Cognitive Developmental Change

Theories, Models and Measurement

Edited by Andreas Demetriou, Athanassios Raftopoulos

Book DOI: http://dx.doi.org/10.1017/CBO9780511489938

Online ISBN: 9780511489938

Hardback ISBN: 9780521825795

Paperback ISBN: 9780521184007

Chapter

10 - Dealing with change: manifestations, measurements and methods pp. 318-353

Chapter DOI: http://dx.doi.org/10.1017/CBO9780511489938.012

Cambridge University Press

10 Dealing with change: manifestations, measurements and methods

Elena L. Grigorenko and Paul A. O'Keefe

The main purpose of this chapter is to discuss briefly major methodologies used for the analysis of change. Throughout the development and maturation of approaches to measuring change, different types of change have been recognized and different methods of quantifying change have been developed (e.g., Collins and Horn 1991; Harris 1963). There is a colloquial reference to so-called old and new approaches to measuring change. 'Old' approaches refer to such conventional indicators of change as the differences between measures in a given time point and a subsequent time point. 'New' approaches refer to ever more complex methodologies for describing and quantifying development, whether spontaneous or occurring in response to intervention.

Clearly, the amount of information on a subject studied by many outstanding scientists that can be introduced in a single chapter is limited. Moreover, a single chapter cannot compete with the comprehensive volumes that have recently been written on the same topic (e.g. Collins and Sayer 2001; Gottman 1995; Moskowitz and Hershberger 2002; von Eye and Niedermeier 1999). Therefore, the strategy selected in this chapter for material presentation is to introduce briefly selected methodological approaches, illustrate them with specific examples, and provide the reader with a wealth of relevant references.

We decided to illustrate various change-related methodologies by reviewing a limited content area – the field of studies related to acquisition, both natural and in response to targeted intervention, of oral and written language. Such a decision is not random, of course. The selection was driven by two considerations. First, the processes of oral and written language acquisition are fundamentally developmental because they (a) mark the emergence of something new (skill, function, or level); (b) assume both procedural continuity and

This research was supported by Grant R206R50001 from the Institute of Educational Sciences (formerly the Office of Educational Research and Improvement), US Department of Education. Preparation of this report was supported by Grant R206R00001 from the same organization. Grantees undertaking such projects are encouraged to express freely their professional judgment. This article, therefore, does not necessarily represent the position or policies of the US Department of Education, and no official endorsement should be inferred. Correspondence should be sent to Elena L. Grigorenko, PACE Center, Yale University, Box 208358, New Haven, CT 06520-8358.

discontinuity; and (c) presume directionality (Pascual-Leone 1995; van Geert 1995; Vygotsky 1982). Second, these acquisitions embody change in its purest form. Both speaking and reading, from a developmental point of view, are characterized by (1) the existence of a zero point (i.e. there are developmental stages that do not include these processes – pre-linguistic and pre-reading stages), (2) unfolding developmental trajectories (i.e. both oral and written languages have to be mastered so that the child's performance meets certain criteria at different stages of development), and (3) their modifiable nature (i.e. speaking and reading interventions can improve the child's performance).

The following procedure was employed in the search for the original publications to be discussed in the present chapter. The article search was conducted using three different databases: PsycInfo, ERIC, and Medline. The terms *development*, *intervention*, *treatment* and *teaching* were identified as major methods of change. Each of these key words was independently cross-searched with the terms *language* and *reading*. Similarly, the four key words signifying change were cross-referenced with the terms *dyslexia*, *developmental dyslexia* and *specific language impairment*. The search was limited to articles that were written in English and published between 1993 and May of 2003. Any articles mentioning such conditions as mental retardation, autism, Turner's, Down's or any other developmental syndrome were omitted. The search resulted in 392 returns, of which sixty-five articles turned out to be review articles and, therefore, were not evaluated. In addition, from the remaining publications reporting empirical data, we excluded all articles that did not contain any reference to change occurring as an outcome of developmental processes or in response to an intervention. Specifically, 138 publications used cross-sectional rather than longitudinal analyses. Although cross-sectional methodologies are informative for understanding development, they are not, strictly speaking, designed to quantify change and, thus, are not in the focus of this chapter. Therefore, cross-sectional articles were excluded from the analysis. These elimination procedures limited the collection of articles for analyses to 189. Of these publications, thirty were deleted from our evaluation due to lack of clarity in describing the methodological procedures applied.

The purpose of this literature evaluation was two-fold. First, we wanted to survey the field in order to provide an adequate review of the methodologies of quantifying change. Second, we wanted to have at least an estimate – even if such an estimate might be biased, since we did not screen book chapters and dissertations – of the 'popularity', as defined by frequency of use, of different methodologies by simply establishing the percentage of usage of a given methodology. A review of methodologies used in the final set of 159 articles resulted in clustering the publications in five major groups of studies utilizing (1) difference scores; (2) techniques describing unsolicited change; (3) growth curves; (4) case analyses; and (5) dynamic systems methodologies. Figure 10.1

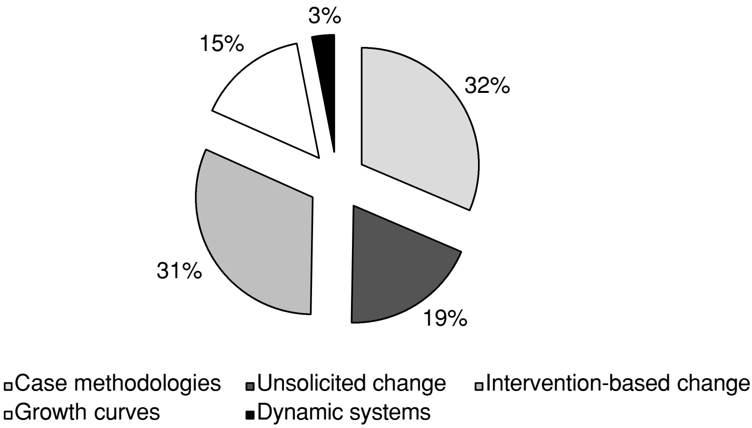

Figure 10.1. Proportional representation of dominant analysis types in the surveyed publications (1993–2003).

illustrates the proportions of these methodologies in the surveyed literature. Consequently, we structured our review around these methodologies. We start, however, with a brief account of the types of change identified in the literature.

Types of change

The concept of *change* is as old as or older than psychology itself. The first debates on the nature and importance of change for understanding humans and the world around them can be traced to the Ancient Greeks Heraklitos and Parmenides. Whereas Heraklitos viewed change as a ubiquitous phenomenon and held that nothing is ever the same (stable), Parmenides argued that stability is the foundation of the world and that change is only perceived, and therefore an illusion. Remarkably, although many theories of change and stability have been developed since, the bottom line of the argument is still not resolved, with the distinction between change and stability being one of the fundamental puzzles of human development (Oyama 2000).

Although no classification of change is universally accepted, many researchers have attempted to describe systematically different types of change. For example, in his work on linear syllogistic reasoning, van Geert (1995) has discussed three parameters characterizing the general growth model of a skill. Specifically, he talks about (a) a growth parameter (i.e. the changing process); (b) the current state of the changing process (i.e. the at-the-moment degree of

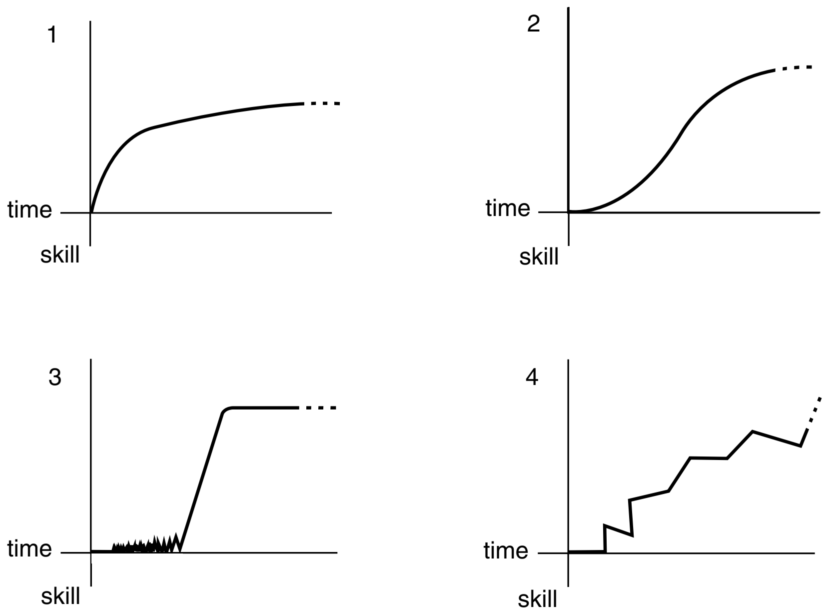

Figure 10.2. Types of change. Drawing by J. O'Keefe.

familiarity or mastery of the skill being acquired); and (c) a set of scaffolding factors immediately available to the child (i.e. motivational and instrumental factors the child has available to him or her in the process of the skill acquisition). Van Geert (1995) stated that various types of growth or change can be represented by combinations of these parameters. Here, following van Geert, we selectively present four types of growth that can be, at least in first approximation, captured by the parameters mentioned above – linear/nonlinear learning curves, S-shaped learning curves, saltatory growth, and stepwise growth (see Figure 10.2, 1–4).

The main assumption of the *linear/non-linear learning curve* is that of continuity of skill acquisition. The essence of this type of growth is that any future state of the skill is a function of the current state of the skill; thus, most skills improve with practice, along the dimension of acquiring competence in a domain (Figure 10.2, 1). Although this general assertion is preserved, what is really remarkable about this type of learning curve is that the initial pattern of rapid improvement is typically followed by lesser improvement with further practice. These curves are referred to as negatively accelerated learning curves; they are typically described by power functions. Based on the omnipresence of this type of learning in human development, 'the power law of practice' is said to be a ubiquitous characteristic of learning. The general learning curve is well described. The following parameters are referred to typically in specifications of learning curves: the range of learning (i.e. how far the initial performance is removed from the individual maximum), the trial number, and individual-specific learning rate parameters. It is fairly easy to see how these parameters map on the (a)–(c) parameters of van Geert. What is important for our discussion here is the statement that the change described by the linear/curvilinear learning curve is continuous. In this sense, it is not essential whether the change is linear (of which there are no examples in lifespan development) or non-linear; what is important is that the change is continuous and can be systematically described by a function depicting the rate of change depending on a given point on the change trajectory.

S-*shaped learning curves*, also referred to as logistic growth model curves, are, in a certain sense, derivatives of the learning curves described above and are also considered to be quite characteristic of a number of learning and developmental processes (Fischer and Rose 1994). The modification of the learning curve into the S-shaped curve results from the realization that the beginning stages of learning or trait acquisition might also occur at a rate different from the middle portion of the curve (see Figure 10.2, 2). Thus, S-shaped curves describe situations in which the learning in the middle of the process occurs at a much higher speed than at either the beginning or the end of the process.

Saltatory learning curves (see Figure 10.2, 3) depict situations when, after a prolonged beginning period of limited change, a rapid and substantial gain of skill occurs (e.g. van der Maas and Molenaar 1992). Saltatory curves are also characterized by a halt at some level of development, which is then specified as the maximum possible level of development for a given individual (or group). A well-studied example of saltatory growth is the mastery of the alphabet. Although there is a lot of intra-individual variation on how the alphabet is acquired by children, once it is mastered the maximum level of possible performance is reached.

Finally, *stepwise curves* assume the presence of random fluctuations in skill during skill acquisition (see Figure 10.2, 4). In other words, although there is a general tendency to excel developmentally, at any given time the curve can represent progress, regress, or halt in the mastery of a skill.

