to allow for other uses (Bergman & Magnusson, 1990).

With respect to personnel issues, it is suggested that the researcher train staff well in advance of the designated time for collecting data (Jordan, 1994). Jordan (1994) further advises against blind testing because of the need to build rapport and a relationship over time, although this view is contrary to that most commonly held by experimental researchers and will likely depend on the nature of the study being conducted. Study personnel who are not blind might introduce expectancy bias into the data, which can affect things like how a construct is operationalized and how raters operate (Bergman & Magnusson, 1990). To prevent staff turnover, it has been suggested that researchers can maintain investment in a longitudinal study by publishing as much as possible (Mednick, Griffith, & Mednick, 1981), although it has been pointed out that the publication of earlier waves of research could potentially impact subsequent behaviors (van der Kamp & Bijleveld, 1998).

In planning a longitudinal study the researcher would be wise to review Rogosa's (1995) "myths" about longitudinal research, particularly in regards to determining the number of times of measurement. One such myth is that two times of measurement constitutes a longitudinal study. Although one can plot the amount of change observed between two points in time, one cannot determine the shape of the growth curve from only two data points. Moreover, if the change function is not a straight-line function then time of measurement can be quite influential. Rogosa (1995) recommends the use of multiple measurement points and growth curve data for the best statistical analysis and examination of individual growth trajectories over time.

Planning is also essential with respect to time of measurement. Goldstein (1979) states that it is almost inevitable that there will be some variation around the targeted sampling age, and this should not be problematic if it is small and random, but if the variation in time of measurement is large it could pose a problem. For example, the findings of a study could be impacted by time of measurement effects if developmental change on some attribute is rapid, if there is skew in the sampling, or if there is some relationship between the time of measurement and the average value of the measurement (e.g., seasonal variations in phenomena such as physical growth). Goldstein recommends sampling throughout the year to avoid this.

An Example of a Simple Longitudinal Study: The Dunedin Study

Silva and colleagues (Silva, 1996; Silva & McCann, 1996) conducted a noteworthy longitudinal study of over 1,000 infants born at one hospital in Dunedin, New Zealand. Infants born between April 1, 1972 and March 31, 1973 were enrolled and assessed at birth, 3, 5, 7, 9, 11, 13, 15, 18, and 21 years of age. The objectives of the study were to examine the health, development, and wellbeing of the participants at each age. This study led to 555 publications by April 1995 (Silva & McCann, 1996). What makes the Dunedin study particularly remarkable is the very low attrition rate: 97 percent of participants were followed at ages 18 and 21. Over the course of the study the participation rate dropped as low as 82 percent at age 13, but the researchers were able to implement aggressive retention measures such as flying participants who had moved away back to New Zealand and using interviewers in other locations such as Australia. Attrition analyses were conducted and revealed no significant differences between dropouts and those who remained in the study between ages 3 and 11 except on socioeconomic status and single motherhood.

Silva (1996) writes that the costs of the study were minimal in the beginning because the services of many volunteers were used, but over the years the costs increased. These were covered in part by government agency funding or grants, but in addition academics and professionals as part of their jobs did work. He notes that costs increased at the age 9 testing period because testing sessions increased from one half-day to a full day. The costs for the age 21 testing sessions were considerable because of the expenses involved in flying participants back to New Zealand and paying incentives. If the entire study had been funded as a yearly contract, it is estimated that it would have cost $1 million a year, which really is not such an unreasonable amount of money considering the wealth of health and developmental data that has been generated by this study.

Thus the simple longitudinal research design can be a powerful method for gaining knowledge of developmental processes, but it is fraught with methodological and practical problems that make it challenging to complete. Clearly one does not embark upon such a study without careful planning and a great deal of patience since the rate of progress will be inevitably slow, especially in the early years of the study. During the planning phase, it is essential that the researcher spend considerable time developing theories and conceptualizing relationships between variables, especially in light of the fact that many developmental researchers have tended to regard time and cohort as confounds while seeking only pure age effects. According to Schaie (1984) this is a "static and ahistorical" (p. 2) approach. A review of the variables laid out by the general developmental model – age, cohort, and time of measurement – and the meaning of each variable within the developmental function might be helpful to the researcher.

