DOCUMENT RESUME

ED 294 923 TM 011 676

AUTHOR Williamson, Gary L.; And Others
TITLE Longitudinal Analyses of Achievement Data from Durham County (NC) Schools.
INSTITUTION North Carolina Univ., Chapel Hill. L.L. Thurstone Psychometric Lab.
SPONS AGENCY North Carolina State Dept. of Public Instruction, Raleigh.
PUB DATE Apr 88
NOTE 37p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 5-9, 1988). Figures and appendix contain marginally legible print.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Ability; *Academic Achievement; Elementary Education; *Individual Development; *Longitudinal Studies; Models; *Psychometrics; Public Schools; Standardized Tests
IDENTIFIERS Aggregation (Data); *Growth Curve Analysis; Hierarchical Linear Modeling; *North Carolina (Durham County)

ABSTRACT

A longitudinal study of achievement data from public schools in Durham County, North Carolina, was undertaken to examine individual growth curves as well as to chart the progress of educational institutions and the effects of aggregation on the study of growth. A two-level hierarchical linear model was used in the study to specify individual growth and to relate the parameters of the growth model to a relevant background variable. Computer analysi was conducted using TIMEPATH, a program written in GAUSS, a matrix programming language developed by Aptech Systems, Inc. Data consistec of eight waves of achievement scores collected in the spring of each year from 1978 to 1985, and three ability scores collected in the fall of 1979 for a cohort of students as they progressed from Grade 1 to 8. Tests included the Prescriptive Reading Inventory, Diagnostic Mathematics Inventory, California Achievement Tests, and tests within the North Carolina Annual Testing Program. Results indicate that the use of growth curves as indices of change has numerous advantages over other methods, and that longitudinal data analyses are feasible for school districts when a suitable longitudinal database has been maintained. Growth curves focus on the use of multi-wave data, provide statistical advantages of multiple occasions of measurement, and emphasize the individual student. Nine tables and two graphs are included. (TJH)

LONGITUDINAL ANALYSES OF ACHIEVEMENT DATA FROM DURHAM COUNTY (NC) SCHOOLS

by

Gary L. Williamson

Mark Appelbaum

and

Alex Epanchin

Paper presented at the annual meeting of the American Educational Research Association, April, 1988 New Orleans, Louisiana.

Longitudinal Analyses of Achievement Data from Durham County (NC) Schools

The purpose of this study is twofold. First, it is intended as an illustration of what can be learned from the analysis of longitudinal data. The emphasis is on examining individual growth curves. Second, the study is intended to show one way of charting the progress of educational institutions, such as schools, and the effects of aggregation on the study of growth.

This project was supported by a contract of the North Carolina State Department of Public Instruction to the L. L. Thurstone Psychometric Laboratory, University of North Carolina at Chapel Hill. Dr. Mark Appelbaum secured the contract and made available the resources of the L. L. Thurstone Psychometric Laboratory for the execution of the study and the completion of this report. Computer programs, data analysis, and report composition were provided by the first author.

Special thanks are due to the Durham County School System for supplying the data. Particular thanks are due Dr. Alex Epanchin, whose vision it was to create and maintain the longitudinal data base that makes this study possible.

Background

Recent Literature

Literature on the measurement of change has traditionally dealt with two occasions (waves) of measurement. The difference score as a measure of change has thus received much of the attention in this literature. However, because of overly restrictive assumptions (see Rogosa and Willett (1983b)) and the lack of a suitable statistical model for change (see Rogosa, Brandt and Zimowski (1982)), many authors have condemned the difference score as a measure of change and have recommended other approaches (e.g., Lord (1956, 1963), Thorndike (1966), Cronbach and Furby (1970), Nunnally (1973), and O'Connor (1972)). Baltes, Reese and Nesselroade (1977) recognized the need for modeling change over time and the fact that perceived problems with traditional approaches to the measurement of change (e.g., low reliability of the difference score) were partly related to the fact that only two occasions had been used to study change.

Recent developments in the measurement of change have clarified many of these issues and have moved a long way toward correcting previous misconceptions. These recent developments have generalized beyond two occasions of measurement to the situation where richer longitudinal data are available.

Blomqvist (1977) used the straight-line growth model to investigate the relationship between change and initial status. Maximum likelihood estimates were provided for the covariance matrix of the parameters of a single measure.

Rogosa, Brandt and Zimowski (1982) emphasized the modeling of individual change and asserted that individual time paths are the "proper focus for the analysis of change." (p. 744) They stated a statistical model for the individual growth curve and examined various assumptions in the measurement of change literature. They also investigated traditional measures of change (the difference score, the improved difference score, the Lord-McNemar regression estimate, Bayesian growth curve estimates, and residual change measures) for two-wave data. They discussed the statistical properties and reliability of each measure. Measures of change for multiwave data were also discussed and the advantages of multiwave data were noted. In their "Mottos for the Measurement of Individual Change", they summarized key points derived from a conceptual and mathematical framework for the measurement of individual change.

Several other references provide new results for the measurement of individual change. Rogosa and Willett (1983b) discussed the reliability of the difference score and pointed out oversights in previous literature. For example, previous investigations of the reliability of the difference score had (perhaps unknowingly) concentrated on cases where individual differences in true change do not exist, or exist only to a small degree. Rogosa and Willett illustrated that the difference score can be a reliable measure, and they showed how restrictive assumptions (e.g., equality of observed score variances at time 1 and time 2; equal reliabilities for each measure) have caused previous investigators to miss this fact.