Clearly, although metaphorically these four types of changes can be described with the set of parameters presented by van Geert, each more complex curve requires more parameters to represent the curve. Moreover, if, as indicated above, the first type of curve, the linear/nonlinear curve, assumes continuity of development, the fourth type of curve allows for discontinuity. The more complex the shape of the curve, the harder it is to describe and, consequently, to quantify the change. What are the methods available to researchers interested in quantifying change?

Difference scores: old and new issues

The absolute majority of the publications that appeared in our search – less than 50 per cent – utilized different variants of the difference-scores methodologies, capturing solicited (i.e. provoked by intervention) and unsolicited (i.e. occurring in the course of typical development) change. Thus, here we review the description of this methodology and quantifying changes.

Conventional approaches to measuring change utilized primarily simple change scores, quantified as the difference between raw or standardized scores on pre- and post-tests. In other words, this type of change quantification is based on evaluating the differences across two or more occasions between measures obtained for one variable on a sample of individuals. Thus, the basic questions here are whether there is a change in the mean of the variable and, if so, whether the change is worthy of noting. To address these questions, roughly speaking, two general types of methodologies have been developed. One approach involves a situation where differences between two time measurements are calculated, in raw or standardized form; the magnitude of the difference is referred to as *gain score*. The second approach involves a collection of methodologies that takes into account multiple time measures, considering simultaneously in the analyses the initial and consequent points of measurement. This set of methodologies typically includes various types of analysis of variance and covariance. It is important to note that difference-score approaches are applied in two general contexts – when change is viewed as a result of deliberate intervention (i.e. a program that was designed to trigger a change) and when change is expected as an outcome of normally unfolding developmental processes (i.e. unsolicited change). Below we discuss a number of points relevant to conducting the difference-scores analyses in contexts of investigating both triggered and unsolicited change.

Simple difference scores

Multiple problems arise with the use of simple change scores (e.g. Bereiter 1963; Lord 1952; Schmitz 2001). These problems are due to (1) the apparent lack of reliability of gain scores as compared to reliability levels of the original variables; (2) the presence of ceiling effects in subgroups of a sample; (3) the disturbance of the equal-interval scale assumption encountered at the higher and lower ends of ability distributions with regard to the quantification of gain (e.g. high initial values tend to demonstrate little differences between multiple measurement occasions, thus producing negative correlations between initial status and change); and (4) regression to the mean (i.e. systematic biases originated by the tendency of extreme values for the first occasion to be followed by more average values for the second occasion.

The first and most apparent problem with gain scores is their lack of reliability. This conclusion was initially made by Lord (1952), based on the assumptions of classical test theory. Lord stated that when two imperfect scores are subtracted from each other, the resulting indicators would be even more imperfect. Even when the reliabilities of the first and second measures are medium or high, the reliability of the difference is low! Furthermore, the reliability of a gain score is substantially impacted by the relative standard deviations of the test in a given sample, the placement of the test on the ability distribution, the distribution of item difficulty in the test, and a number of other factors.

The second weakness of the gain scores is the presence of a ceiling effect due to learning that occurs between the two occasions of test administration. Clearly, when the test is administered for the first time, it is often more difficult because of its novelty than during its second administration, when it is more familiar. Thus, it is not unusual that the variation at the second testing occasion is suppressed. This suppression creates a number of problems for interpreting the meaning of gain scores. Thus, if a researcher is interested in working with gain scores, he or she would do well to insure that the pre-test is easy enough for lower-ability students to be differentiated (rather than to floor) and hard enough for the high-ability students not to be differentiated at the post-test (rather than to ceiling).

Third, the nature of the scale on which change is measured is not well understood. What has been clearly demonstrated, however, is that the units of change are not independent of the initial level of performance (i.e. measurement at first occasion). For example, it is a well-known observation that highly able individuals tend to demonstrate smaller gains. A quantitative interpretation of this observation refers to the introduction of a negative bias to the analyses of gain scores. By itself, this is not good, but it can be dealt with statistically. A qualitative interpretation of this observation, however, is more troublesome, because more than one explanation can account for the negative correlation between initial status and change for highly able individuals. The first possibility is that, given the initially high level of performance of highly able individuals at the pre-test, these individuals do not benefit from the intervention administered to the group in between pre-test and post-test. The second possibility is, however, that as a result of training, high-ability individuals tinker with their strategies, demonstrating more efficient and possibly quicker solutions. The third possibility is that high-ability individuals, by the time of the second testing, demonstrate lowered motivation or are bored with the test, which results in their lowered performance. Unfortunately, these three possibilities cannot be distinguished within the simple arithmetic gain-score paradigm of group data analyses. Of the three problems discussed above, the third appears to be the most fundamental one. Apparently, this problem cannot be compensated for by means of the classical test theory approach, but it can be addressed by means of modern psychometric theories, specifically item response theory (IRT, Embretson and Reise 2000; Hambleton, Swaminathan, and Rogers 1991; Hambleton and Slater 1997). IRT is a rapidly developing field of research, introduced to the field of measurement in the early 1950s (Lord 1952), which initially gained popularity in the 1960s, 1970s and 1980s (e.g. Lord 1980; Wright and Stone 1979), and today is the major tool in test development (e.g. Embretson and Reise 2000; Hambleton, Swaminathan, and Rogers 1991). There are also applications of IRT developed for quantifying change (e.g. Embretson 1991).

Finally, another major factor to consider while working with gain scores is the detrimental impact on regression to the mean on their interpretation and understanding (e.g. Campbell and Kenny 1999). The problem is that, in order to establish the presence and to quantify the impact of regression to the mean, one usually needs to have more than two time points observing the performance in the sample on a particular test. Clearly, the need to have more than two time points challenges the very reason for using gain scores – they are easy to compute and understand because they are based on only two observations, and, if more than two observations are available, other methodologies can be used that quantify change more reliably!

Much professional attention has been devoted to ways of compensating for these weaknesses in change scores (e.g. Cohen and Cohen 1975; Cronbach and Furby 1970; Rasch 1980), but none of the procedures that have been developed have been universally accepted (e.g. Campbell and Kenny 1999; Embretson 1994, 1996; Rogosa 1995). Yet, it has been stated that gain scores can be used to characterize differences in performance on two occasions (Schmitz 2001). But in using gain scores, it is necessary to be aware of the criticisms of this methodology.

Due, in part, to multiple warnings in the literature with regard to simple difference scores, researchers today rarely use this methodology. In fact, in our search of the literature, we have not observed a single publication that used simple difference scores. Yet, given the purpose of this chapter, we thought that the discussion of simple difference scores was warranted. Having done that, we can move on to the discussion of the results of our literature search.

Quantifying triggered change: how do we know that an intervention worked?

As indicated above, a set of methodologies has been developed to take into account simultaneously pre- and post-intervention scores (note that post-intervention scores can be obtained more than once, right after the intervention).

A number of recent volumes offer excellent treatment of families of relevant methodologies (Campbell and Kenny 1999; Cohen, Cohen, West and Aiken 2003; Moskowitz and Hershberger 2002; Reise and Duan 2003), so here we limit our discussion only to the summary of our literature search.

In the reviewed publications, we encountered a number of methodologies used to quantify the change occurring as a result of an intervention of some kind. The majority of the studies (Benson et al. 1997; Blachman et al. 1999; Berninger et al. 1999, 2000; Chambers et al. 1998; Chera and Wood 2003; Churches et al. 2002; Das et al. 1995; Dryer et al. 1993; Elkind et al. 1993; Facoetti et al. 2003; Gillon 2000, 2002; Goldstein and Obrzut 2001; Greaney et al. 1997; Greenway 2002; Guyer et al. 1993; Habib et al. 2002; Hart et al. 1997; Hatcher 2000; Hatcher, Hulme and Ellis 1994; Hecht and Close 2002; Ho et al. 2001; Lovett and Steinbach 1997; Lovett et al. 2000; Lundberg 1995; McCarthy et al. 1995; Morris et al. 2000; Nelson et al. 1996; Oakland et al. 1998; O'Shaughnessy and Swanson 2000; Pogorzelski and Wheldall 2002; Poskiparta et al. 1999; Post et al. 2001; Schneider et al. 1997, 1999, 2000; Uhry and Shepherd 1997; Vadasy et al. 2002; van Daal and Reitsma 1999; Wheldall 2000) used the traditional pre-/post-test intervention design, with some studies reporting multiple follow-up points and both immediate and delayed effects of intervention (e.g. Blachman et al. 1999; Gillon 2002; McCarthy et al. 1995; Schneider et al. 1997, 1999, 2000) and performance on transfer tasks (e.g. Benson et al. 1997).

With regard to incorporating both pre- and post-test data, three approaches were dominant – one using repeated analysis of variance including time and group variables as factors in the analysis (e.g. O'Shaughnessy and Swanson 2000; Uhry and Shepherd 1997), one covarying the pre-test scores using analysis of variance for group comparison (e.g. Chambers et al. 1998), and one using paired t-tests (e.g. Bouldoukian et al. 2002). There also were studies in which, although reported, the performance at the baseline was not controlled for (e.g. Churches et al. 2002). The majority of the studies contained a no-treatment control group, but there were studies comparing effectiveness of various treatments (e.g. Berninger et al. 1999, 2000; Dryer et al. 1993; Goldstein and Obrzut 2001; Graham and Wong 1993; Greaney et al. 1997; Hatcher, Hulme, and Ellis 1994; Lovett and Steinbach 1997; Lovett et al. 2000; Nelson et al. 1996; Post et al. 2001; Schneider et al. 2000; van Strien et al. 1995), effectiveness of the same treatment for various groups (e.g. Hatcher 2000; Schneider et al. 1999; van Daal and Reitsma 1999), or added effectiveness of programs administered consecutively (e.g. Vadasy et al. 2002). Some studies reported only pre- to post-test differences, without using control groups (e.g. Greenway 2002; Uhry and Shepherd 1997).

In addition, a number of researchers used the criterion-based approach, where training was administered until the a-priori criterion was reached (in such cases the change was quantified as the number of sessions necessary to master the criterion – e.g. Camarata, Nelson and Camarata 1994) or training was administered for a certain duration and the percentage of responses or participants reaching the criterion was presented (e.g. Stone and Connell 1993; Swisher et al. 1995).

Finally, a number of publications, in addition to tracking change, explored physiological correlates/causes of the observed psychological change (e.g. Richards et al. 2000, 2002; Simos et al. 2002; Stein, Richardson and Fowler 2000).

Clearly, a variety of methodologies are used to quantify change in response to an intervention or multiple interventions. In completing this section on the analysis of change scores in response to intervention, we want to comment on issues of summative interpretations of indicators of change. Since we systematically reviewed the literature in two developmental domains, language and reading, we encountered a number of studies in which the same outcome variables (e.g. reading comprehension) were targeted through a variety of interventions (e.g. word-level versus sub-word-level interventions). Not surprisingly, the reports presented a variety of outcomes. In the context of interpreting and discussing difference-scores techniques, we would like to mention the methodology of meta-analysis – a way to summarize data from a number of diverse intervention studies in an attempt to describe a 'meta-change'. Here we provide a brief illustration of meta-analyses using a portion of the data we encountered in our literature search. Again, there are a number of recent publications presenting technical foundations (Arthur, Bennet and Huffcutt 2001; Lipsey and Wilson 2001) and illustrations (e.g. Swanson 1999) of meta-analysis.

