Part 8

Extending Design Research Methodologically
22 Design Research and the Study of Change
Conceptualizing Individual Growth in Designed Settings

Finbarr C. Sloane
Arizona State University

Anthony E. Kelly
George Mason University

Introduction
All educational research involves design choices. For design researchers, the nature of the choice is most explicitly about the character of the designed artifact (e.g., software or learning environment) or about the students’ navigation of some content terrain (Cobb et al., 2003). Less obvious, and perhaps unconscious, are the education researcher’s beliefs about the nature of change. These beliefs affect the researcher’s choice of theoretical frame, choice of measure, and choice of analytic tool. These choices interact one with another and critically affect the way inferences are drawn (either qualitatively or quantitatively). The goal of this chapter is to highlight some of the defining features of change implicit in education research to guide design researchers as they move to quantify student growth over time. We note that the challenges posed to design researchers in modeling change afflict, equally, the modeling of change even by those with mastery of current statistical modeling formalisms. Models of change over time intersect with and are grounded in larger construct validity issues facing all education researchers that are not resolved (but at least made more explicit) by quantitative techniques.

Assessing Change in Designed Settings
In some areas of social science research such as those in developmental psychology (e.g., Slater & Bremner, 2003), the concept and assessment of change is explicitly the focus of study. In much design research, the concept of change over time is less explicit but no less fundamental. For example, the Design-Based Research Collective (2003) focused on the design of technological artifacts that support and change student learning, while Cobb et al. (2003) focused explicitly on changes of student learning of content with less emphasis placed on the “designed intervention.” In both cases, change over time is fundamental. Accordingly, it is important that design researchers possess a good grasp of the basic but often ignored issues relating to its conceptualization and measurement.

Standard Pre-Post Testing Models
The standard model for measuring change in educational research is the pre-post testing model. Generally, pre-post testing models are two time-point models and are found
in general studies of learning (Shadish et al., 2004; Singer & Willett, 2003; Willett, 1989), and in design research studies (e.g., Fishman et al., 2004). Data from such a design are analyzed by some two-wave analytic technique such as computing a difference (or change) score, a residual change score, or a regression estimate of true change.

Technical Concerns

The ubiquity of two-wave analytic techniques in applied educational research would suggest that these techniques are uncontroversial. Actually, the difference score has been criticized on technical and substantive grounds. The difference score frequently has negative correlation with initial status (i.e., the change score is often negatively correlated with the pretest score). Its relatively low reliability (Bereiter, 1963; Cronbach & Furby, 1970; Linn & Slinde, 1977) is related to a number of factors including measurement error. For a more complete treatment, including cases in which difference scores can be viewed as unbiased estimates of change over time, see Rogosa and Willett (1985). The lesson for design researchers is that unreliability of measurement (and consequently its validity) poses problems even for quantitative researchers with sophisticated instruments. The measures used by design researchers (sometimes using a very small number of items or subjective judgments based on observations of complex classroom processes) cannot be assumed to be either reliable or valid. This clouds the quality of the claims that emerge at each time point, and the change (or learning) that occurs over time.

Substantive Issues

Two time-point designs generate problems because these data provide no precise information on intra-individual change over time. As Willett (1989: 347) noted, it is a:

... conceptualization that views individual learning, not as a process of continuous development over time, but as a quantized acquisition of skills, attitudes, and beliefs. It is as though the individual is delivered a quantum of learning in the time period that intervenes between the pretest measure and the posttest measure, and that our only concern should be with the size of the acquired chunk.

Because many design researchers use two time-point studies few non-content oriented design researchers detail their work in this manner.

Framed analytically, with only two snapshots, individual growth curves cannot be characterized with certainty (Bryk & Weisberg, 1977). The simplest approach is to assume that any growth from time 1 to time 2 is linear (thus allowing the difference score calculation). Mathematically, however, an infinite number of curves could pass through two points. Yet, some design researchers may find reasons to challenge the simple linear model by demonstrating changes over time in terms of apparent mastery by students of increasingly difficult content (e.g., hypothetical learning trajectories [Cobb et al., 2003]).

Considering More Than Two Time-Points

Multiwave (three or more time-points) repeated measurement within a longitudinal design allows for the possibility of better mapping to intra-individual change as it unfolds (Raudenbush & Bryk, 2002; Singer & Willett, 2003). However, the majority of
published methodological work in this area is highly technical, requiring quantitative methodological expertise and valid reliable quantitative measures of learning, which typically are not available given the prospective and the iterative character of design research. To bridge the gap between these statistical advances and design research in substantive areas we present a non-technical presentation of several fundamental questions concerning the conceptualization and analysis of change over time.

Fundamental Questions Regarding the Understanding of Change

An adequate methodology for building and assessing a theory of change can be evaluated in terms of the extent to which the theory and commensurate methodology can address the following problems, which are derived from discussions in Collins (1991), Golembiewski et al. (1976), Nesselroade (1991), and Pellegrino et al. (2001).