In more recent work, Willett (1985) and Rogosa and Willett (1985) discussed correlates of change via models for systematic individual differences in growth. Their approach incorporates a model for growth and a model for individual differences in growth. A model with constant rate of growth and various models with nonconstant rates of growth were discussed. (Models with nonconstant rates of growth include polynomial, linear state dependence, logistic and simplex models.) Rogosa, Floden and Willett (1984) and Rogosa and Willett (1983a) considered the use of tracking indices to assess the stability of individual differences over time.

Rogosa, Willett and Williamson (1986) used many of the above techniques to analyze achievement scores on the Comprehensive Tests of Basic Skills (CTBS) for a cohort of approximately two hundred high school students. They also illustrated how some common indices from the profile similarity literature can be used to help examine growth on multiple measures. Finally, they introduced a prototypical student achievement report that highlights academic growth on multiple measures.

Williamson (1986) investigated growth on multiple measures. Profiles of individual achievement and growth were investigated using using a straight-line growth model for individual longitudinal data sequences on several academic measures. A specific data analysis procedure was defined that provides (1) univariate and multivariate descriptions of achievement and growth, and (2) identification of intraindividual strengths and weaknesses in achievement and rate of growth.

Most recently, Bryk and Raudenbush (1987) advanced the study of change by formulating a general model framework (hierarchical linear models) and using more sophisticated techniques to estimate the parameters of the models. (The pair of models used by Rogosa and Willett (1985) is a special case of the Bryk and Raudenbush approach.) Hierarchical linear models have been used by Bryk (1987) and

Raudenbush (1987) to study the effects of educational programs on student academic growth.

The recent literature on the measurement of change focuses attention on models for individual growth. Questions about level and rate of learning on a single measure are phrased in terms of quantities derived from a specified model. This desirable property lends itself well to answering questions about growth since estimated parameters provide descriptions of individual performance and can be aggregated to provide descriptions of institutional performance.

Mathematical Model

A two-level hierarchcal linear model is used in this study. First, a model for individual growth is specified. Second, a model is specified to relate the parameters of the first model to some relevant background variable, W, that is thought to be related to growth. Through this approach, relevant growth parameters are identified and their relationship to other variables is made explicit.

The simplest type of individual growth is straight-line growth. In this case, each individual exhibits a constant rate of change. The mathematical model for straight-line growth for individual p is

$$\xi_p(t) = \xi_p(0) + \theta_p t \qquad (1)$$

where $\xi_p(t)$ is the true score of individual p at time t and θ_p is the true rate of change of individual p.

There are numerous reasons for using a straight-line growth model. They include parsimony, ease of interpretation, frequent use by researchers, robustness (i.e., linear models often provide a good approximation even when growth is nonlinear), and the fact that linear models often fit empirical data well.

There are of course limitations involved with the choice of such a simple model. Using a straight-line growth model, the researcher implicitly assumes that growth (however scaled) is constant in time, that the scale is continuous and extends infinitely in one or both directions (positive, negative). In addition, a straight-line model may be inappropriate when certain kinds of measurement problems exist. For example, ceiling bounded effects are not modeled by a straight-line growth curve.

When such problems are detected, alternative score scales or alternative models may be more appropriate. In any case, fitting a straight-line growth model is a reasonable first step in investigating individual growth patterns.

The second level of the hierarchical linear model gives an explicit representation of the relationship of θ_p to some exogenous background variable, W (assumed to be constant over time). As noted by Rogosa and Willett (1985), "Individual differences in growth exist when different individuals have different values of θ_p. Systematic individual differences in growth exist when individual differences in a growth parameter such as θ_p can be linked with one or more W's." (p. 205) A simple relationship is expressed by the model,

$$E(\theta|W) = \mu_\theta + \beta_{\theta W} (W - \mu_W) \qquad (2)$$

where μ_θ is the population mean of θ_p; μ_W is the population mean of W_p; and $\beta_{\theta W}$ is the population regression coefficient. A nonzero value of $\beta_{\theta W}$ indicates that W is a predictor of growth. The correlation, $\rho_{\theta W}$, is often more convenient, and is used in this report.

The two-part model given in (1) and (2) makes it possible to describe individual growth. In addition, it allows the investigation of systematic individual differences in growth.

Computer Programs

Most of the analyses presented in this study were produced by TIMEPATH, a computer program for fitting individual straight-line growth curves to longitudinal data. The program was developed at Stanford University under the supervision and guidance of Professor David R. Rogosa. Ordinary least squares and maximum likelihood estimation are used to accomplish state-of-the art longitudinal data analyses.

TIMEPATH is written in GAUSS, a matrix programming language developed by Aptech Systems, Inc. (see Edlefson and Jones, 1984) for the IBM/PC and compatibles. TIMEPATH is used to fit individual straight-line growth curves, investigate the aptness of the straight-line model, and produce comprehensive statistical summaries relating to individual growth and systematic individual differences in growth. The program also produces an output file that can be used for subsequent analyses.

The Statistical Analysis System for the PC (SAS/PC) was used to perform some descriptive analyses. In addition, SAS/PC proved useful for preprocessing the data before using TIMEPATH, and for file management.

All analyses were done on an IBM/PC AT using SAS/PC or GAUSS Given the general availability of both software and hardware, this study serves as a demonstration of the feasibility of studying achievement and growth using longitudinal data in a school system setting.

Application

Data

The data were supplied by Durham County Schools. They consist of eight waves of achievement scores collected in the spring of each year from 1978 to 1985, and three ability scores collected in the fall of 1979 for a cohort of students as they progressed from grade one to grade eight. Special thanks go to Dr Alex Epanchin for supplying the data in a convenient form and for describing the nature of the data.