In our illustrative mini-meta-analysis, we identified all publications that used educational interventions in an attempt to enhance reading comprehension. The data in these publications were reported in a variety of ways ranging from means and standard deviations at pre- and post-tests (a number of publications presented delayed follow-up evaluations as well) to inferential statistics (e.g. t- and F-tests). The data were uniformly converted into d-statistics (standard difference). In addition, the studies were coded for the following variables: (1) type of target population (typically developing children, children at-risk for academic failure, and children with special academic needs); (2) type of study (studies registering comparative gains in the intervention and control groups or studies registering change from pre- to post-test); and (3) type of outcome (immediate versus delayed). The combined sample included thirteen studies and 1086 participants (Chambers et al. 1998; Dryer et al. 1993; Elkind et al. 1993; Gillon 2000; Goldstein and Obrzut 2001; Graham and Wong 1993; Greenway 2002; Hatcher et al. 1994; Lovett et al. 2000; Morris 2000; O'Shaughnessy et al. 2000; Schneider et al. 2000; Wheldall 2000). The outcomes of the studies were not uniform with standardized differences ranging from −0.083 to 1.60.

In conducting these summary analyses, we wanted to ask three questions. First, we wanted to estimate the degree of modifiability of comprehension skills in response to *any* kind of intervention. Second, we were interested in a comparison of the effect size of change across different groups of participants involved in the reviewed intervention studies. Finally, we were interested in the impact of intervention on the immediate versus delayed performance. Correspondingly, three types of meta-analyses were carried out. First, we explored the effects of targeted intervention on reading comprehension. Whether quantified in a controlled experimental design involving multiple intervention groups or in a quasi-experimental pre-/post-test design, the triggered change was statistically significant (d=0.362, Z=4.704, $p < 0.001$ and d=0.808, Z=4.81, $p < 0.001$) for controlled and quasi-experiments, respectively). However, interesting moderator effects were identified in this meta-analysis. Apparently, the interventions were much more effective for children with special needs and at-risk children than for typical readers. Finally, the effect size estimated immediately after the intervention was substantial and greatly exceeded the effect size estimated with a time delay (d=0.439, Z=6.28, $p < 0.001$ and d=0.038, Z=0.502, $p > 0.1$ for immediate and delayed outcomes, respectively). Thus, meta-analytic approaches are extremely helpful in deriving summative estimates of change from a number of studies and stratifying these estimates for subsamples and subconditions in which the change occurs.

Quantifying unsolicited change: observing natural development

A large portion (24) of the screened papers reported unsolicited developmental changes (i.e. changes not caused by experimental manipulations originated by the experimenter) in various language- and reading-related psychological processes. Methodology-wise, these articles were heterogeneous, not illustrating, in particular, any specific statistical approach, but providing a glance at the types and distributions of the techniques utilized in the attempt to describe and quantify self-occurring developmental changes.

The champion analytical methodology is that of analysis of variance, both univariate and multivariate. These techniques were employed in thirteen publications surveyed (Fazio, Naremore and Connell 1996; Hadley 1998; Hadley and Rice 1996; Johnston et al. 2001; Joseph et al. 2002; Lyytinen et al. 2001, 2003; Manis et al. 1993; Pennington and Lefly 2001; Snowling et al. 1996; 2000; Sprenger-Charolles et al. 2000; Viholainen et al. 2002), with the majority of publications employing the repeated measures analysis of variance (Fazio, Naremore and Connell 1996; Hadley 1998; Hadley and Rice 1996; Johnston et al. 2001; Joseph et al. 2002; Lyytinen et al. 2001, 2003; Manis et al. 1993; Pennington and Lefly 2001; Snowling et al. 1996; Viholainen et al. 2002).

In addition, there were applications where the follow-up performance was evaluated with other techniques suitable for analysis of change (e.g. Mann-Whitney test, Korhonen 1995; paired *t*-test, Boulet et al. 1998; Pharr et al. 2000; Rescorla and Roberts 2002; Rescorla 2000, 2002).

One of the most commonly encountered methodological tasks in this set of publications is that of predicting the future status of the variable of interest based on the same outcome variable as measured at the baseline (e.g. predicting reading comprehension in Grade 5 based on reading comprehension in Grade 2) or on a set of related variables, which themselves predict the outcome variable both concurrently and longitudinally. There, if the outcome variable was continuous, the analytical technique used the most for this type of task was that of linear regression. In studying the relevant literature we found both use of theory-driven, hierarchical regression (e.g. Gallagher, Frith and Snowling 2000; Lyytinen et al. 2001; Manis and Custodio 1993; McGee et al. 2002; Mirak and Rescorla 1998; Rescorla 2002; Wessling and Reitsma 2001) and data-driven, stepwise regression (e.g. Fazio, Naremore and Connell 1996) approaches. We also found examples of simultaneous regression (Lewis et al. 2000; Olofsson and Niedersoe 1999). A number of studies utilized methodologies of path analyses (Olofsson and Niedersoe 1999) and structural equation modelling (Laakso et al. 1999). If the outcome variable was categorical (e.g. whether the child is diagnosable with developmental dyslexia in Grade 2 based on some measure collected in kindergarten), then authors used logistic regression (e.g. Gallagher, Frith and Snowling 2000).

Another oft-observed application, which, in essence, is an extension of the prediction application above, was that of prediction of group membership (i.e. whether the child is diagnosable with the same condition, non-condition, or some other condition) at the follow-up, given the membership in a certain group at baseline (e.g. being diagnosed with Specific Language Impairment, SLI). Two preferred techniques are used in establishing group membership based on developmental data – logistic regression and discriminant analyses (e.g. Hurford et al. 1993, 1994, 2002; Pennington and Lefly 2001). These applications included both predicting membership at follow-up time(s) based on relevant indicators at baseline (e.g. Hurford et al. 1993, 1994, 2002; Pennington and Lefly 2001; Snowling et al. 2000), and predicting change in the group membership from baseline to follow-up (Manis et al. 1999).

Finally, we encountered a number of studies that used developmental frequency analyses (e.g. changes in percentage of certain types of linguistic errors over time). The majority of these publications presented the data in forms of percentages and frequencies and carried out qualitative analyses of these data (Hadley and Rice 1996; Joseph et al. 2002; Rescorla and Roberts 2002).

In addition to the articles mentioned above, yet another good source of examples of capturing change in the context of unsolicited (e.g. normal) development is a collection of essays describing major longitudinal studies in the United States (Phelps, Furstenberg and Colby 2002).

Case studies

A substantial number of articles (50) reported change data from individual cases; these data were collected either in the process of unfolding developmental change (e.g. Anderson 1999; Cipriani et al. 1998; Eyer and Leonard 1995) or as an outcome of intervention (e.g. Berninger 2000; Brooks 1995; Brunsdon et al. 2002; Butler et al. 2000; Daly and Martens 1994; Hillis 1993; Louis et al. 2001; Miller and Felton 2001; Ottem 2001; Peach 2002; Silliman et al. 2000; Yampolsky and Waters 2002). Not surprisingly, a variety of analytical approaches were used to summarize and present these data. Arguably, all modes of quantifying change described in this chapter can be applied to analysing case study data; what matters for the choice of an analytical approach is the number and frequency of measurements. Among the evaluated publications, many researchers used presentation of assessment data before, during and after the intervention (Butler et al. 2000; Daly and Martens 1994; Daly et al. 2002; Miller and Felton 2001; Ottem 2001; Yampolsky and Waters 2002) or at various stages of longitudinal evaluation (e.g. Cipriani et al. 1998), and graphical representations of change in time (e.g. Anderson 1999; Brooks 1995; Daly and Martens 1994; Daly et al. 2002; Hillis 1993; Louis et al. 2001; Peach 2002; Yampolsky and Waters 2002). Furthermore, in a number of publications, the change was captured through tracking performance on individual items (i.e. treating items as observational units). In this regard, a number of analytic techniques assessing change categorically are relevant; specifically, logistic regression (e.g. Brunsdon et al. 2002), contingency tables analyses (e.g. Brunsdon et al. 2002; Louis et al. 2001; Yampolsky and Waters 2002) and repeated measures analysis of variance (Louis et al. 2001) can be utilized for the analysis of change. Researchers also analyse differences in quality and quantity of errors prior to and after intervention (e.g. Brunsdon et al. 2002; Hillis 1993). Finally, much attention is given to qualitative analyses of change (e.g. Anderson 1999; Berninger 2000; Eyer and Leonard 1995; Silliman et al. 2000).

Although we have not encountered a realization of the so-called person-oriented longitudinal statistical analysis in our review, we find it necessary to mention this approach here. This type of analysis addresses changes at the individual level by using configural statistical analyses and latent transition analyses, especially relevant to applications dealing with individual and small-group data and for studying short-term development (for review, see Bergman, Magnusson and El-Khouri 2003).

As apparent from this brief summary, a variety of methodologies are used for quantifying change, both solicited (arising in response to intervention), and unsolicited (attributable to the normal course of development). What is characteristic of the studies described above is that the majority of them used a two-timepoint design; some of the presented studies dealt with more than two points, but treated these additional time points as delayed outcomes rather than points on time trajectories. Below we present a number of methodologies utilized for analyses of change through multiple time points.

Growth curves

When there are more than two points in a timeframe within which change is evaluated, growth-curve analyses are often utilized. The logic of growth-curve modelling is, maybe surprisingly so, quite similar to that of the difference scores. Although, initially, the term *growth curve* simply referred to a graphical representation of change over time (e.g. changes in height across years from birth in an individual as compared to the group mean, Scammon 1927), now, commonly, growth curve models are referred to as slopes-as-outcomes models, and this reference reveals the meaning of these models. In other words, growth curve models are designed to measure the rate of change, just as the difference-scores approaches are, but they do so much more reliably than the gain-scores models because they are able to incorporate multiple repeated measures minimizing the measurement error. It is fair to say that growth curves, within the last 20 to 30 years, have gained sufficient popularity and now appear to be one of the most widely studied and applied analytical techniques (McArdle 2001).

However, this popularity of the growth curve methodologies creates a certain terminological confusion. Specifically, in this chapter, when we presented theoretical models of change, we also used the term *growth curves.* Just in the paragraph above, we implied that any graphical representation of the change in a trait over time could be considered a growth curve. Finally, below, we briefly summarize various statistical methodologies referred to as growth curve analyses. This multiplicity of the meaning of the term *growth curve*, although unfortunate, is inevitable. Thus, it is important to keep the context of the discussion in mind when talking about growth curves.

In addition to the brief summary of types of theoretical growth curves developed to capture different types of change, one more piece of theory needs to be introduced prior to the following discussion of growth curve methodologies of statistical analyses. This piece of theory relates to two fundamental concepts of change – absolute and normative change and/or stability (Baltes et al. 1977). Normative stability or change is defined exclusively at the group level and relates to the concept of developmental stage. Here what matter are the milestones of development established for humanity as a whole or its particular subgroup (culture, nation, tribe and so on). Absolute stability or change can be defined at both the individual and group levels. In its limit, absolute stability at the individual level assumes a lack of any fluctuation on a trait between different occasions of the trait measurement. Absolute stability at the group level assumes a lack of fluctuation of the group mean between different measurement occasions. To appreciate the relevance of this theoretical distinction to our discussion, consider the graphical representation of growth curves produced by Scammon (1927) – he plotted absolute change at the individual level against normative change at the group level.

In the majority of studies of change (especially in developmental psychology), it is typically assumed that all members in a sample of interest change in correspondence with some underlying common trajectory (i.e. in reference to normative change or stability), but each participant might follow this trajectory with specific deviations. For example, for a typical sample of seven-year-olds, during the period of reading acquisition it is assumed that reading will be mastered; what is of interest is individual variation in the mastery of the skill.

Following McArdle (2001), in this brief review of relevant issues in growth curve analyses, we structure the discussion below along the following three types of growth curves: (1) linear models, (2) non-linear models, and (3) multivariate growth curves.