Special Considerations Regarding Age, Cohort, and Time of Measurement

Age as a variable

Age is commonly used as an independent variable in developmental research and as "the central marker of development in biological and psychological research on developmental phenomena" (Bergman & Magnusson, 1990, p. 26). Hertzog (1996) warns, however, that chronological age only imperfectly maps onto maturational, psychological, and social aging processes. It is not necessarily a correlate of time of onset or duration of a particular behavior, and is not the same thing as biological age (Bergman & Magnusson, 1990; Schaie, 1988). Indeed, development needs to be understood as a function of both chronological and biological age, as well as birth cohort and time of measurement. Age thus should not be conceptualized as a causal variable, but rather as a proxy variable for a host of co-occurring, co-varying processes and events that can be more meaningfully used to account for age-related change. Variables such as biological maturation, years of schooling, and specific experiences are some prime examples (Miller, 1998).

It is clear, therefore, that age is not a singular variable but is multiply defined by a number of variables. Wohlwill (1970a) makes this case strongly in stating that age is:

> at best a shorthand for the set of variables acting over time, most typically identified with experiential events or conditions, which are in a direct functional relationship with observed developmental changes in behavior; at worst it is merely a cloak for our ignorance in this regard (p. 30).

Cohort as a variable

Similarly, Baltes, Cornelius, and Nesselroade (1979) state that developmental researchers should not automatically begin a study by assuming that age is the most important explanatory variable for a phenomenon under investigation. This is true especially for samples of adolescents and adults, when cohort effects could potentially be much more explanatory than age. However, for the most part, the role of cohort effects in developmental processes has been presented descriptively rather than empirically (Baltes, Cornelius, & Nesselroade, 1979).

Baltes, Cornelius, and Nesselroade (1979) also note that the researcher's approach to dealing with the cohort variable depends on the nature of the research question. It could be seen as error, which is most probably true of child researchers who examine basic processes such as learning and cognition. It also could be seen as a dimension of external validity/generalization. Only a sequential strategy will separate out cohort-specific or between-cohort differences from "true" ontogenetic development (see Schaie & Caskie, this volume). The cohort variable also could be treated as a theoretical and process variable, but this is difficult to do because it requires explication of the "form and nature of cohort change that is judged to be developmental, the need for such concepts as stages or transitions in representing cohort change, and the types of explanatory mechanisms involved in producing cohort change" (Baltes, Cornelius, & Nesselroade, 1979, p. 80).

Kosloski (1986) makes a persuasive argument that defining a birth cohort simply by a shared discrete period of time does not automatically mean that members of that cohort shared experiences, so that the "cohort effect" might be meaningless in some instances. At other times, however, it makes sense to expect that historical events will impact individuals of varying ages differentially. Defining cohort by birth says little about what specific events "define" that cohort, and is likely to be of little real use (Kosloski, 1986).

Cohort can be related theoretically to development through the use of a model explicated by Baltes and Nesselroade (1979). They assert that there are three influences on behavioral development: normative age-graded, normative history-graded, and nonnormative. Normative age-graded influences are those most highly correlated with chronological age, and include processes such as biological maturation and socialization processes that are widely experienced across time and cohorts. Normative history-graded influences are those biological and social processes that are more culturally based and that are presumed to affect most members of a cohort, such as entering school. Nonnormative influences are those biological and social processes that do not impact most members of a cohort, such as illness, disability, divorce, and unemployment. These all operate simultaneously over time, which leads to between-cohort differences in developmental change as well as within-cohort differences (Baltes, Cornelius, & Nesselroade, 1979). Schaie (1986) adapts this framework when he proposes a method of composing cohorts that result in cohorts free from chronological age, thus allowing the researcher to more fully examine the parameters of development as set forth by the general developmental model without the constraint of calendar time.