The tests and forms are listed in Table 1. Appropriate levels of the Prescriptive Reading Inventory (PRI) and Diagnostic Mathematics Inventory (DMI) were administered to grades one and two in conjunction with the North Carolina Annual Testing Program. Various levels of the California Achievement Tests (CAT) were administered to grades three through eight in conjunction with the North Carolina Annual Testing Program and the local school district testing program. The PRI, DMI and CAT are produced by CTB/McGraw-Hill.

Table 1

Achievement and Ability Tests

Year	Grade	Test*	Form	Level
Achievement Tests (Spring Testing)				
1978	1	PRI		2
		DMI		Red
1979	2	PRI		A-Red
		DMI		Green
1980	3	CAT	C	13
1981	4	CAT	C	14
1982	5	CAT	C	15
1983	6	CAT	C	16
1984	7	CAT	C	17
1985	8	CAT	C	18
Ability Test (Fall Testing)				
1979	3	Cognitive Abilities Test	1	A

*PRI=Prescriptive Reading Inventory
DMI=Diagnostic Mathematics Inventory
CAT=California Achievement Tests

Reading Total Scale Scores and Mathematics Total Scale Scores were chosen as the achievement measures to be analyzed. Scale scores were chosen because they reportedly have properties that make them suitable for longitudinal studies. Scale scores are a normal part of the reporting provided by CTB/McGraw-Hill for the CAT and were routinely reported at grades three through eight. The scale scores for grades one and two were estimated by CTB/McGraw-Hill from the PRI and DMI scores. [Analyses for Language Total and Total Battery are not included since the subtest (and consequently the total battery score) was unavailable for grades one and two.]

Scores on the Cognitive Abilities Tests (Riverside Publishing Company) were collected at several times during the eight years. The scores on these tests during 1979 were chosen as background exogenous variables for the illustration of analyses of systematic individual differences in growth. The theory used in the application of equation (2) requires that background variables be constant over time. While this is probably not true for ability test scores, they provide a convenient illustration of the techniques. The substantive results found here regarding systematic individual differences in growth have meaning only to the extent that ability can be regarded as a constant attribute of the individual.

From the data base, there were 667 observations that had student records for each of the eight years. When cases with missing data were eliminated this reduced to 529 individuals with complete Reading Total records and 527 individuals with complete Mathematics Total records. The individuals were spread among twelve schools in the district. Table 2 shows the frequency of students from each school (identified by an arbitrary two-digit school identification code) and the number of males and females in the total group. These two data files (one for reading, one for math) are the ones used for this study.

Analyses Performed

For each of the two data files, preliminary descriptive analyses were run to check the score distributions for obvious outliers or data errors. Frequency tables showing the number of individuals in each school and gender group were produced. Because of a limitation on the amount of space available to GAUSS on and IBM/PC AT, it was decided to analyze gender groups separately. Thus some statistics (for example, estimates of the reliability of the empirical rates) are reported only for gender groups and not for the total group.

The TIMEPATH program was run for Reading Total Scale Score (male, female) and Mathematics Total Scale Score (male, female). An output file was created containing the estimated individual rates of change and other summary statistics for each analysis group (reading, males; reading, females; mathematics, males; mathematics, females). The four output files were combined into two files (Reading, total group; Mathematics, total group) and SAS/PC was used to determine distribution statistics for the total group and for each school in the district.

Finally effects of aggregation were investigated as follows First the mean growth rate and various centiles in the empirical

Table 2

Sample Distribution for Reading and Mathematics

School	Reading N	Mathematics N
10	50	49
11	43	43
12	27	27
13	53	51
14	72	74
15	31	31
16	4	4
17	25	26
18	70	70
19	57	58
20	41	40
21	56	54
District	529	527
Males	278	277
Females	251	250

distributions of growth rates were calculated for the district and each school. These provide a sound picture of the nature of growth in reading and mathematics at the district and school level. Next, the means and corresponding centiles of *scale scores* were calculated on each occasion of measurement for the district and for each school Then straight-line growth curves were fit to each of these "aggregated" scores (i.e., the scale score means, or centiles). The resulting growth rates, determined from fitting an "institutional growth model", were compared with mean (or centile) growth rates from the empirical distribution of individual growth rates.

The procedure described in the previous paragraphs (particularly the application of the TIMEPATH program) produces a wealth of information about growth. The results described in the next section illustrate what is possible when suitable models for growth are used and longitudinal data are properly maintained. In addition, the analyses make it possible to sensibly chart institutional growth.

Descriptive summaries of scale scores are presented first. Next the aptness of the straight-line model is investigated and results are summarized. Individual growth is discussed, statistical and psychometric characteristics are described, and group summaries are provided. Last, the effects of aggregation are discussed.

Results

Descriptive summary. An examination of the empirical distributions of scores revealed no peculiarities. This was expected since the school district has likely done extensive data cleaning prior to this study

Table 3 shows means and standard deviations for Reading Total and Mathematics Total scale scores at each grade, and for the ability (IQ) test scores in grade three The achievement test means show a steady increase from grade one to grade eight. There appears to be an increase of about 35-40 scale score points per year on both reading and mathematics. The ability test scores indicate that the students were slightly above average in verbal, mathematical and nonverbal ability near the beginning of their school career.

Reporting achievement test score means as in Table 3 is the typical presentation made by educational institutions when trying to demonstrate growth. If each individual in the district is growing in straight-line fashion, then the trend in the means of Table 3 is actually informative about the mean rate of growth in the district. However, the table does not make the rate of growth explicit. Furthermore, this presentation of the longitudinal data does not inform about the growth of individuals in the district, or about the variability in growth across individuals in the district. In the worst case, it is possible that the growth function implied by the means is entirely different from the growth exhibited by individuals in the district. Finally, the table does not show an explicit connection between the growth implied by the achievement test means and the status in ability represented by the ability test means.