Linear models

The essence of these types of models is in the assumption that a simple straight line can be fitted into a set of measurements. If, however, various nonlinear curvatures occur, then a small set of power polynomials could be used to describe them. Under this type of modelling, each individual is assumed to demonstrate the trajectory of the skill acquisition that resembles a straight line, which can be characterized by the intercept and slope. These individual growth curves can be averaged to represent the group growth curve. If curvatures are observed, then they can be characterized by higher-order parameters (e.g. acceleration), and these parameters can also be characterized at individual and group levels. The main idea here is that, for individual differences in growth to be captured and characterized accurately, individual data should be collected within some normative samples (e.g. a bunch of seven-year-olds mastering reading). An example of differentiating individual and group data comes from the work on the vocabulary growth in a normative sample of two-year-olds: the individual growth curves in this study inexorably demonstrated upward curvature (acceleration), but showed many individual differences in the rate of change (velocity) (Huttenlocher, Haight, Bryk and Seltzer 1991).

The relevant simplicity of this approach resulted in its widespread applications. However, like any approach, this approach has a number of weak points. Among these are difficulties associated with (1) estimations of individual parameters, especially for higher-order terms and incomplete data; (2) inability of even higher-order polynomials to capture complex developmental shifts; (3) lack of interpretability of many individual parameters; and (4) instability of meaning of individual parameters (McArdle 2001). There are a number of attempts in the field to respond to these criticisms and yet preserve the simplicity of linear modelling. Such attempts refer to ideas of dividing the development into phases or stages connected by critical points, where the dynamics of development changes its course (e.g. Bryk and Raudenbush 1992), or whether introducing a way to reorganize the observed data through finding some 'latent curves' driving the measurements (e.g. Meredith and Tisak 1990).

Nonlinear models

The psychological literature contains many examples of nonlinear growth (e.g. Bock and Thissen 1980; Seber and Wild 1989; Zeger and Harlow 1987). The main thrust behind nonlinear application of growth curve modelling is the attempt to model observations over a wide range of normative groups (e.g. different ages or different life periods), minimizing the number of parameters in the model. These types of models are referred to as composite models because they are based on multiple functions, each of which describes a specific part of the modelled growth process (e.g. Bock and Thissen 1980; Hauspie et al. 1991; Preece and Baines 1978). However, the interpretation of the fits of alternative models and individual parameters still remains a challenge (e.g. Browne and du Toit 1991).

Multivariate models

Relatively few applications illustrating analysis of multiple variables using growth curves are available in developmental psychology. However, a number of recent examples are of great importance to this emerging field. For example, McArdle and Woodcock (1997) demonstrated, using data on multiple cognitive abilities, that a model conceptualized as a latent growth model of a single factor provides a poor fit for the data. To develop a methodology allowing developmental explorations of a multivariate system, McArdle (2001) suggested an application of structural equation modelling. Muthén et al. (2003) provide yet another example illustrating how sets of reading-development-related variables can be modelled by means of growth mixture modelling. What is especially interesting in this application is the capacity of the model to handle different observed variables collected at different times by linking them to sequential pathways of reading development; specifically, in this application, multiple measures of word recognition in Grade 1 are predicted by multiple measures of phonemic awareness in kindergarten.

http://dx.doi.org/10.1017/CBO9780511489938.012

Current methodological work is resulting in the development of increasingly sophisticated growth curves, encompassing both the models' general (Seber and Wild 1989) and specific (e.g. Sayer and Cumsille 2001) aspects. The sophistication relates to multiple levels of hierarchy (Raudenbush 2001), dealing with unbalanced, incomplete or missing data (Bryk and Raudenbush 1992; Pinherio and Bates 2000; Verbeke and Molenberghs 2000), and attempting to capture the possible discontinuity of growth (Osgood 2001; Rovine and Molenaar 2001). Thus, growth curve modelling approaches have strong representation in theoretical literature. What about empirical literature?

Our literature search resulted in identifying twenty-four studies using growth curve approaches. Two of these studies (Aro et al. 1999; Hick et al. 2002) used the graphical variant of growth curve approaches, meaning that they graphed individual trajectories on children across time. Twenty-two articles utilized different statistical approaches to growth curve analyses; the majority of these models were univariate growth curve models; both linear (e.g. Tressoldi et al. 2001) and non-linear (e.g. Kemper, Rice and Chen 1995) models were fitted. All articles presented group data, with the exception of one that analysed changes in performance of a single boy (Robinson and Mervis 1999). Based on the research questions answering which growth curve analyses were implemented, these publications can be divided into three groups.

The first group of papers ($N = 6$) used growth curve analyses for the purposes of determining the impact of a given intervention. The central question here is whether, in response to intervention, children show more growth than is expected by chance (e.g. Abbott et al. 1997; Campbell et al. 2001; Stage et al. 2003). Similarly, often the question of interest is the one comparing the growth in different groups subjected to different interventions (Abbott and Berninger 1999; Foorman et al. 1997; Kappers 1997). To obtain answers to these questions, group statistics (e.g. the slope of the group curve) are analysed. Moreover, the analyses of group curves are helpful in understanding the role of other mediating or moderating variables (Foorman et al. 1997). In addition, a number of other conclusions can be derived from the analyses of individual curves. For example, in analysing slopes of individual growth curves, authors engaged in the discussion of who did and did not gain from the intervention and why (Abbott et al. 1997; Abbott and Berninger 1999; Kappers 1997) and what third factors, other than baseline performance and the type of intervention, influence the gain (Foorman et al. 1997; Kappers 1997). A number of researchers reported that the best fits were the models obtained when higher-order polynomial terms were included (e.g. Campbell et al. 2001). Some researchers found that the linear models were not adequate for the representation of growth rate across larger chunks of lifespan; multiple curves needed to be fitted to accommodate the data (e.g. Campbell et al. 2001).

The second group of articles (N=10) employed growth curve analyses to investigate unsolicited changes across time (Burchinal et al. 2002; Flowers et al. 2000; Francis et al. 1996; Jacobson 1999; Landry et al. 1997; Meyer et al. 1998; Rescorla, Mirak and Singh 2000; Rice et al. 1998, 2000; Shaywitz et al. 1999) or for program evaluation, where all students were provided with a curriculum (e.g. Stage 2001). Typical research questions in these studies had to do with identification of third variables (demographic-, ability-, personality-based) that somehow impacted or differentiated developmental pathways of interest. Similarly to the group of papers above, group and individual growth curves were analysed. A number of studies (e.g. Burchinal et al. 2002; Flowers et al. 2000; Francis et al. 1996; Landry et al. 1997; Meyer et al. 1998) utilized multi-level models, investigating associations between variables collected within nested designs. Similarly, a number of papers included higher-order polynomial terms (Burchinal et al. 2002; Francis et al. 1996; Landry et al. 1997; Rescorla, Mirak and Singh 2000; Rice et al. 1998, 2000; Shaywitz et al. 1999). Some researchers used growth curve modelling to verify various theoretical developmental models (Francis et al. 1996; Jacobson 1999).

We found only one study (Compton 2000) that combined the first and second types of growth curve modelling by using the data obtained from modelling unsolicited change to divide the sample of children into subgroups so that the outcome of intervention was maximized; consequently, the interventions themselves were modelled with growth curve approaches.

In sum, growth curves are well represented in both theoretical and empirical literature dealing with change.

Time series

Collectively, approaches utilizing many measurement points distributed in time are referred to as time-series approaches. But how many is enough? The rule of thumb is at least twenty. Fifty is better. Whereas fifty and more time measures are typical for specific branches in psychology (e.g. psychophysiology), they are not seen as often in developmental psychology. Clearly, both difference-scores approaches (when only two points in time are used) and growth curve methodologies are variants of time series with a limited number of observations.

Time-series approaches exist in an overwhelming variety of shapes and forms (Brillinger 2001). They are used for purposes of describing sequential data, estimating various parameters with the goal of generating a stochastic model capturing the dynamic of the time series, and identifying the system behind the sequential data. One other special case of general time-series approaches is chaos modelling (e.g. Alligood et al. 1997). The distinct feature of chaos models is their sensitive dependence upon initial condition, but the dependence is such that the generated time paths appear to be random (Brock 2001). If all parameters of the system are within their pre-determined limits, the system converges to homeostasis. However, a slight violation of parameter limitations can result in the system's manifestation of chaotic behaviours.

Time-series models constructed for a single variable are characterized by trend and rhythm. The indicator of trend shows the nature of change of quantity associated with the variable of interest across time (e.g. a variable demonstrates a linear trend if the increase from one measurement occasion to the next is constant). The indicators of rhythm capture the periodicity of the change in the variable (e.g. the amount of office noise is usually higher during work days and lower during weekends). If models contain more than one variable changing in time, these time series can be characterized by synchronicity of the relationships between variables (e.g. typically the development of reading skill and vocabulary are synchronous, but in cases of specific reading disabilities, the development of these skills is desynchronized). The majority of time-series applications deal with data stretched in time on the same scale structured by a time-like parameter (for a review, see Brillinger 2001). However, the assumption of interval is not crucial for time series. For example, Markov approaches allow the modelling of data in which the assumptions of interval data are not realized (e.g. Gottman and Roy 1990).

Like other sciences dealing with issues of change over time, developmental and education psychology are gradually incorporating various ideas generated by theories addressing sequential data (for a review, see Collins and Sayer 2001). In the context of this chapter, we will review briefly only one general methodological approach encountered in our literature search – that of dynamic systems.

Boker (2001) distinguishes growth curve models and dynamic systems models by stating that the former models generate predictions regarding a single trajectory of central tendency (referred to as the attractor, a single point) whereas the latter models generate predictions regarding multiple trajectories (referred to as a basin of attraction, a vector field plot). If the growth curve models use multiple measurements as a way to characterize the entire curve, the dynamic systems models use multiple measurements to hypothesize and verify hidden patterns characteristic of a system changing in time. In doing so, the dynamic systems models rely on the current value at any particular point in the curve and its first derivative (the rate of change or the velocity) to predict the second derivative of any particular point (the acceleration) (Piccinin 2001).

The technique used for these types of analyses is referred to as differential structural equation modelling (Boker 2001; Piccinin 2001). Although a new and exciting direction, differential structural equation modelling (dSEM) has some theoretical and practical problems that need to be addressed both theoretically and empirically (Piccinin 2001). Specifically, little is known about the validity of assumptions central to dSEM, such as the homogeneity of shape of attractor basin across individuals and variability in the estimates of first- and second-order derivatives within and across individuals. Similarly, there are no certain answers to questions such as how many measurement points are needed, how they should be spaced in time, and for what psychological processes these models are of use.

Although the dynamic system theory (also referred to as systems theory) is gaining popularity in the developmental literature, very few studies have utilized the applied methodologies developed on this theory's basis (e.g. Gogate and Walker-Andrews 2001; Thelen 1989; Thelen and Smith 1994). A special interest in the dynamic systems approaches in psychology is attributable, in part, to the realization made in a large body of developmental literature that states that development, largely, is discontinuous (e.g. van Geert 1997; Pascual-Leone and Baillargeon 1994) and includes abrupt changes, stabilizations and non-linear and linear gains and losses (e.g. Wimmers, Beek, Savelsbergh and Hopkins 1998). Yet, the demands of data collection (many data points are needed) and data analyses (rather sophisticated data-analytical approaches are utilized) prevent, at this point, these methodologies from wide-scale adoption.