Problem 1

Do researchers assume the measure of change to be systematic or random? Change can refer to interpretable systematic differences or random fluctuation. As educational and design researchers, we are almost always substantively interested in interpretable systematic change (as opposed to making sense of random fluctuation). However, most, if not all, measuring instruments used in educational research have measurement error. That is, they are not perfectly reliable or valid and as such there are many discrepancies between the constructs and the measures used to operationalize them. As we noted, the difficulty of establishing high reliability in measures can create difficulties for the field in conceptualizing an adequate model of change. An adequate change assessment methodology should account for measurement error and allow observed variance to be partitioned into true construct variance, nonrandom (systematic) error variance, and random error variance.

When the same student or set of students is measured, repeatedly, then time-specific and time-related errors have to be considered carefully. When consecutive measurements, or measurement occasions, are closely spaced, as in teaching experiments, measurement errors will also be correlated. This is especially true when the same measurement tool, or set of items, is used by the researcher. We need to be cognizant of these possible dilemmas and ways to deal with them in our analyses. Where sound quantitative measures are available, autocorrelated error regression models can be used to address this problem (Hedeker & Gibbons, 1997). The solution for mixed method or design research studies is not yet clear. Researchers in each methodological tradition will likely find unique, but somewhat different, answers.

Problem 2

Do researchers consider the change to be reversible? This assumption or realization has important implications for the functional form of the growth curves to be modeled. While it is simple to assume that the growth trajectory may be monotonically increasing or decreasing (e.g., linear), perhaps most psychological interest resides in U-shaped or inverted U-shaped learning curves. These curves may show growth–decline–growth, or decline–growth–decline, respectively. Learning scientists are also interested in curves with plateaus (e.g., stages in growth), or curves that display cubic growth (growth–plateau–further growth). Additionally, an adequate change assessment methodology should allow two things. First, the researcher should be able to specify a priori one or
more functional forms. Second, the researcher should be able to assess the goodness-of-fit of each form and the incremental fit of one form over another. The learning theory helps the researcher specify the anticipated functional form of the learning to be measured and modeled. Moreover, such a theory also helps the researcher specify an adequate number of measurement occasions, and the spacing of such measurements.

Problem 3

Change may be assumed to be proceeding in a fixed pathway between sampled time-points. Multiple paths may also occur as students proceed from one time-point to another. For example, some students may follow a linear trajectory, while others followed a quadratic trajectory. An adequate theory of learning should be able to identify and describe why certain subgroups of students follow different paths. In parallel, an adequate assessment methodology should also be able to identify these subgroups, and when data should be optimally collected for each group. Qualitative studies bring these concerns to our attention. Design research studies that purposefully (re)intervene to effect changes in learning can be assumed to promote multi-pathway growth, adding complexity to the statistical and theoretical modeling problems.

Problem 4

Do researchers view change: (a) to be continuous and gradual, (b) to have large magnitude shifts on quantitative variable(s), or (c) to progress through a series of qualitatively distinct stages? Gradual or large shifts can be captured by multi-point analyses. Where there are sharp qualitative shifts in the conceptualization of the phenomenon being measured between time-points (as may be expected with researcher-induced interventions), continuous metric models may be inadequate. One candidate approach to handling such eventualities is Wilson’s *Saltus* model (for Piagetian and other stage dependent developmental theory). This model will allow for the detection of stages when these stages have been built into the careful writing and testing of the item pool (Wilson, 1989). A theory for such discrete changes would need to be developed (see Case, 1985) and articulated before the measurement items can be developed, piloted, and tested. But even here the discrete shifts are embedded as an extra parameter in the Rasch model: a psychometric model that forces continuity for fit to the psychometric assumptions of the model to occur (Rasch, 1980).

Sharp, qualitative, and substantive changes in the assumed measurement construct pose significant problems for quantitative modeling. Under certain conditions, not discussed here, growth models can be used to analyze change in non-continuous outcomes such as counts, dichotomies (whether a student persists in a content area or not), and ordinal outcomes. These models are decidedly more sophisticated and draw on the statistical theory of generalized linear models (McCullagh & Nelder, 1989). Only recently, have these statistical models been extended to allow for the analysis of nested data structures (Raudenbush & Bryk, 2002; Singer & Willett, 2003).

Problem 5

Is student growth occurring in what Golembiewski et al. (1976) would consider as *alpha*, *beta*, and *gamma* change? Alpha change is assumed to be measured against a reasonably constant knowledge base—one that is reliably and validly measured. Measurement
invariance across time exists when the numerical values across time are on the same measurement scale (Drasgow, 1984, 1987). Alpha change refers to changes in absolute differences given a constant conceptual domain and a constant measuring instrument (much like data gathered in the Longitudinal Study of American Youth [Miller et al., 2000]).

Beta change refers to changes in the measuring instruments given a constant conceptual domain. Beta change occurs when there is recalibration of the measurement scale. That is, in beta change the observed change results from an alteration in the respondent’s subjective metric rather than from actual change in the construct of interest. When beta change occurs there is a stretching or shrinking of the measurement scale, making direct pretest/post-test comparisons problematic.