It is much better to investigate individual growth first, before attempting to depict institutional performance. Then some of the

Table 3

Means and Standard Deviations

Grade	Year	Reading Cohort (N = 529)		Mathematics Cohort (N = 527)	
		Mean	S. D.	Mean	S. D.
		Achievement Scores			
1	1978	320.7	31.9	333.3	25.9
2	1979	383.6	40.1	385.0	33.7
3	1980	426.7	51.0	416.2	32.7
4	1981	470.2	55.9	454.6	43.2
5	1982	505.6	60.6	493.7	48.9
6	1983	536.0	63.8	525.7	54.7
7	1984	565.1	67.9	558.3	63.5
8	1985	593.6	74.1	588.9	68.1
		Ability Scores (Fall 1979)			
Verbal		104.9	14.3	104.7	14.4
Mathematics		104.6	15.2	104.4	15.2
NonVerbal		105.3	14.3	105.3	14.2

above concerns can be addressed in a productive manner. The first step is the choice of a suitable model for individual growth. The model represented by (1) and (2) is employed with the Durham County longitudinal data. The results follow.

<u>Adequacy of the model</u>. The model in (1) was fit to each individual's longitudinal data on Reading Total Scale Score. The same approach was taken with Mathematics Total Scale Score. Examining the individual values of the squared multiple correlation, R^2, is one way of judging the adequacy of fit for each individual. Table 4 shows a seven-number summary of the distributions of R^2 for reading and mathematics.

Note that 95% of the individuals have R^2 values greater than .85 for reading and .86 for mathematics. The median values of R^2 are .95 and .96 for reading and mathematics, respectively! This seems to indicate that straight lines are a remarkably good choice for the individual growth model with this cohort of students.

However, R^2 can be misleading. There can be large values of R^2 when the true growth model is actually nonlinear. Plots of the scores of selected individuals with largest values of R^2 and values of R^2 near the median confirmed the linear trend. Plots of scores for individuals with lowest values of R^2, or with high sums of squared residuals (SSRES), are of course less clear.

In order to quantify departures from linearity, a quadratic component was tested for each individual fit. Only 33 individuals (6.3%) showed a significant ($\alpha = .10$) quadratic component for mathematics. Ninety-three individuals (17.6%) showed a significant

Table 4

Seven-Number Summaries of R^2

#529			#527		
median	.945		median	.961	
25%	.915	.964	25%	.938	.977
5%	.853	.985	5%	.863	.990
extreme	.559	.997	extreme	.451	.996
	Reading Total			Mathematics Total	

($\alpha = .10$) quadratic component for reading. Plots of scores for selected students revealed that curvature may indeed be observed for a few individuals. For other students, the curvature might be due to one stray point or to erratic scores. Usually, these cases were also identified by a large value of SSRES.

While a number of individuals might better be modeled by a quadratic curve for their reading scores, it appears that the straight-line model is appropriate for the majority of individuals for both reading and mathematics.

Figure 1 shows a histogram of the SSRES from each individual straight-line regression. These histograms show that relatively few individuals have large values of SSRES. This is true of both reading and mathematics, and supports the conclusion that the straight-line model is appropriate.

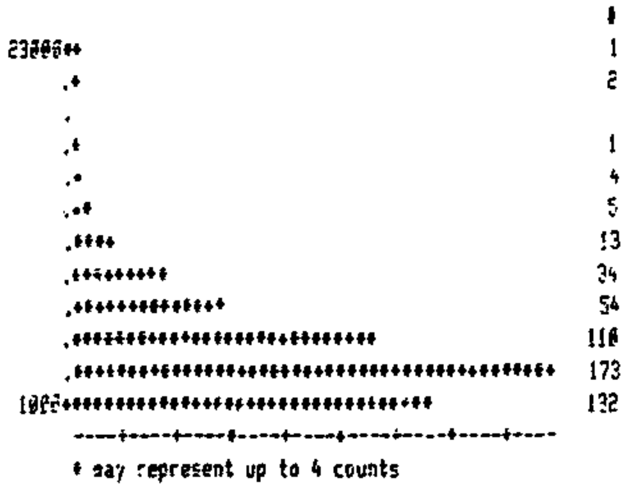
The ordinary least squares estimate of the rate of growth in the straight-line model is robust in the sense that even when the true model is quadratic, the straignt-line slope provides an estimate of the average growth rate of the individual. This, along with the fact that most of the individuals seem to fit the straight-line model extremely well, implies that the straight-line model is suitable for these data and should serve eminently well with these data as an illustration of the growth curve approach to the measurement of change.

Individual growth. The great advantage of fitting individual growth curves to longitudinal data is that growth rates can be determined for each student and the quality of fit can be examined

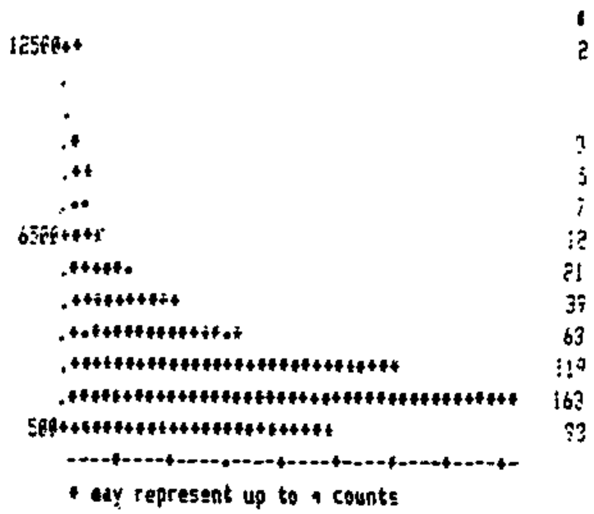
Figure 1

Histograms of Sums of Squared Residuals

Reading



Mathematics



individually. It is also frequently possible to identify aberrations in patterns of growth by examining the scores in conjunction with the model fit for an individual. This can lead to useful hypotheses about the observed behavior. For example, sometimes a poor fit can be due to one peculiar point. This can raise questions about the testing on that administration, or about the validity of the student's performance on that particular day.