Dynamic systems models operate under the general assumption that under what are mostly stable pressures of independent variables, including biological foundations such as genetic makeup, or environmental context such as educational patterns or SES, dependent variables (cognitive, behavioural, and social–emotional variables) can attain relatively stable states (attractors). Attractors can change suddenly, however, with changes to independent variables. For our purposes, that is to say, if the relative stability of reading-acquisition skill is challenged by an appropriate intervention, the reading-acquisition skill could fall apart, or be transformed into a different (higher-order) skill in a discontinuous manner. Methodologies based on nonlinear dynamic systems claim well-defined ways to establish linkages between changes in independent variables and changes in attractors.

Let us return to van Geert (1997) for an illustration. In syllogistic reasoning there supposedly exists a method of constructing syllogisms that combines previous outputs of reasoning development (a child's ability to form a conjunction with the logical operator *and*) with an external intervention (instruction on how to solve linear syllogisms). An adequate intervention would transfer the first stage of reasoning development into a new stage of reasoning development (knowing how to solve linear syllogisms). According to the dynamic systems approach, such a transformation could be of a discontinuous nature. After the skill has been mastered, however, each subsequent intervention would strengthen the reasoning but would keep it within the attractor stage, thus safeguarding the principle of finding the solution to a syllogism task. Van Geert provides yet another example of a dynamic systems model, the so-called Verhulst model (van Geert 1991, 1993, 1994), which has been explored primarily in the domain of language development (Ruhland and van Geert 1998). In this model, a variable called the 'grower' (whose change in time can be observed) departs from an arbitrary level (or 'seed'), and grows (increases or decreases) to reach a state of equilibrium (which in turn can change as an outcome of some disturbance).

While the dynamic systems approach continues to gain a modicum of popularity, it is still a minor player in the developmental psychology literature. One such example of a dynamic model is Thelen and Smith's dynamic model of motor development (1994, 1997). Another example of a dynamic systems approach is linked to the model of transition in catastrophe theory (Thom 1975), the mathematical theory that allows the detection of phase shifts in dynamic systems (i.e. a system in which time-based change is inherent). Eight mathematically defined indicators define the catastrophe (or transformation) stage of a dynamic system (Gilmore 1981; van der Maas and Molenaar 1992). Certain of these indicators (bimodality of the trait distribution, inaccessibility of the skill, sudden jump, anomalous variance, and critical slowing down) have been studied in the context of developmental research (e.g. the development of analogical reasoning, Hosenfeld, van der Maas and van den Boom 1997b and the conservation research, Hartelman et al. 1998).

Although the number of specific applications of the models described above is somewhat limited in psychological literature, researchers (Hosenfeld, van der Maas and van den Boom 1997a; Thomas 1989; Thomas and Lohaus 1993; Thomas and Turner 1991) have had success in locating several indicators of bimodality (separating those who have and have not mastered the skills) in developmental data in performance on tasks such as conservation, classification, the understanding of horizontality and verticality, and analogical reasoning. The presence of bimodality stresses the importance of taking non-linearity of development into account.

Currently, assumptions of linearity form the basis of most testing models (and corresponding data-analytic procedures) – that is, most models suppose that any effect is proportional to the magnitude of the input of some controlling variables (the better the intervention, the better the outcome). However, numerous reported observations have shown that interventions lead to proportional effects only up to a certain point (or starting from a certain point). When (or before) a certain threshold is reached, however, the impact of the intervention may change both qualitatively and quantitatively. And here is the point most relevant to our discussion: non-linear effects of intervention and the nonlinear nature of parameters are often ignored in the context of traditional group-difference-based approaches to quantifying change. For example, van der Maas and Molenaar (1992) discuss the possibility that the impact of standardized training might be especially substantial for those children who are close to the acquisition of the skill, but very small for those children who are far away from mastery of the skill. Specifically, if the distance to the 'mastery point' is short, it is possible for interventions to successfully introduce the necessary dynamics into the system (instability, expressed as increased variability in performance, in terms of van der Maas and Molenaar 1992) to insure transition to mastery.

Moreover, non-linear dynamic systems can be used to model the impact of an intervention where complex dynamic forms can be realized from relatively simple equations (Glass and Mackey 1988; Newell and Molenaar 1998; Robinson and Mervis 1998). Here, the identification of relevant developmental variables that might act as critical variables in the shaping of system dynamics over time is a primary goal. Two assumptions underlie these analyses: (1) the mastery of a skill is the product of the coalescence of numerous constraints to action imposed by critical variables (e.g. Newell and Molenaar 1998); (2) large-scale qualitative changes in emerging skills can result from small qualitative changes in critical variables; and (3) the stability-transformation dynamics of the emergence and the transformation of skills are linked to the interplay between sets of cooperating and competing critical variables. Researchers constantly work on the development, realization and application of newly developed methodologies (e.g. van Geert and van Dijk 2002).

Concluding thoughts

In this chapter, we briefly reviewed major methodologies currently applied in the fields of developmental and educational psychology for the quantification of change. We started by describing types of change distinguishable theoretically and proceeded with providing illustrations for methodologies available for quantification of change. Four different types of change were described; each type of change can be characterized by a different parameter set addressing the change's continuity–discontinuity, linearity–non-linearity, and rhythm of change. We stated that simpler models of change require fewer time points and use less sophisticated analytical procedures; the demand for great detail of measurement in time and the complexity of the analytic technique used correspondingly increases with the complexity of the modelled change. To illustrate this assertion, we structured the review with an empirical analysis of the citations generated in a systematic literature search within two domains of development – the acquisition of language and reading. The frequency analysis of the dominant methodologies in the encountered citations indicates that the most popular type of analysis is that of difference scores, followed by growth curves. The dynamic-system methodologies are still relatively infrequent in studies of change. Simpler analytical approaches are linked to more global interpretations of the data (e.g. whether the change occurred), whereas more sophisticated approaches are linked to the generation of complex models of development (e.g. what kind of change and under what conditions it occurred).

The main objective of this chapter was to evaluate the degree of 'penetration' of the methodological developments regarding measurements of change into the domain of presentation of empirical results obtained in educational and developmental psychology. As it appears from the observations described above, there are a limited number of applications of complex methodologies. There are multiple reasons for this situation. First, complex methodologies are designed to meet the needs of complex datasets. Such complex datasets are difficult to collect and require much time and personnel commitment; therefore, the number of illustrations of applications of such methodologies is limited by the lack of complex datasets suitable for these methodologies. In other words, the introduction of change from utilization of conventional data analytic strategies to capitalizing on novel analytic strategies requires a systemic change, allowing for time and effort in collecting the data structures suitable for novel methodologies.

Second, there are certain areas of developmental and educational psychology in which the 'call for' methodologies of dealing with change is comparatively greater than in others. Often, however, these areas are somewhat remote from mainstream fashionable areas of psychology. One such area is that of dynamic testing and assessment. The dynamic assessment approach, by its very meaning, implies quantification of change: in this paradigm what is typically looked at is the modification in performance between first and second administration of a test (or a testing item) occurring in response to an intervention. Yet, at this stage of its existence, dynamic testing is a methodology utilized primarily in clinical settings and with relatively small sample sizes (for a review, see Sternberg and Grigorenko 2002). There appears to be a disconnect between the theoretical thought on developing complex methods of quantification of change and the applied development of the field in which precise methods of quantification of change are most needed.

Third, we limited our review to publications on reading and language. Clearly, this is a limited selection of domains of development; as we pointed out earlier, the literature on motor development has more examples of the utilization of complex methodologies of change quantification. Thus, it is possible that the distribution of frequencies of different methodologies illustrated in Figure 10.1 will be different if other developmental domains are surveyed.

Finally, the reality of empirical research in the fields of developmental and educational psychology is such that it calls for a variety of methodologies applicable to different tasks, datasets, and contexts. In this chapter we attempted to illustrate how different methods of quantifying change can be applied in a variety of different situations and how and why it is important to take into consideration their respective strengths and weaknesses.

REFERENCES

Abbott, S. P. and Berninger, V. W. (1999). It's never too late to remediate: teaching word recognition to students with reading disabilities in grades 4–7. *Annals of Dyslexia, 49*, 223–50.

Abbott, S. P., Reed, E., Abbott, R. D. and Berninger, V. W. (1997). Year-long balanced reading/writing tutorial: a design experiment used for dynamic assessment. *Learning Disability Quarterly, 20*(3), 249–63.

Alligood, K. T., Sauer, T. D. and Yorke, J. A. (1997). *Chaos. An introduction to dynamic systems.* New York: Springer.

Anderson, R. (1999). Impact of first language loss on grammar in a bilingual child. *Communication Disorders Quarterly, 21*(1), 4–16.

Aro, M., Aro, T., Ahonen, T., Raesaenen, T., Hietala, A. and Lyytinen, H. (1999). The development of phonological abilities and their relation to reading acquisition: case studies of six Finnish children. *Journal of Learning Disabilities, 32*(5), 457–63, 478.

Arthur, W., Jr, Bennett, W., Jr and Huffcutt, A. I. (2001). *Conducting meta-analysis using SAS*. Mahwah, NJ: Erlbaum.

Baltes, P. B., Reese, H. W. and Nesselrode, J. R. (1997). *Life-span developmental psychology: introduction to research methods.* Monterey, CA: Brooks/Cole.

Benson, N. J., Lovett, M. W. and Kroeber, C. L. (1997). Training and transfer-of-learning effects in disabled and normal readers: evidence of specific deficits. *Journal of Experimental Child Psychology, 64*(3), 343–66.

Bereiter, C. (1963). Some persisting dilemmas in the measurement of change. In C. W. Harris (ed.) *Problems in measuring change* (pp. 3–20). Madison: University of Wisconsin Press.

Bergman, L. R., Magnusson, D. and El-Khouri, B. M. (2003). *Studying individual development in an interindividual context.* Mahwah, NJ: Erlbaum.

Berninger, V. W. (2000). Dyslexia the invisible, treatable disorder: the story of Enstein's ninja turtles. *Learning Disability Quarterly, 23*(3), 175–95.

Berninger, V. W., Abbott, R. D., Brooksher, R., Lemos, Z., Ogier, S., Zook, D. et al. (2000). A connectionist approach to making the predictability of English orthography explicit to at-risk beginning readers: evidence for alternative, effective strategies. *Developmental Neuropsychology, 17*(2), 241–71.

Berninger, V. W., Abbott, R. D., Zook, D., Ogier, S., Lemos-Britton, Z. and Brooksher, R. (1999). Early intervention for reading disabilities: teaching the alphabet principle in a connectionist framework. *Journal of Learning Disabilities, 32*(6), 491–503.

Blachman, B. A., Tangel, D. M., Ball, E. W., Black, R. and McGraw, C. K. (1999). Developing phonological awareness and word recognition skills: a two-year intervention with low-income, inner-city children. *Reading and Writing, 11*(3), 239–73.

Bock, R. D. and Thissen, D. (1980). Statistical problems of fitting individual growth curves. In F. E. Johnston, A. F. Roche, C. Susanne (eds.) *Human physical growth and maturation: methodologies and factors* (pp. 265–90). New York, NY: Plenum.

Boker, S. M. (2001). Differential structural equation modelling of intraindividual variability. In L. M. Collins and A. G. Sayer (eds.) *New methods for the analysis of change* (pp. 3–27). Washington, DC: APA.

Bouldoukian, J., Wilkins, A. J. and Evans, B. J. (2002). Randomised controlled trial of the effect of coloured overlays on the rate of reading of people with specific learning difficulties. *Ophthalmic and Physiological Optics, 22*(1), 55–60.

Boulet, D., Bryce, S., Bliss, L. S. and Linebaugh, C. (1998). Listener judgments of impaired narration at two age levels. *Journal of Children's Communication Development, 20*(1), 1–8.

Brillinger, D. R. (2001). Time series: general. In N. J. Smelser and P. B. Baltes (eds.) *Social and behavioral sciences* (vol. 23, pp. 15724–31). Amsterdam, Holland: Elsevier.