Gamma change refers to changes in the conceptual domain, e.g., those involving dramatic qualitative shifts in understanding by the students. Lord (1963) highlights this problem in the context of multiplication: “he argued that a multiplication test may be a valid measure of a mathematical skill for young children, but it becomes a measure of memory among teenagers” (Singer & Willett, 2003: 14). Gamma change can take a variety of forms. For example, in the context of factor analysis, the number of factors assessed by a given set of measures may change from one point in time to another. Alternatively, the number of factors may remain constant across time, but a differentiation process may occur so that the factor inter-correlations vary over time. When there is gamma or beta change over time, it is unlikely that a simple growth model will provide usable insight.

Problem 6
Do researchers consider the change to be a shared characteristic of a group of individuals over time, what occurs within individuals over time, or both (Lave & Wenger, 1990)? This question originally derives from Allport’s (1937) distinction between the nomothetic research orientation, which focuses on lawful relations that apply across individuals, and the idiographic research orientation, which focuses on uniqueness of individuals. In design research, this issue of coordinating analyses at the individual and group levels poses significant methodological challenges, particularly when it is difficult to characterize contingent versus the necessary processes affecting learning (Kelly, 2004).

Problem 7
Do researchers assume that there are systematic inter-individual differences in the values of the individual growth parameters (e.g., initial status, and rate of change) that define the individual trajectory (Huttenlocher, et al., 1991), assuming that all individuals have trajectories of the same functional form (e.g., linear, quadratic, etc.)? If so, how can we predict and increase our understanding of these inter-individual differences? The rate of change is a critical individual growth parameter that has been, until recently, neglected in the conceptualization and measurement of inter-individual differences. In many content domains of design research (e.g., learning in mathematics and science), the rate of change is of theoretical and practical importance, and is in need of further study. Moreover, as was demonstrated by Sloane, Helding, and Kelly (this volume) and Sloane (this volume), these rates can now be measured, carefully planned for, and analyzed.
Problem 8
Do researchers assume that there are cross-domain relationships (e.g., issues between content, content knowledge for teaching, and teacher instructional performance) in change over time? Is the relationship between inter-individual differences in intra-individual change over time and the predictors of those differences invariant across domains? These questions can only be addressed statistically with a multivariate, multi-level type analytic tool (see Thum, 1997).

Problem 9
Finally, in order to draw differential claims, the models of change assumed in a group of students under scrutiny may need to be compared to the behavior and learning of students in some other group (e.g., matched comparison students or students in a randomized cohort). The question of interest is whether a specific change pattern found in one group is equal to or differs from (in either magnitude or form) the change pattern found in the comparison group. The drawbacks of non-randomized controls or even matched groups have been documented elsewhere in advocacy pieces for randomized clinical trials in education, but the focus in design research represents a new challenge, methodologically, given the desire to recursively change instruction or artifact design in response to student changes.

Conclusions
If the parameters of growth models are to be valuable for researchers (e.g., educational and design researchers), the actual trajectories being modeled must make conceptual and theoretical sense from the perspective of the researcher and be consistent with a particular learning theory (e.g., the construction of Vergnaud’s [1988] theory of additive and multiplicative changes in students’ understanding of mathematical constructs). Moreover, the measures must map to the goals of the study. Longitudinal analysis imposes three additional conditions because the metric, validity and precision of the measured outcome must be preserved across the time period under study. The conditions include: (a) the outcome scores must be equitable across time, (b) the outcomes must be equally valid across occasions, and (c) the precision of the measure should also be preserved across time. Meeting these working conditions in educational research, however, is non-trivial, and generates a major challenge for future research.

Given their propensity to actively adapt instructional interventions, to change artifact design repeatedly in response to student learning, to involve subjects (both teacher and students) in iterative design cycles, and to adapt measures and concepts over time, design researchers pose significant challenges to methodologists and psychometricians. In keeping with the creative commissive space of design research (Kelly, 2006), we suggest cooperation across the fields of applied research and methodology in order to provide design researchers with the tools they need, rather than simply fault them for not comporting with models of change in education that are often simplistic or chosen primarily for their mathematical tractability.

Acknowledgments
The first author would like to thank Dan Battey (Mary Lou Fulton College of Education, Arizona State University) for his comments on an earlier draft of this chapter.
Additionally, the first author acknowledges funding from the National Science Foundation, NSF 0634103 (Moving from Arithmetic to Algebra) and NSF 0412537 (i.e., MSP: Project Pathways). Opinions expressed here are those of the author and not necessarily those of the Foundation.

Note
1 Wilson (1989) designed the Saltus model as a developmental extension of the Rasch (1980) model to measure discontinuous stage changes in persons. Saltus is the Latin word for leap. The Saltus model measures state changes using multiple tasks at each of the various developmental levels. Wilson (1989) described the Saltus model in terms of distinctions between first-order and second-order discontinuities. First-order discontinuities are sudden or abrupt changes in a single ability, whereas second-order discontinuities, or shifts, are sudden or abrupt changes in at least two dimensions. The Rasch model is used for the first-order discontinuities, and the Saltus model, which estimates parameters for persons, dimensions, and levels, is used for the second-order discontinuities.

References