Nesselroade and Baltes (1974) conducted a study that demonstrated a cohort effect. Longitudinal sequences of cohorts born in 1954, 1955, 1956, and 1957 and tested every year from 1970 to 1972 were used. Over 1,800 subjects were drawn from public schools in West Virginia and given personality and ability tests. Data analyses showed significant main effects of time of measurement on 7 of 10 personality variables and significant main effects of cohort on 2 of 10 personality variables. The main effects of cohort on the personality variables Independence and Achievement can be seen in Figure 1.1. The 14-year-olds tested in 1972 scored much higher in Independence than 14-year-olds tested in 1970 or 1971, while the 14-year-olds tested in 1970 scored higher in Achievement than 14-year-olds tested in 1971 and 1972. The researchers interpreted these findings as suggesting that the social-cultural context is more influential than maturation in adolescent personality development. This study is also important in that it demonstrated retest effects for the mental abilities testing, and attrition influenced the findings, such that those who remained in the study performed better than those who dropped out.

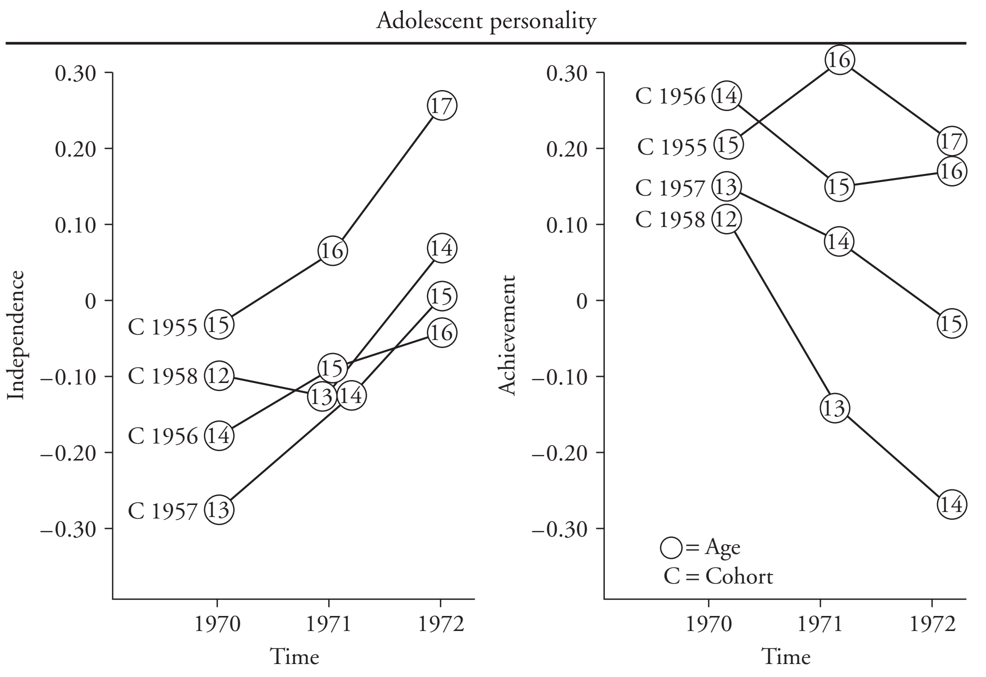

Figure 1.1 Differences in adolescent personality development as a function of the cohort effect. (From Baltes, Cornelius, & Nesselroade (1979). Copyright 1979 by Academic Press. Reprinted by permission.)

Time as a variable

The contribution of time to the developmental function has perhaps been the least understood, and, like the cohort variable, it has tended to be treated as a confound rather than an integral part of development (Schaie, 1984, 1986). Simply acknowledging the influence of time, which can be defined as a marker for historical events (Kosloski, 1986), does not give it explanatory power; it begs the question of what is the underlying psychological process (Caspi & Bem, 1990). Historical time is an essential parameter of developmental research, according to Schaie (1986), because it provides an important context for development.