Abbreviated output appears in Appendix A showing fits to individual regressions for the first 56 male students in the reading analysis Students are identified by a numeric code. Then values of their ability scores are listed as W_1 through W_3. These correspond to the Cognitive Abilities Test scores on Verbal, Mathematics and Nonverbal, respectively. Then each student's OLS growth rate is listed along with the associated R^2 value. The seventh column contains the individual increments to R^2 (labeled d_rsq) due to the addition of a quadratic component to the straight-line model. The last eight columns show the scale scores from grades one through eight, labeled X_1 through X_8. Though not shown on the printout, SSRES and other diagnostics were computed for each individual.

By examining distributions of the growth rates, R^2, SSRES or other quantities, it is possible to identify individuals with singular characteristics. For example, it is possible to determine that individual 774311 has one of the poorest fits among male students. It appears that the quadratic component is significant. A glance at the scores would suggest that this is largely due to the eighth grade score, which shows a decided drop from the previous scores. Another student, 734810, has one of the fastest growth rates (55.8) among males. The value of R^2 is high for this individual, and a glance at the scores confirms that the estimated growth rate is an accurate indication of fast and steady growth.

Table 5 shows five-number summaries of the growth rates on reading and mathematics for the total group of students and for males and females separately. For the total group, it is apparent that the median growth rates are nearly the same on reading and math, but that there is slightly more variability in mathematics growth, i.e., the extreme growth rates are more extreme for mathematics. This is true among males and among females as well. Females and males show very similar distributions for both reading and mathematics.

Statistical and psychometric characteristics. Because of the nature of the models (1) and (2), it is possible to estimate several important quantities. One key quantity is the variance of the true rates of growth in the population, σ_θ^2. Once this estimate is available, it is possible to estimate the reliability of the estimated rate of change, $\hat{\theta}$, and the standard error of $\hat{\theta}$. These quantities are shown in Table 6. Separate estiamates are provided for reading and mathematics by gender group. (It would be preferable to provide these for the total group, but space limitations with GAUSS on the IBM/PC make this more difficult.) The estimated variances of the true rates of change range from 40.8 to 65.9. The reliability of $\hat{\theta}$ ranges from .731 to .858. (Recall that much traditional literature focused on the low reliability of change scores!) The reliability is

Table 5

Five-Number Summaries of Rate

Total Group

Reading (#529)			Mathematics (#527)		
median	37.3		median	36.2	
25%	32.5	43.1	25%	30.7	41.0
extreme	16.3	59.7	extreme	9.7	64.2

Females

Reading (#278)			Mathematics (#277)		
median	36.8		median	36.4	
25%	32.6	41.5	25%	31.5	41.1
extreme	16.3	59.7	extreme	9.7	64.2

Males

Reading (#251)			Mathematics (#250)		
median	37.7		median	35.6	
25%	32.1	44.4	25%	29.7	41.2
extreme	17.3	59.0	extreme	11.6	62.2

Table 6

Estimates of Statistical and Psychometric Quantities

Quantity Estimated	Reading		Mathematics	
	Females	Males	Females	Males
σ_θ^2	40.812	58.056	46.028	65.883
Reliability of $\hat{\theta}$	.731	.755	.827	.858
Standard Error of $\hat{\theta}$	3.872	4.341	3.099	3.301
Tracking index, γ	.757	.742	.721	.687
Correlation between θ and initial status	.708	.692	.653	.483
Correlation between θ and Verbal ability	.681	646	.686	.614
Correlation between θ and Mathematics ability	.557	.627	.700	.666
Correlation between θ and Nonverbal ability	.560	546	598	.534

somewhat higher for mathematics than for reading, but it is very respectable for both subjects. Standard errors are also provided.

Another interesting quantity listed in Table 6 is an estimate of the tracking index of Foulkes and Davis (1981). Denoted γ, this tracking index reflects the consistency of individual differences in growth across individuals. Alternatively, the index indicates the degree to which individual growth curves are parallel (or "tracking" each other). The index ranges between zero and one, with higher values indicating a greater degree of tracking. Gamma is the empirical probability that two randomly chosen growth curves do not cross in the observed range of time. Values of γ in excess of .5 are taken to indicate tracking.

As can be seen from Table 6, tracking occurs for both males and females on both reading and mathematics. This can be interpreted to mean that there are relatively few crossings of growth curves within each group. An alternative interpretation is that academic ranking within each group is being maintained across time.

The last four rows of Table 6 relate to the investigation of systematic individual differences in growth. Initial status is often taken to be an important correlate of rate of growth. For these data, initial status is taken to be the achievement level at grade one. There are three other candidate predictors as well. They are the three ability scores from the Cognitive Abilities Test--Verbal, Mathematics, and Nonverbal scores. The last four rows of Table 6 show the estimated correlations between rate of change and the four potential predictors of change mentioned above.

The correlation between change and initial status is moderate (.48) for males on the mathematics test. However, for females it is .65 The correlation between change and initial status is higher on the reading test--.69 among males and .71 among females

The correlations between rate of change and the ability test scores range from .534 (the correlation among males between rate of change in mathematics achievement and Nonverbal ability in grade three) and .700 (the correlation among females between the rate of change in mathematics achievement and Mathematics ability in grade three). Thus the ability scores are reasonably good predictors of rate of growth in these data.