Brock, W. A. (2001). *Chaos theory*. In N. J. Smelser & P. B. Baltes (eds.) *International encyclopedia of the social and behavioral sciences* (pp. 1643–6). Amsterdam, Holland: Elsevier.

Brooks, P. (1995). A comparison of the effectiveness of different teaching strategies in teaching spelling to a student with severe specific difficulties/dyslexia. *Educational and Child Psychology, 12*(1), 80–8.

Browne, M. W. and Du Toit, S. H. C. (1991). Models for learning data. In L. M. Collins and J. L. Horn (eds.) *Best methods for the analysis of change* (pp. 47–68). Washington: American Psychological Association.

Brunsdon, R. K., Hannan, T. J., Coltheart, M. and Nickels, L. (2002). Treatment of lexical processing in mixed dyslexia: a case study. *Neuropsychological Rehabilitation, 12*(5), 385–418.

Bryk, A. and Raudenbush, S. (1992). *Hierarchical linear models in social and behavioral research: applications and data analysis methods*. Newbury Park, CA: Sage.

Burchinal, M. R., Peisner-Feinberg, E., Pianta, R. and Howes, C. (2002). Development of academic skills from preschool through second grade: family and classroom predictors of developmental trajectories. *Journal of School Psychology, 40*(5), 415–36.

Butler, D. L., Elaschuk, C. L. and Poole, S. (2000). Promoting strategic writing by postsecondary students with learning disabilities: a report of three case studies. *Learning Disability Quarterly, 23*(3), 196–213.

Camarata, S. M., Nelson, K. E. and Camarata, M. N. (1994). Comparison of conversational-recasting and imitative procedures for training grammatical structures in children with specific language impairment. *Journal of Speech and Hearing Research, 37*(6), 1414–23.

Campbell, D. T. and Kenny, D. A. (1999). *A primer on regression artifacts*. New York: Guilford Press.

Campbell, F. A., Pungello, E. P., Miller-Johnson, S., Burchinal, M. and Ramey, C. T. (2001). The development of cognitive and academic abilities: growth curves from an early childhood educational experiment. *Developmental Psychology, 37*(2), 231–42.

Chambers, B., Abrami, P. C., Massue, F. M. and Morrison, S. (1998). Success for all: evaluating an early-intervention program for children at risk of school failure. *Canadian Journal of Education, 23*(4), 357–72.

Chera, P. and Wood, C. (2003). Animated multimedia 'talking books' can promote phonological awareness in children beginning to read. *Learning and Instruction, 13*(1), 33–52.

Churches, M., Skuy, M. and Das, J. P. (2002). Identification and remediation of reading difficulties based on successive processing deficits and delay in general reading. *Psychological Reports, 91*(3, pt. 1), 813–24.

Cipriani, P., Bottari, P., Chilosi, A. M. and Pfanner, L. (1998). A longitudinal perspective on the study of specific language impairment: the long-term follow-up of an Italian child. *International Journal of Language and Communication Disorders, 33*(3), 245–80.

Cohen, J. and Cohen, P. (1975). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Cohen, J., Cohen, P., West, S. G. and Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences.* Mahwah, NJ: Lawrence Erlbaum.

Collins, L. M. and Horn, J. L. (eds.) (1991). *Best methods for the analysis of change.* Washington, DC: American Psychological Association.

(2001). *Best methods for the analysis of change.*Washington, DC: APA.

Collins, L. M. and Sayer, A. G. (eds.) (2001). *New methods for analysis of change.* Washington, DC: APA.

Compton, D. L. (2000). Modeling the response of normally achieving and at-risk first grade children to word reading instruction. *Annals of Dyslexia, 50*, 53–84.

Cronbach, L. and Furby, L. (1970). How should we measure change – or should we? *Psychological Bulletin, 74*, 68–70.

Daly, E. J. and Martens, B. K. (1994). A comparison of three interventions for increasing oral reading performance: application of the instructional hierarchy. *Journal of Applied Behavior Analysis, 27*(3), 459–69.

Daly, E. J., Murdoch, A., Lillenstein, L., Webber, L. and Lentz, F. E. (2002). An examination of methods for testing treatments: conducting brief experimental analyses of the effects of instructional components on oral reading fluency. *Education and Treatment of Children, 25*(3), 288–316.

Das, J. P., Mishra, R. K. and Pool, J. E. (1995). An experiment on cognitive remediation of word-reading difficulty. *Journal of Learning Disabilities, 28*(2), 66–79.

Dryer, R., Beale, I. L. and Lambert, A. J. (1993). The balance model of dyslexia and remedial training: an evaluative study. *Journal of Learning Disabilities, 32*(2), 174–86.

Elkind, J., Cohen, K. and Murray, C. (1993). Using computer-based readers to improve reading comprehension of students with dyslexia. *Annals of Dyslexia, 43*, 238–59.

Embretson, S. E. (1991). A multidimensional latent trait model for measuring learning and change. *Psychometrika, 56*, 495–516.

(1994). Comparing changes between groups: Some perplexities arising from psychometrics. In D. Laveault, B. D. Zumbo, M. E. Gessaroli and M. W. Boss (eds.) *Modern theories of measurement: problems and issues*. Ottawa: Edumetric Research Group, University of Ottawa.

(1996). Item response theory models and inferential bias in multiple group comparisons. *Applied Psychological Measurement, 20*, 201–12.

Embretson, S. E. and Reise, S. R. (2000). *Item response theory for psychologists.* Mahwah, NJ: Erlbaum Lawrence.

Eyer, J. A. and Leonard, L. B. (1995). Functional categories and specific language impairment: a case study. *Language Acquisition: A Journal of Developmental Linguistics, 4*(3), 177–203.

Facoetti, A., Lorusso, M. L., Paganoni, P., Umilta, C. and Mascetti, G. G. (2003). The role of visuospatial attention in developmental dyslexia: evidence from a rehabilitation study. *Cognitive Brain Research, 15*(2), 154–64.

Fazio, B. B., Naremore, R. C. and Connell, P. J. (1996). Tracking children from poverty at risk for specific language impairment: a three-year longitudinal study. *Journal of Speech and Hearing Research, 39*(3), 611–24.

Fischer, K. W. and Rose, S. P. (1994). Development of coordination of components in brain and behavior: a framework for theory and research. In G. Dawson and K. W. Fisher (eds.) *Human behavior and the developing brain* (pp. 3–66). New York: Guilford.

Flowers, L., Meyer, M., Lovato, J., Wood, F. and Felton, R. (2000). Does third grade discrepancy status predict the course of reading development? *Annals of Dyslexia, 51*, 49–74.

Foorman, B. R., Francis, D. J., Winikates, D., Mehta, P., Schatschneider, C. and Fletcher, J. M. (1997). Early interventions for children with reading disabilities. *Scientific Studies of Reading, 1*(3), 255–76.

Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A. and Fletcher, J. M. (1996). Developmental lag versus deficit models of reading disability: a longitudinal, individual growth curves analysis. *Journal of Educational Psychology, 88*(1), 3–17.

Gallagher, A., Frith, U. and Snowling, M. J. (2000). Precursors of literacy delay among children at genetic risk of dyslexia. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 41*(2), 202–13.

Gillon, Gail T. (2000). The efficacy of phonological awareness intervention for children with spoken language impairment. *Language, Speech and Hearing Services in the Schools, 31*(2), 126–41.

(2002). Follow-up study investigating the benefits of phonological awareness intervention for children with spoken language impairment. *International Journal of Language and Communication Disorders, 37*(4), 381–400.

Gilmore, R. (1981). *Catastrophe theory for scientists and engineers.* New York: Wiley.

Glass, L. and Mackey, M. (1988). *From clocks to chaos.* Princeton, NJ: Princeton University Press.

Gogate, L. J. and Walker-Andrews, A. (2001). The intersensory origins of word comprehension: an ecological-dynamic systems view. *Developmental Science, 4*, 1–37.

Goldstein, B. H. and Obrzut, J. E. (2001). Neuropsychological treatment of dyslexia in the classroom setting. *Journal of Learning Disabilities, 34*(3), 276–85.

Gottman, J. M. (ed.) (1995). *The analysis of change*. Mahwah, NJ: Erlbaum.

Gottman, J. M. and Roy, A. K. (1990). *Sequential analysis: a guide for behavioral researchers.* Cambridge: Cambridge University Press.

Graham, L. and Wong, B. Y. (1993). Comparing two modes of teaching a question-answering strategy for enhancing reading comprehension: didactic and self-instructional training. *Journal of Learning Disabilities, 26*(4), 270–9.

Greaney, K. T., Tunmer, W. E. and Chapman, J. W. (1997). Effects of rime-based orthographic analogy training on the word recognition skills of children with reading disability. *Journal of Educational Psychology, 89*(4), 645–51.

Greenway, C. (2002). The process, pitfalls and benefits of implementing a reciprocal teaching intervention to improve the reading comprehension of a group of year 6 pupils. *Educational Psychology in Practice, 18*(2), 113–37.

Guyer, B. P., Banks, S. R. and Guyer, K. E. (1993). Spelling improvement for college students who are dyslexic. *Annals of Dyslexia*, 43, 186–93.

Habib, M., Rey, V., Daffaure, V., Camps, R., Espesser, R., Joly-Pottuz, B. et al. (2002). Phonological training in children with dyslexia using temporally modified speech: a three-step pilot investigation. *International Journal of Language and Communication Disorders, 37*(3), 289–308.

Hadley, P. A. (1998). Early verb-related vulnerability among children with specific language impairment. *Journal of Speech Language and Hearing Research, 41*(6), 1384–97.

Hadley, P. A. and Rice, M. L. (1996). Emergent uses of BE and DO: evidence from children with specific language impairment. *Language Acquisition: A Journal of Developmental Linguistics, 5*(3), 209–43.

Hambleton, R. K. and Slater, S. C. (1997). Item response theory models and testing practices: current international status and future directions. *European Journal of Psychological Assessment, 13*, 21–8.

Hambleton, R. K., Swaminathan, H. and Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park, CA: Sage.

Harris, C. W. (ed.) (1963). *Problems in measuring change*. Madison; University of Wisconsin Press.

Hart, T. M., Berninger, V. M. and Abbott, R. D. (1997). Comparison of teaching single or multiple orthographic-phonological connections for word recognition and spelling: implications for instructional consultation. *School Psychology Review, 26*(2), 279–97.

Hartelman, P. A. I., van der Maas, H. L. J. and Molenaar, P. C. M. (1998). Detecting and modelling developmental transitions. *The Journal of Developmental Psychology, 16*, 97–122.

Hatcher, P. J. (2000). Sound links in reading and spelling with discrepancy-defined dyslexics and children with moderate learning difficulties. *Reading and Writing, 13*(3–4), 257–72.

Hatcher, P. J., Hulme, C. and Ellis, A. W. (1994). Ameliorating early reading failure by integrating the teaching of reading and phonological skills: the phonological linkage hypothesis. *Child Development, 65*(1), 41–57.

Hecht, S. A., and Close, L. (2002). Emergent literary skills and training time uniquely predict variability in responses to phonemic awareness training in disadvantaged kindergarteners. *Journal of Experimental Child Psychology, 82*(2), 93–115.

Hick, R. F., Joseph, K. L., Conti-Ramsden, G., Serratrice, L. and Faragher, B. (2002). Vocabulary profiles of children with specific language impairment. *Child Language Teaching and Therapy, 18*(2), 165–80.

Hillis, A. E. (1993). The role of models of language processing in rehabilitation of language impairments. *Aphasiology, 7*(1), 5–26.

Ho, C. S., Lam, E. Y. and Au, A. (2001). The effectiveness of multisensory training in improving reading and writing skills of Chinese dyslexic children. *Psychologia, 44*(4), 269–80.