Schaie (1986) suggests redefining time of measurement in terms of the impact of events on life-span development, which would separate the variable from calendar time. When attempting to determine what influences and processes might be important in terms of historical time, Schaie (1984) suggests looking for "societal changes in technology, customs, and cultural stereotypes that might constrain behavior" (p. 8). The researcher needs to make some conceptualization of historical causation as being distal or proximal as this affects the spacing of observations in longitudinal research (Baltes & Nesselroade, 1979). Finally, it is probably most important for the researcher studying adult/life-span development to hypothesize about the contribution of historical time to the developmental process since it is most likely influential in adulthood (Schaie, 1984, 1986). Indeed, the adult development researcher studying individual differences is really studying cohort and period effects according to Schaie (1986).

Conclusions

Donaldson and Horn (1992, p. 213) noted that "age, cohort, and time constitute a muddle. They are redundant quantities that cannot be independently varied to produce unique contributions to a dependent variable." Psychologists, they argue, tend to ignore cohort and time variables because they are the domains of other disciplines and assert that psychologists alone will not be able to construct a general model of development which takes into account the effects of age, cohort, and period. This is a strong call for interdisciplinary scholarship in developmental research. Indeed, the complexities seen in separating out the influences of age, cohort, and time of measurement on human development mandate the creation of complex models of development which will require the expertise of many disciplines (Baltes, Cornelius, & Nesselroade, 1979; Donaldson & Horn, 1992). At the very least, one undertaking a cross-sectional or simple longitudinal study should recognize that the best solution to the problem of separating age, period, and cohort effects is to measure directly those things that the variables index (Kosloski, 1986).

As indicated above, simple cross-sectional and longitudinal designs have been criticized severely over the years. At the same time, we would argue that, despite the flaws inherent in simple developmental designs, they remain high on the list of design choices when one is interested either in testing or generating hypotheses about developmental phenomena. Indeed, there is no better design than the longitudinal design for identifying age-related developmental change, and competent use of this and the cross-sectional design requires that one understand, in a proactive way, the issues and pitfalls associated with each, and plan data collection strategies and choose dependent variables so as to minimize concerns about threats to validity. We endorse Miller's (1998) point that nearly everything we know about developmental processes is the result of cross-sectional or longitudinal designs. Despite their limitations, we expect these methods will continue to be used by students of human development for many years to come.

References

Achenbach, T. M. (1978). *Research in developmental psychology: Concepts, strategies, and methods*. New York: Free Press.

Baltes, P. B. (1968). Longitudinal and cross-sectional sequences in the study of age and generational effects. *Human Development*, *11*, 145–71.

Baltes, P. B. & Nesselroade, J. R. (1979). History and rationale of longitudinal research. In J. R. Nesselroade & P. B. Baltes (eds.), *Longitudinal research in the study of behavior and development* (pp. 1–39). New York: Academic Press.

Baltes, P. B., Cornelius, S. W., & Nesselroade, J. R. (1979). Cohort effects in developmental psychology. In J. R. Nesselroade & P. B. Baltes (eds.), *Longitudinal research in the study of behavior and development* (pp. 61–87). New York: Academic Press.

Baltes, P. B., Reese, H. W., & Nesselroade, J. R. (1977). *Life-span developmental psychology: Introduction to research methods*. Monterey, CA: Brooks/Cole.

Bergman, L. R. & Magnusson, D. (1990). General issues about data quality in longitudinal research. In D. Magnusson & L. R. Bergman (eds.), *Data quality in longitudinal research* (pp. 1–27). New York: Cambridge University Press.

Bergman, L. R., Eklund, G., & Magnusson, D. (1991). Studying individual development: Problems and methods. In D. Magnusson, L. R. Bergman, G. Rudinger, & B. Torestad (eds.), *Problems and methods in longitudinal research: Stability and change* (pp. 1–31). New York: Cambridge University Press.