Summarizing school growth. By examining the empirical distributions of the estimated rates of change within each school, it is possible to produce an informative picture of the growth in each school in the district. Table 7 shows the mean growth rate in each school and the district. In addition, Table 7 shows the five number summary of growth rates in each school. It is apparent that School 20 has the highest average growth rate on Reading Total Scale Scores and School 16 has the lowest. There are however substantial differences in the number of students in each school. For example, School 16 has only four students in this cohort. Nevertheless, all the schools demonstrate positive average growth.

The five-number summaries are more informative. Scanning the minimum values of the rate of growth, it is apparent that all students are showing progress on both Reading Total and Mathematics Total. The smallest individual growth rate is 9.7 in School 18 for

Table 7

Means and Selected Quantiles of the Empirical Rate Distribution for each School and the District

School	N	Mean	Minimum	25%	Median	75%	Maximum
			Reading				
10	50	36.1	26.0	31.3	34.6	39.9	55.8
11	43	35.5	18.0	30.9	35.8	40.5	50.1
12	27	36.6	19.2	29.3	36.4	40.7	58.8
13	53	37.7	21.4	30.8	35.8	44.2	59.7
14	72	37.6	17.3	31.5	38.6	42.4	56.5
15	31	42.6	32.5	38.2	43.2	45.5	55.7
16	4	23.5	16.3	19.1	24.0	27.9	29.9
17	25	38.6	25.0	33.2	37.0	42.6	56.1
18	70	36.1	20.7	31.2	35.1	39.7	58.4
19	57	36.2	19.3	31.3	36.4	40.1	53.6
20	41	43.3	28.7	37.3	42.8	50.5	57.7
21	56	40.4	20.7	34.9	38.6	46.1	59.0
Total	529	37.9	16.3	32.5	37.3	43.1	59.7
			Mathematics				
10	49	32.5	11.6	27.7	33.6	36.7	56.2
11	43	34.7	18.5	29.6	35.4	39.5	49.6
12	27	33.4	18.8	25.8	35.0	40.2	50.5
13	51	37.5	19.6	31.4	37.0	42.2	56.3
14	74	37.0	12.9	33.3	37.2	42.6	53.8
15	31	41.1	32.2	36.9	39.0	44.1	64.2
16	4	23.2	14.2	18.1	22.2	28.4	34.3
17	26	36.7	25.6	31.0	35.4	42.9	55.8
18	70	33.4	9.7	27.9	33.5	38.9	49.6
19	58	33.8	15.2	28.6	34.2	38.3	54.4
20	40	39.6	22.2	34.5	39.1	44.4	62.2
21	54	39.4	21.3	34.1	37.8	46.6	56.9
Total	527	36.0	9.7	30.7	36.2	41.0	64.2

mathematics. By using the individual fits, this individual could be identified and perhaps hypotheses could be offered about the relatively low performance. On the opposite end of the spectrum, the highest individual growth rate is 64.2 in School 15, and again, it is on mathematics.

By scanning the five-number summaries for each school, it is apparent (as it was from the mean growth rates) that there is variability across schools in the central tendency of the distributions of growth or learning. In addition, there is some variability in the range of growth within a school. One school shows a range of more than forty scale score points (i.e., School 18, mathematics) while another has a range of less than 30 scale score points (i.e., School 20, reading). Still, all schools seem to be characterized by consistently strong growth patterns. This is of course reflected in the District summaries for reading and mathematics.

Effects of aggregation. When polynomial growth curves are used, the collection of individual growth curves has the property that the mean of the individual growth rates is equal to the growth rate of the mean scores. This property has been called *dynamic consistency*. However, this property does not necessarily characterize other types of growth curves, nor does it necessarily apply when other types of summary statistics are applied (e.g., the median, or other quantiles). The purpose of this section is to empirically investigate the effects of using the quantiles of the score distributions at each occasion (rather than the quantiles of the distribution of individual growth rates) as summaries of longitudinal performance.

Table 8 shows the growth rates determined from fitting a straight-line model to various summary quantities. Six different summary statistics are used. They are the mean, the minimum, the 25th quantile, the median, the 75th quantile, and the maximum. These growth rates should be compared to the growth rates in Table 7 to determine how closely growth rates based on summary statistics reflect the actual empirical distribution of growth.

Note first of all that the column of means in Table 8 is exactly the same as the column of means in Table 7. This is true for both reading and mathematics. This reflects the dynamic consistency of polynomial growth curves mentioned above.

Next note the column of "median" growth rates in Table 8 and compare this with the actual median growth rates in Table 7. The two sets of "medians" are not equal. The values for each school in Table 8 are sometimes the same as the corresponding values in Table 7; but more often, they are different. There is no discernible pattern. Table 8 gives higher values for some schools and lower values for other schools.

Comparing the "minimum" and "25th quantile" values in Table 8 with those in Table 7, it is notable that the values in Table 8 are most often larger than those in Table 7. Generally it appears that fitting growth curves to the minimum values of the scale scores (or the 25th quantiles) yields an overestimate of the minimum (or 25th quantile) of the empirical distributions of the growth rates. This is true regardless of whether the level of "aggregation" is at the school or district level.