Hosenfeld, B., van der Maas, H. L. J. and van den Boom, D. C. (1997a). Detecting bimodality in the analogical reasoning performance of elementary schoolchildren. *International Journal of Behavioral Development, 20*, 529–47.

(1997b). Indicators of discontinuous change in the development of analogical reasoning. *Journal of Experimental Child Psychology, 64*, 367–95.

Hurford, D. P., Darrow, L. J., Edwards, T., Howerton, C. J., Mote, C. R., Schauf, J. D. et al. (1993). An examination of phonemic processing abilities in children during their first-grade year. *Journal of Learning Disabilities, 26*(3), 167–77.
Hurford, D. P., Potter, T. S. and Hart, G. S. (2002). Examination of three techniques for identifying first-grade children at risk for difficulty in word identification with an emphasis on reducing the false negative error rate. *Reading Psychology, 23*(3) 2, 159–80.
Hurford, D. P., Schauf, J. D., Blaich, T., Bunce, L. and Moore, K. (1994). Early identification of children at risk for reading disabilities. *Journal of Learning Disabilities, 27*, 371–82.
Huttenlocher, J., Haight, W., Bryk, A. and Seltzer, M. (1991). Early vocabulary growth: relation to language input and gender. *Developmental Psychology, 27*, 236–49.
Jacobson, C. (1999). How persistent is reading disability? Individual growth curves in reading. *Dyslexia: The Journal of the British Dyslexia Association. 5*(2), 78–93.
Johnston, J. R., Miller, J. and Tallal, P. (2001). Use of cognitive state predicates by language-impaired children. *International Journal of Language and Communication Disorders, 36*(3), 349–70.
Joseph, K. L., Serratrice, L. and Conti-Ramsden, G. (2002). Development of copula and auxiliary BE in children with specific language impairment and younger unaffected controls. *First Language, 22*(65, pt. 2), 137–72.
Kappers, E. J. (1997). Outpatient treatment of dyslexia through stimulation of the cerebral hemispheres. *Journal of Learning Disabilities, 30*(1), 100–24.
Kemper, S., Rice, K. and Chen, Y. (1995). Complexity metrics and growth curves for measuring grammatical development from five to ten. *First Language, 15*(44, pt. 2), 151–66.
Korhonen, T. T. (1995). The persistence of rapid naming problems in children with reading disabilities: a nine-year follow-up. *Journal of Learning Disabilities, 28*(4), 232–9.
Laakso, M., Poikkeus, A. and Lyytinen, P. (1999). Shared reading interaction in families with and without genetic risk for dyslexia: implications for toddlers' language development. *Infant and Child Development, 8*(4), 179–95.
Landry, S. H., Smith, K. E., Miller-Loncar, C. L. and Swank, P. R. (1997). Predicting cognitive-language and social growth curves from early maternal behaviors in children at varying degrees of biological risk. *Developmental Psychology, 33*(6), 1040–53.
Lewis, B. A., Freebairn, L. A. and Taylor, H. G. (2000). Academic outcomes in children with histories of speech sound disorders. *Journal of Communication Disorders, 33*(1), 11–30.
Lipsey, M. W. and Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
Lord, F. M. (1952). A theory of mental test scores. *Psychometric Monograph No 7.*
(1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum.
Louis, M., Espesser, R., Rey, V., Daffaure, V., Di Cristo, A. and Habib, M. (2001). Intensive training of phonological skills in progressive aphasia: a model of

brain plasticity in neurodegenerative disease. *Brain and Cognition, 46*(1–2), 197–201.

Lovett, M. W. and Steinbach, K. A. (1997). The effectiveness of remedial programs for reading disabled children of different ages: does the benefit decrease for older children? *Learning Disability Quarterly, 20*(3), 189–210.

Lovett, M. W., Lacerenza, L., Borden, S. L., Frijters, J. C., Steinbach, K. A. and De Palma, M. (2000). Components of effective remediation for developmental reading disabilities: combining phonological and strategy-based instruction to improve outcomes. *Journal of Educational Psychology, 92*(2), 263–83.

Lundberg, I. (1995). The computer as a tool of remediation in the education of students with reading disabilities: a theory-based approach. *Learning Disability Quarterly, 18*(2), 89–99.

Lyytinen, P., Eklund, K. and Lyytinen, H. (2003). The play and language behavior of mothers with and without dyslexia and its association to their toddlers' language development. *Journal of Learning Disabilities, 36*(1), 74–86.

Lyytinen, P., Poikkeus, A., Laakso, M., Eklund, K. and Lyytinen, H. (2001). Language development and symbolic play in children with and without familial risk for dyslexia. *Journal of Speech Language and Hearing Research, 44*(4), 873–85.

Manis, F. R., Custodio, R. and Szeszulski, P. A. (1993). Development of phonological and orthographic skill: a two-year longitudinal study of dyslexic children. *Journal of Experimental Child Psychology, 56*(1), 64–86.

Manis, F. R., Seidenberg, M. S., Stallings, L., Joanisse, M., Bailey, C., Freedman, L. et al. (1999). Development of dyslexic subgroups: a one-year follow up. *Annals of Dyslexia, 49*, 105–34.

McArdle, J. (2001). Growth curve analysis. In N. J. Smelser and P. B. Baltes (eds.) *Social and behavioral sciences* (vol. 9, pp. 6439–45). Amsterdam, Holland: Elsevier.

McArdle, J. and Woodcock, J. (1997). Expanding test-retest designs to include developmental time-lag components. *Psychological Methods, 2*, 403–35.

McCarthy, P., Newby, R. F. and Recht, D. R. (1995). Results of an early intervention program for first grade children at risk for reading disability. *Reading Research and Instruction, 34*(4), 273–94.

McGee, R., Prior, M., Williams, S., Smart, D. and Sanson, A. (2002). The long-term significance of teacher-rated hyperactivity and reading ability in childhood: findings from two longitudinal studies. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 43*(8), 1004–16.

Meredith, W. and Tisak, J. (1990). Latent curve analysis. *Psychometrika, 55*, 107–22.

Meyer, M. S., Wood, F. B., Hart, L. A. and Felton, R. H. (1998). Longitudinal course of rapid naming in disabled and nondisabled readers. *Annals of Dyslexia, 48*, 91–114.

Miller, L. L. and Felton, R. H. (2001). 'It's one of them . . . I don't know': case study of a student with phonological, rapid naming, and word-finding deficits. *Journal of Special Education, 35*(3), 125–33.

Mirak, J., and Rescorla, L. (1998). Phonetic skills and vocabulary size in late talkers: concurrent and predictive relationships. *Applied Psycholinguistics, 19*(1), 1–17.

Morris, D., Tyner, B., and Perney, J. (2000). Early steps: replicating the effects of a first-grade reading intervention program. *Journal of Educational Psychology, 92*(4), 681–93.

Moskowitz, D. S. and Hershberger, S. L. (eds.) (2002). *Modeling intraindividaul variability with repeated measures data: methods and applications.* Mahwah, NJ: Erlbaum.

Muthén, B., Khoo, S.-T., Francis, D. J. and Boscardin, C. K. (2003). Analysis of reading skills development from kindergarten through first grade: an application of growth mixture modeling to sequential process. In S. P. Reise and N. Duan (eds.) *Multilevel modeling*. Mahwah, NJ: Erlbaum.

Nelson, K. E., Camarata, S. M., Welsh, J., Butkovsky, L. and Camarata, M. (1996). Effects of imitative and conversational recasting treatment on the acquisition of grammar in children with Specific Language Impairment and younger language-normal children. *Journal of Speech and Hearing Research*, *39,* 850–9.

Newell, K. M., and Molenaar, P. C. M. (1998). *Applications of nonlinear dynamics to developmental process modeling*. Mahwah, NJ: Lawrence Erlbaum.

Oakland, T., Black, J. L., Stanford, G., Nussbaum, N. L. and Balise, R. R. (1998). An evaluation of the dyslexia training program: a multisensory method for promoting reading in students with reading disabilities. *Journal of Learning Disabilities, 31*(2), 140–7.

Olofsson, A. and Niedersoe, J. (1999). Early language development and kindergarten phonological awareness as predictors of reading problems: from 3 to 11 years of age. *Journal of Learning Disabilities, 32*(5), 464–72.

Osgood, D. W. (2001). Application of multilevel models to the analysis of change. In L. M. Collins and A. G. Sayer (eds.) *New methods for the analysis of change* (pp. 97–104). Washington, DC: APA.

O'Shaughnessy, T. E. and Swanson, H. L. (2000). A comparison of two reading interventions for children with reading disabilities. *Journal of Learning Disabilities, 33*(3), 257–77.

Ottem, E. (2001). Use of pictographic-articulatory symbols to promote alphabetic reading in a language-impaired boy: case study. *AAC: Augmentative and Alternative Communication, 17*(1), 52–60.

Oyama, S. (2000). *The ontogeny of information: developmental systems and evolution.* Durham, NC: Duke University Press.

Pascual-Leone, J. (1995). Learning and development as dialectical factors in cognitive growth. *Human Development, 38*, 338–48.

Pascual-Leone, J. and Baillargeon, R. (1994). Developmental measurement of mental attention. *International Journal of Behavioral Development, 17*, 161–200.

Peach, R. K. (2002). Treatment for phonological dyslexia targeting regularity effects. *Aphasiology, 16*(8), 779–89.

Pennington, B. F. and Lefly, D. L. (2001). Early reading development in children at family risk for dyslexia. *Child Development, 72*(3), 816–33.

Pharr, A. B., Ratner, N. B. and Rescorla, L. (2000). Syllable structure development of toddlers with expressive specific language impairment. *Applied Psycholinguistics, 21*(4), 429–49.

Phelps, E., Furstenberg, F. F. and Colby, A. (eds.) (2002). *Looking at lives: American longitudinal studies of the twentieth century.* New York: Russell Sage Foundation.

Piccinin, A. M. (2001). Differential structural equation modeling of intraindividual variability. In L. M. Collins and A. G. Sayer (eds.) *New methods for the analysis of change* (pp. 29–32). Washington, DC: APA.
Pinherio, J. C. and Bates, D. M. (2000). *Mixed-effect models in S and S-PLUS.* New York, NY: Springer.
Pogorzelski, S., and Wheldall, K. (2002). Do differences in phonological processing performance predict gains made by older low-progress readers following intensive literacy intervention? *Educational Psychology, 22*(4), 413–27.
Poskiparta, E., Niemi, P. and Vauras, M. (1999). Who benefits from training in linguistic awareness in the first grade, and what components show training effects? *Journal of Learning Disabilities, 32*(5), 437–46, 456.
Post, Y. V., Carreker, S. and Holland, G. (2001). The spelling of final letter patterns: a comparison of instruction at the phoneme and the rime. *Annals of Dyslexia, 51*, 121–46.
Preece, M. A. and Baines, M. K. (1978). A new family of mathematical models describing the human growth curve. *Ann. Hum. Biol., 7*, 507–28.
Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press (original work published in 1960: Rasch, G. *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danmarks Paedagogiske Institut).
Raudenbush, S. W. (2001). Toward a coherent framework for comparing trajectories of individual change. In L. M. Collins and A. G. Sayer (eds.) *New methods for the analysis of change* (pp. 33–64). Washington, DC: APA.
Reise, S. P. and Duan, N. (2003). *Multilevel modeling*. Mahwah, NJ: Lawrence Erlbaum.
Rescorla, L. (2000). Do late-talking toddlers turn out to have reading difficulties a decade later? *Annals of Dyslexia, 50*, 87–102.
(2002). Language and reading outcomes to age nine in late-talking toddlers. *Journal of Speech Language and Hearing Research. 45*(2), 360–71.
Rescorla, L. and Roberts, J. (2002). Nominal versus verbal morpheme use in late talkers at ages three and four. *Journal of Speech Language and Hearing Research, 45*(6), 1219–31.
Rescorla, L., Mirak, J. and Singh, L. (2000). Vocabulary growth in late talkers: lexical development from 2;0 to 3;0. *Journal of Child Language, 27*(2), 293–311.
Rice, M. L., Wexler, K. and Hershberger, S. (1998). Tense over time: the longitudinal course of tense acquisition in children with specific language impairment. *Journal of Speech Language and Hearing Research, 41*(6), 1412–31.
Rice, M. L., Wexler, K., Marquis, J. and Hershberger, S. (2000). Acquisition of irregular past tense by children with specific language impairment. *Journal of Speech Language and Hearing Research, 43*(5), 1126–45.
Richards, T. L., Berninger, V. W., Aylward, E. H., Richards, A. L., Thomson, J. B., Nagy, W. E. et al. (2002). Reproducibility of proton MR spectroscopic imaging (PEPSI): comparison of dyslexic and normal-reading children and effects of treatment on brain lactate levels during language tasks. *American Journal of Neuroradiology, 23*(10), 1678–85.
Richards, T. L., Corina, D., Serafini, S., Steury, K., Echelard, D. R., Dager, S. R. et al. (2000). Effects of a phonologically driven treatment for dyslexia on lactate levels

measured by proton MR spectroscopic imaging. *American Journal of Neuroradiology, 21*(5), 916–22.