Buss, A. R. (1979). Toward a unified framework for psychometric concepts in the multivariate developmental situation: Intraindividual change and inter- and intraindividual differences. In J. R. Nesselroade & P. B. Baltes (eds.), *Longitudinal research in the study of behavior and development* (pp. 41–59). New York: Academic Press.

Campbell, R. T. (1988). Integrating conceptualization, design, and analysis in panel studies of the life course. In K. W. Schaie, R. T. Campbell, W. Meredith, & S. C. Rawlings (eds.), *Methodological issues in aging research* (pp. 43–69). New York: Springer-Verlag.

Caspi, A. & Bem, D. J. (1990). Personality continuity and change across the life course. In L. Pervin (ed.), *Handbook of personality theory and research* (pp. 549–75). New York: Guilford.

Cook, T. D. & Campbell, D. T. (1979). *Quasi-experimental design and analysis issues for field settings*. Chicago, IL: Rand McNally.

Donaldson, G. & Horn, J. L. (1992). Age, cohort, and time developmental muddles: Easy in practice, hard in theory. *Experimental Aging Research*, *18*, 213–22.

Farrington, D. P. (1991). Longitudinal research strategies: Advantages, problems, and prospects. *Journal of the American Academy of Child and Adolescent Psychiatry*, *30*, 369–74.

Flavell, J. H., Beach, D. R., & Chinsky, J. M. (1966). Spontaneous verbal rehearsal in a memory task as a function of age. *Child Development*, *37*, 283–99.

Flick, S. N. (1988). Managing attrition in clinical research. *Clinical Psychology Review*, *8*, 499–515.

Friedman, S. L., Haywood, H. C., & Livesy, K. (1994). From the past to the future of developmental follow-up research. In S. L. Friedman & H. C. Haywood (eds.), *Developmental follow-up: Concepts, domains, and methods* (pp. 3–26). San Diego, CA: Academic Press.

Goldstein, H. (1979). *The design and analysis of longitudinal studies: Their role in the measurement of change*. New York: Academic Press.

Gopnik, A. & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development*, *59*, 26–37.

Hartmann, D. P. & George, T. P. (1999). Design, measurement, and analysis in developmental research. In M. H. Bornstein & M. E. Lamb (eds.), *Developmental psychology: An advanced textbook* (4th edn., pp. 125–95). Mahwah, NJ: Lawrence Erlbaum Associates.

Hertzog, C. (1996). Research design in studies of aging and cognition. In J. E. Birren & K. W. Schaie (eds.), *Handbook of the psychology of aging* (4th edn., pp. 24–37). San Diego, CA: Academic Press.

Jordan, T. E. (1994). The arrow of time: Longitudinal study and its applications. *Genetic, Social, and General Psychology Monographs*, *120*, 469–531.

Kosloski, K. (1986). Isolating age, period, and cohort effects in developmental research: A critical review. *Research on Aging*, *8(4)*, 460–79.

Kraemer, H. C. (1994). Special methodological problems of childhood developmental follow-up studies: Focus on planning. In S. L. Friedman & H. C. Haywood (eds.), *Developmental follow-up: Concepts, domains, and methods* (pp. 259–76). San Diego, CA: Academic Press.

Kraemer, H. C., Yesavage, J. A., Taylor, J. L., & Kupfer, D. (2000). How can we learn about developmental processes from cross-sectional studies, or can we? *American Journal of Psychiatry*, *157*, 163–71.

Mayer, K. U. & Huinink, J. (1990). Age, period, and cohort in the study of the life course: A comparison of classical A-P-C analysis with event history analysis or Farewell to Lexis? In D. Magnusson & L. R. Bergman (eds.), *Data quality in longitudinal research* (pp. 211–32). New York: Cambridge University Press.