Table 8

Rates Estimated by Fitting a Straight-Line Growth Model to the Means or Various Quantiles Within Each School and the District

School	Mean	Minimum	25%	Median	75%	Maximum
			Reading			
10	36.1	30.3	34.2	36.0	37.2	52.2
11	35.5	21.0	31.2	35.4	39.6	51.7
12	36.6	23.2	28.9	36.5	41.3	52.8
13	37.7	19.1	32.5	37.2	40.9	55.2
14	37.6	19.5	31.9	38.7	41.5	52.6
15	42.6	36.2	40.3	40.9	43.9	52.2
16	23.5	16.3	19.3	23.2	27.7	31.3
17	38.6	30.9	33.5	38.6	41.7	55.3
18	36.1	30.0	33.2	36.4	38.9	50.4
19	36.2	25.4	36.3	35.6	35.8	52.5
20	43.3	31.5	39.1	42.1	47.4	55.7
21	40.4	29.6	37.5	41.4	41.7	53.7
District	37.9	23.3	34.2	37.6	39.8	58.1
			Mathematics			
10	32.5	16.4	30.0	34.2	34.6	50.8
11	34.7	25.9	31.2	35.5	37.7	48.2
12	33.4	21.0	25.4	35.0	39.4	50.0
13	37.5	28.6	33.0	37.0	41.2	53.6
14	37.0	19.3	35.8	38.2	41.6	47.6
15	41.1	35.6	37.2	40.8	43.3	51.9
16	23.2	20.1	20.3	21.1	26.1	30.6
17	36.7	27.8	32.3	34.7	41.2	56.5
18	33.4	19.5	31.4	34.2	36.4	47.3
19	33.8	17.6	29.3	34.6	38.1	52.5
20	39.6	24.6	37.5	39.7	41.9	53.9
21	39.4	22.1	35.1	38.2	43.3	51.3
District	36.0	16.1	32.6	36.5	39.8	56.5

Finally, compare the "75^{th} quantile" and "maximum" columns in Table 8 with the corresponding columns in Table 7. Again, there are a few exceptions, but in general, the numbers in Table 8 are smaller than those in Table 7. Determining growth from the upper quantiles of the school or district scale score distributions appears to give an underestimate of the empirical growth rates at corresponding points in the school or district distributions of growth.

Concluding Remarks

Recent literature in the measurement of change focuses on the use of growth curves for modeling individual growth. This approach has numerous advantages over previous methods. Foremost, it focuses attention on the use of multi-wave data for determining characteristics of growth among individuals. There are statistical rewards for using multiple occasions of measurement. The rewards include improved estimation, increased precision and better reliability Just as important, however, the use of an individual growth model focuses attention on individual students and makes them the first level of analysis. Understanding institutional growth is secondary to understanding individual growth.

The results presented in this report show that longitudinal data analyses are feasible for school districts when a suitable longitudinal data base has been maintained. The payoff for maintaining such data bases is a wealth of information about individual growth that has not been available to school personnel before This includes estimates of individual growth; assessments of the quality of the growth model for each individual, diagnostic information that may help teachers better understand the performance of each student; and, information about correlates of learning. In addition, new ways of describing institutional performance over time become available. These include estimates of "typical" growth and descriptions of the distribution of growth in an institution.

Failing to focus first on individual growth has its costs. Examining the *mean* level of performance in schools will give an accurate assessment of the average growth rate when polynomial growth models are used. However, such analyses give very limited information about growth in the schools, and their accuracy is not guaranteed when other (non-polynomial) growth models are used. Finally, attempting to determine institutional growth characteristics from quantiles of the scale score distributions can be misleading and seems to introduce systematic error at the extremes of the score distributions.