Robinson, B. F. and Mervis, C. B. (1998). Disentangling early language development: modeling lexical and grammatical acquisition using an extension of case-study methodology. *Developmental Psychology, 34*, 363–75.

(1999). Comparing productive vocabulary measures from the CDI and a systematic diary study. *Journal of Child Language, 26*(1), 177–85.

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

Rogosa, D., Brandt, D. and Zimowskyk, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin, 92*, 726–48.

Rovine, M. J. and Molenaar, P. C. M. (2001). A structural equations modeling approach to the general linear mixed model. In L. M. Collins and A. G. Sayer (eds.) *New methods for the analysis of change* (pp. 65–96). Washington, DC: APA.

Ruhland, R. and van Geert, P. (1998). Jumping into syntax: transitions in the development of closed class words. *British Journal of Developmental Psychology, 16*, 65–95.

Sayer, A. G. and Cumsille, P. R. (2001). Second-order latent growth models. In L. M. Collins and A. G. Sayer (eds.) *New methods for the analysis of change* (pp. 177–200). Washington, DC: APA.

Scammon, R. E. (1927). The first seriatim study of human growth. *American Journal of Physical Anthropology, 10*, 329–36.

Seber, G. A. F. and Wild, C. J. (1989). *Nonlinear models.* New York, NY; Wiley.

Schmitz, B. (2001). Change: methods of studying. In N. J. Smelser and P. B. Baltes (eds.) *International encyclopedia of the social and behavioral sciences, vol. 3* (pp. 1640–3). Amsterdam, Holland: Elsevier.

Schneider, W., Ennemoser, M., Roth, E. and Kuespert, P. (1999). Kindergarten prevention of dyslexia: does training in phonological awareness work for everybody? *Journal of Learning Disabilities, 32*(5), 429–36.

Schneider, W., Kuespert, P., Roth, E., and Vise, M. (1997). Short- and long-term effects of training phonological awareness in kindergarten: evidence from two German studies. *Journal of Experimental Child Psychology, 66*(3), 311–40.

Schneider, W., Roth, E. and Ennemoser, M. (2000). Training phonological skills and letter knowledge in children at risk for dyslexia: a comparison of three kindergarten intervention programs. *Journal of Educational Psychology, 92*(2), 284–95.

Shaywitz, S., Fletcher, E., Holahan, J. M., Shneider, A. E., Marchione, K. E., Stuebing, K. K. et al. (1999). Persistence of dyslexia: the Connecticut Longitudinal Study at adolescence. *Pediatrics, 104*, 1351–9.

Silliman, E. R., Jimerson, T. L. and Wilkinson, L. C. (2000). A dynamic systems approach to writing assessment in students with language learning problems. *Topics in Language Disorders, 20*(4), 45–64.

Simos, P. G., Fletcher, J. M., Bergman, E., Breier, J. I., Foorman, B. R., Castillo, E. M. et al. (2002). Dyslexia-specific brain activation profile becomes normal following successful remedial training. *Neurology, 58*(8), 1203–13.

Snowling, M. J., Bishop, D. V. M. and Stothard, S. E. (2000). Is preschool language impairment a risk factor for dyslexia in adolescence? *Journal of Child Psychology and Psychiatry and Allied Disciplines, 41*(5), 587–600.

Snowling, M. J., Goulandris, N. and Defty, N. (1996). A longitudinal study of reading development in dyslexic children. *Journal of Educational Psychology, 88*(4), 653–69.

Sprenger-Charolles, L., Cole, P., Lacert, P. and Serniclaes, W. (2000). On subtypes of developmental dyslexia: evidence from processing time and accuracy scores. *Canadian Journal of Experimental Psychology, 54*(2), 87–104.

Stage, S. A. (2001). Program evaluation using hierarchical linear modeling with curriculum-based measurement reading probes. *School Psychology Quarterly, 16*(1), 91–112.

Stage, S. A., Abbott, R. D., Jenkins, J. R., and Berninger, V. W. (2003). Predicting response to early reading intervention from verbal IQ, reading-related language abilities, attention ratings, and verbal IQ-word readings discrepancy: failure to validate discrepancy method. *Journal of Learning Disabilities, 36*(1), 24–33.

Stein, J. F., Richardson, A. J. and Fowler, M. S. (2000). Monocular occlusion can improve binocular control and reading in dyslexics. *Brain, 123*, 164–70.

Sternberg, R. J. and Grigorenko, E. L. (2002). *Dynamic testing.* New York, NY: Cambridge University Press.

Stone, C. A. and Connell, P. J. (1993). Induction of a visual symbolic rule in children with specific language impairment. *Journal of Speech and Hearing Research, 36*(3), 599–608.

Swanson, H. L. (1999). Reading research for students with LD: a meta-analysis of intervention outcomes. *Journal of Learning Disabilities, 32*, 504–32.

Swisher, L., Restrepo, M. A., Plante, E. and Lowell, S. (1995). Effect of implicit and explicit 'rule' presentation on bound-morpheme generalization in specific language impairment. *Journal of Speech and Hearing Research, 38*(1), 168–73.

Thelen, E. (1989). Self-organization in developmental process: can systems approaches work? In M. Gunnar and E. Thelen (eds.) *Systems and development. The Minnesota Symposium on Child Psychology, 22* (pp. 77–117). Mahwah, NJ: Erlbaum.

Thelen, E. and Smith, L. B. (1994). *A dynamic systems approach to the development of cognitions and action.* Cambridge, MA: Bradford Books, MIT.

(1997). Dynamic systems theories. In W. Damon (series ed.) and R. M. Lerner (vol. ed.), *Handbook of child psychology, vol. 1: theoretical models of human development* (5th edn, pp. 563–634). New York, NY: Wiley.

Thom, R. (1975). *Structural stability and morphogenesis*. Reading, MA; Benjamin.

Thomas, H. (1989). A binomial mixture model for classification performance: a commentary on Waxman, Chambers, Yntema, and Gelman. *Journal of Experimental Child Psychology, 48*, 423–30.

Thomas, H. and Lohaus, A. (1993). Modeling growth and individual differences in spatial tasks. *Monographs of the Society for Research in Child Development, 58*, (9, serial no. 237).

Thomas, H. and Turner, G. F. W. (1991). Individual differences and development in water-level task performance. *Journal of Experimental Child Psychology, 51*, 171–94.

Tressoldi, P. E., Stella, G. and Faggella, M. (2001). The development of reading speed in Italians with dyslexia: a longitudinal study. *Journal of Learning Disabilities, 34*(5), 414–17.

Uhry, J. K. and Shepherd, M. J. (1997). Teaching phonological recoding to young children with phonological processing deficits: the effect on sight-vocabulary acquisition. *Learning Disability Quarterly, 20*(2), 104–25.

Vadasy, P. F., Sanders, E. A., Peyton, J. A. and Jenkins, J. R. (2002). Timing and intensity of tutoring: a closer look at the conditions for effective early literacy tutoring. *Learning Disabilities Research and Practice, 17*(4), 227–41.

van Daal, V. H. P. and Reitsma, P. (1999). Effects of outpatient treatment of dyslexia. *Journal of Learning Disabilities, 32*(5), 447–56.

van der Maas, H. L. J. and Molenaar, P. M. C. (1992). Stagewise cognitive development: an application of catastrophe theory. *Psychological Review, 99*, 395–417.

van Geert, P. (1991). A dynamic systems model of cognitive and language growth. *Psychological Review, 98*, 3–53.

(1993). A dynamic systems model of cognitive growth: competition and support under limited resource conditions. In L. B. Smith and E. Thelen (eds.) *A dynamic systems approach to development: applications.* (pp. 265–331). Cambridge, MA: MIT Press.

(1994). *Dynamic systems of development: change between complexity and chaos.* New York: Prentice Hall/Harvester Wheatsheaf.

(1995). Dimensions of change: a semantic and mathematical analysis of learning and development. *Human Development, 38*, 322–31.

(1997). Variability and fluctuation: a dynamic view. In E. Amsel and K. A. Renninger (eds.) *Changes and development: issues of theory, method, and application* (pp. 193–212). Mahwah, NJ: Erlbaum.

van Geert, P. and van Dijk, M. (2002). Focus on variability: new tools to study intra-individual variability in developmental data. *Infant Behavior and Development, 25*, 340–74.

van Strien, J. W., Stolk, B. D. and Zuiker, S. (1995). Hemisphere-specific treatment of dyslexia subtypes: better reading with anxiety-laden words? *Journal of Learning Disabilities, 28*(1), 30–4.

Verbeke, G. and Molenberghs, G. (2000). *Linear mixed models for longitudinal data.* New York, NY: Springer.

Viholainen, H., Ahonen, T., Cantell, M., Lyytinen, P. and Lyytinen, H. (2002). Development of early motor skills and language in children at risk for familial dyslexia. *Developmental Medicine and Child Neurology, 44*(11), 761–9.

Von Eye, A. and Niedermeier, K. E. (1999). *Statistical analysis of longitudinal categorical data in the social and behavioral sciences.* Mahwah, NJ: Lawrence Erlbaum.

Vygotsky, L. S. (1982). *Myshlenie i rech'* [Thought and language]. Moskva: Pedagogika.

Wessling, R. and Reitsma, P. (2001). Preschool phonological representations and development of reading skills. *Annals of Dyslexia, 51*, 203–29.

Wheldall, K. (2000). Does Rainbow Repeated Reading add value to an intensive literacy intervention program for low-progress readers? An experimental evaluation. *Educational Review, 52*(1), 29–36.

Wimmers, R. H., Beek, P. J., Savelsbergh, G. J. P. and Hopkins, B. (1998). Developmental changes in action: theoretical and methodological issues. *British Journal of Developmental Psychology, 16*, 45–63.

Wright, B. D. and Stone, M. H. (1979). *Best test design.* Chicago: MESA.

Yampolsky, S. and Waters, G. (2002). Treatment of single word oral reading in an individual with deep dyslexia. *Aphasiology, 16*(4–6), 455–71.

Zeger, S. L. and Harlow, S. D. (1987). Mathematical models from laws of growth to tools for biological analysis: fifty years of growth. *Growth, 51*, 1–21.