McCall, R. B. (1977). Challenges to a science of developmental psychology. *Child Development*, *48*, 333–44.

Mednick, B. R., Mednick, S. A., & Griffith, J. J. (1981). Some recommendations for the design and conduct of longitudinal investigations. In F. Schulsinger, S. A. Mednick, & J. Knop (eds.), *Longitudinal research: Methods and uses in behavioral science* (pp. 285–95). Boston, MA: Martinus Nijhoff.

Mednick, S. A., Griffith, J. J., & Mednick, B. R. (1981). Problems with traditional strategies in mental health research. In F. Schulsinger, S. A. Mednick, & J. Knop (eds.), *Longitudinal research: Methods and uses in behavioral science* (pp. 3–15). Boston, MA: Martinus Nijhoff.

Miller, S. A. (1998). *Developmental research methods* (2nd edn.). Upper Saddle River, NJ: Prentice-Hall.

Nesselroade, J. R. & Baltes, P. B. (1974). Adolescent personality development and historical change: 1970–1972. *Monographs of the Society for Research in Child Development*, *39* (1, serial no. 154).

Pellegrini, A. D. (1996). *Observing children in their natural worlds: A methodological primer*. Mahwah, NJ: Lawrence Erlbaum Associates.

Rogosa, D. (1995). Myths and methods: "Myths about longitudinal research" plus supplemental questions. In J. M. Gottman (ed.), *The analysis of change* (pp. 3–66). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Schaie, K. W. (1977). Quasi-experimental designs in the psychology of aging. In J. E. Birren & K. W. Schaie (eds.), *Handbook of the psychology of aging*. New York: Van Nostrand.

Schaie, K. W. (1983). 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.

Schaie, K. W. (1984). Historical time and cohort effects. In K. A. McCluskey & H. W. Reese (eds.), *Life-span developmental psychology: Historical and generational effects* (pp. 107–21). New York: Academic Press.

Schaie, K. W. (1986). Beyond calendar definitions of age, time, and cohort: The general developmental model revisited. *Developmental Review*, *6*, 252–77.

Schaie, K. W. (1988). Methodological issues in aging research: An introduction. In K. W. Schaie, R. T. Campbell, W. Meredith, & S. C. Rawlings (eds.), *Methodological issues in aging research* (pp. 1–11). New York: Springer-Verlag.

Schaie, K. W. (1996). *Intellectual development in adulthood: The Seattle Longitudinal Study*. New York: Cambridge University Press.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Siegler, I. C. & Botwinick, J. (1979). A long-term longitudinal study of intellectual ability of older adults: The matter of selective subject attrition. *Journal of Gerontology*, *34*, 242–45.

Silva, P. A. (1996). The future of the Dunedin Study. In P. A. Silva & W. R. Stanton (eds.), *From child to adult: The Dunedin Multidisciplinary Health and Development Study* (pp. 259–66). Auckland, New Zealand: Oxford University Press.

Silva, P. A. & McCann, M. (1996). An introduction to the Dunedin Study. In P. A. Silva & W. R. Stanton (eds.), *From child to adult: The Dunedin Multidisciplinary Health and Development Study* (pp. 1–23). Auckland, New Zealand: Oxford University Press.

van der Kamp, L. J. T. & Bijleveld, C. C. J. H. (1998). Methodological issues in longitudinal research. In C. C. J. H. Bijleveld & L. J. T. van der Kamp (eds.), *Longitudinal data analysis: Designs, models, and methods* (pp. 1–45). London: Sage.

Wohlwill, J. F. (1970a). The age variable in psychological research. *Psychological Review*, *77*, 49–64.

Wohlwill, J. F. (1970b). Methodology and research strategy in the study of developmental change. In L. R. Goulet & P. B. Baltes (eds.), *Life-span developmental psychology: Research and theory* (pp. 149–91). New York: Academic Press.

Wohlwill, J. F. (1973). *The study of behavioral development*. New York: Academic Press.