APPENDIX A

Reading Scale Score--Males

Fits to individual regressions

id	w1	w2	w3	rate	rsq	d_rsq	x1	x2	x3	x4	x5	x6	x7	x8
314815	114	101	110	32.5	97.4	0.5	368	376	445	465	513	521	[illegible]	593
704810	98	96	102	31.3	96.7	1.3	306	339	373	438	456	464	496	520
713810	78	82	105	52.3	98.0	0.2	312	331	378	448	496	535	626	647
715310	100	108	123	40.7	93.6	2.0	316	435	418	483	531	564	612	[illegible]
715810	96	97	100	30.8	95.1	0.0	355	369	398	412	499	508	518	558
718910	113	108	109	32.7	88.8	6.2 *	344	435	465	500	548	530	606	579
722910	114	102	92	34.3	93.6	4.0	332	356	454	498	508	539	537	595
726310	93	90	89	36.3	98.2	0.4	324	379	392	448	476	535	548	574
727310	106	106	106	42.8	98.7	0.0	328	385	423	451	504	571	600	623
727810	109	118	108	37.7	97.5	0.5	341	365	413	451	489	557	569	581
734810	106	113	123	55.8	96.8	2.0	335	441	492	558	639	645	687	751
735810	114	130	135	35.7	94.3	2.1	348	414	478	475	518	557	612	595
736310	109	98	104	45.3	98.5	0.8	303	340	413	469	504	540	559	622
737810	102	97	102	26.1	85.8	13.2 *	291	352	398	456	456	489	493	468
738310	106	109	99	36.3	94.6	2.3	326	396	423	488	508	545	555	599
740310	91	91	110	34.1	97.8	0.1	289	343	347	404	439	459	517	521
740810	101	94	102	37.7	94.8	2.9	296	385	437	466	489	535	542	591
741810	95	97	105	27.7	91.5	4.3	335	352	423	441	459	508	490	525
748310	96	88	81	32.5	92.3	1.4	303	349	383	440	476	485	409	558
755810	112	106	97	39.3	95.8	1.2	372	363	409	443	476	564	577	619
756311	85	94	73	28.4	78.8	10.1 *	303	400	437	414	479	540	528	495
757311	122	125	126	34.7	84.3	5.3 *	338	414	509	563	504	540	606	645
757811	117	109	125	27.1	91.4	0.2	364	448	478	469	499	512	559	591
760811	97	87	84	30.9	94.3	1.9	315	382	391	438	489	496	500	545
761811	126	111	114	45.8	91.4	2.5	359	367	445	559	539	633	612	656
763311	108	119	103	33.8	95.9	2.3	372	400	454	475	489	549	563	583
764311	107	124	119	42.8	97.8	0.1	364	374	423	472	524	557	590	648
767811	95	83	89	24.2	90.1	4.0	296	359	395	445	475	444	493	525
769311	98	103	93	37.4	95.3	2.6	313	389	429	461	446	571	531	583
771311	131	115	131	43.5	91.5	0.6	341	429	492	483	539	633	561	686
771811	117	119	102	36.7	98.8	2.2	335	414	478	451	525	551	617	587
773811	114	100	107	37.3	73.9	21.0 *	335	409	509	518	639	508	600	597
774311	94	94	99	20.8	73.7	5.7 *	316	361	395	399	412	450	515	430
774811	130	116	118	37.7	91.0	1.4	372	400	509	518	539	545	647	527
776811	96	87	100	25.4	91.5	0.2	241	336	381	464	432	460	517	[illegible]
778311	120	130	123	50.1	92.4	0.1	355	455	492	456	540	666	572	714
780011	101	102	97	20.8	87.4	10.0 *	341	369	429	451	462	425	430	496
781811	131	139	135	44.9	95.5	2.8	355	455	476	553	600	629	659	692
783811	116	127	122	43.3	92.1	5.0 *	340	385	492	518	600	587	619	645
785811	166	117	98	33.0	96.4	0.2	355	405	405	446	512	521	555	613
787311	111	121	105	31.8	91.7	2.0	372	392	492	463	531	557	592	595
788812	74	75	69	19.2	82.3	4.4	313	299	347	378	404	464	450	467
789312	99	117	117	40.5	98.1	0.7	305	376	336	458	492	535	555	595
793812	91	82	102	30.0	92.5	5.8 *	254	336	375	401	426	447	469	494
796812	102	104	113	33.8	85.8	1.0	396	361	409	451	485	583	452	595
797312	110	98	91	40.7	96.6	0.8	365	363	418	427	469	526	534	574
798312	106	109	103	37.6	98.2	4.7 *	334	405	409	529	513	571	577	591
798812	90	94	103	26.9	90.5	0.0	287	325	355	401	373	450	-27	500
799812	96	83	92	24.8	93.5	0.7	300	350	362	414	441	431	504	493
802812	81	87	91	25.0	96.2	0.4	293	322	324	378	404	435	431	464
804812	86	84	100	29.3	95.7	1.7	313	356	379	430	459	460	518	509
811312	116	112	106	52.4	91.6	2.5	326	403	454	577	639	600	672	731
832813	89	107	107	35.2	80.9	14.0 *	289	425	445	529	529	543	551	577
839313	111	107	109	30.4	95.0	2.8	338	374	445	459	539	564	577	591
839813	99	95	97	34.2	88.9	7.2 *	291	324	369	478	451	499	528	525
840312	109	110	110	44.2	89.7	9.4 *	319	395	455	543	594	557	619	[illegible]
844813	105	110	105	52.0	96.3	0.1	322	409	392	435	571	609	612	[illegible]
845313	71	57	88	21.4	55.9	1.9	313	326	[illegible]	[illegible]	[illegible]	479	[illegible]	496

REFERENCES

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

Blomqvist, N. (1977). On the relation between change and initial value. Journal of the American Statistical Association, 72(360), 746-749

Bryk, A. S. (1987). A paradigm for assessing the effects of educational programs on student growth. Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C.

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. Psychological Bulletin, 101(1), 147-158.

Cronbach, L. J., & Furby, L. (1970). How we should measure "change"--or should we? Psychological Bulletin, 74, 68-80.

Lord, F. M. (1956). The measurement of growth. Educational and Psychological Measurement, 16, 421-437.

Lord, F. M. (1963). Elementary models for measuring change. In C. W. Harris (Ed.), Problems in measuring change (pp. 21-38). Madison: The University of Wisconsin Press.

Nunnally, J. C. (1973). Research strategies and measurement methods for investigating human development. In J. R. Nesselroade & H. W. Reese (Eds.), Life-span developmental psychology: Methodological issues (pp. 87-109). New York: Academic Press.

O'Connor, E. F., Jr. (1972). Extending classical test theory to the measurement of change. Review of Educational Research, 42, 73-97.

Raudenbush, S. (1987). Applications of hierarchical linear models to the study of educational programs: A review. Paper presented at the annual meeting of the American Educational Research Association, Washington, D. C.

Rogosa, D. R., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. Psychological Bulletin, 92, 726-748.

Rogosa, D. R., Floden, R., & Willett, J. B. (1984). Assessing the stability of teacher behavior. Journal of Educational Psychology, 76, 1000-1027.

Rogosa, D. R., & Willett, J. B. (1983a). Comparing two indices of tracking [Letter to the editor]. Biometrics, 39, 795-796

Rogosa, D. R., & Willett, J. B. (1983b). Demonstrating the reliability of the difference score in the measurement of change. Journal of Educational Measurement, 20, 335-343.

Rogosa, D. R., & Willett, J B. (1985). Understanding correlates of change by modeling individual differences in growth Psychometrika, 50, 203-228.

Rogosa, D. R., Willett, J. B., & Williamson, G. L. (1986). Tools for the study of school learning: Implementation of statistical methods for the analysis of student growth in achievement. Unpublished manuscript.

Thorndike, R. L. (1966). Intellectual status and intellectual growth. Journal of Educational Psychology, 57, 121-127.

Willett, J. B. (1985). Investigating systematic individual differences in academic growth. Unpublished doctoral dissertation, Stanford University.

Williamson, G. L. (1986). Assessing Academic Growth on Multiple Measures. Unpublished doctoral dissertation, Stanford University.