Using the Longitudinal Guttman Simplex as a Basis for Measuring Growth

Linda M. Collins and Norman Cliff
J. P. Guilford Laboratory of Quantitative Psychology, University of Southern California

Many difficulties inherent in the measurement of growth stem from the use of traditional measurement methodologies. The longitudinal Guttman simplex (LGS), an alternative approach based on a model of growth, is discussed in this article. The LGS has several advantages over traditional methodology. First, interindividual differences in developmental rates are a part of the model. Second, the LGS procedure can easily handle any number of occasions of measurement. Third, the LGS is suited to nonlinear as well as linear monotonic growth. Fourth, a consistency index associated with the LGS methodology, CL, indicates the extent to which cumulative, unitary development characterizes a particular latent variable. Finally, and perhaps most important, because a model of the growth undergone by the latent variable being measured is incorporated in the LGS model the resulting instruments enjoy a high level of construct validity. The LGS is limited to cumulative, unitary development; additional measurement theories are needed for other kinds of development.

Although the ultimate goal of measuring growth is to infer functional curves for individuals, in practice this is a difficult goal to attain. Much of the difficulty stems from the methodology in current use for developing instruments to measure growth in a latent variable. For the most part traditional measurement theory, including classical test theory and item response theory, is based on the idea of an unchanging true score and thus is poorly suited to development of measures of growth. Yet psychological research on growth generally relies on traditional psychometric methodology. One purpose of this article is to demonstrate how traditional measurement theory is lacking in relevance to the concerns of researchers who wish to measure monotonic individual growth, and that the requirements for developing and selecting items to make up an instrument measuring monotonic individual growth of a latent variable are different from those commonly used for single-occasion instruments.

An alternative to traditional methodology is needed, one offering criteria for development of instruments that are more relevant to the measurement of growth. Such an alternative is suggested in part by the work of those researchers—such as Rogosa, Brandt, and Zimowski (1982), Rogosa and Willett (1985), and Bryk and Raudenbush (1987)—who are advancing a new perspective on the long-standing debate about the statistical treatment of individual growth data. A cornerstone of this new perspective is the philosophy that an explicit a priori model of individual growth is an indispensable part of the study of growth (e.g., Rogosa et al., 1982). One important implication of this point of view is that the key to measurement of individual growth lies in using an explicit model of individual growth to guide measurement theory.

No single approach to measurement could be general enough to embrace all the different models of growth and change appearing in the psychological literature. However, cutting across many functional curve models is one simple, important, and reasonably general aspect of growth that serves as the focus of this article. Regardless of its exact form, the change featured by most growth models is *monotonic*. Thus, a fundamental characteristic of scales that purport to measure growth is that they change monotonically across time *within individuals*. A second purpose of this article is to describe a new measurement theory based on a model of monotonic individual growth. This alternative measurement theory is called the longitudinal Guttman simplex.

Problems With Traditional Approaches to Measurement of Monotonic Growth

The problems with traditional approaches to measurement of monotonic individual growth arise in two areas. First, traditional approaches are not based on a model of individual growth in the latent variable, limiting the construct validity of the resulting measures. Second, traditional approaches place a heavy emphasis on within-time interindividual variability in latent variables to the exclusion of across-time intraindividual variability. Yet the latter is the essence of individual growth, whereas growth in a population can and does occur in the absence of within-time interindividual differences. The purpose of this section is to show how, for these reasons, traditional psychometric methodology is inadequate for and largely irrelevant to development of scales to measure monotonic growth.

This research was supported by National Institute on Drug Abuse grants DA3673 and DA04111 and National Science Foundation Grant BNS8403126.

Correspondence concerning this article should be addressed to Linda M. Collins, J. P. Guilford Laboratory of Quantitative Psychology, Department of Psychology, University of Southern California, University Park, Los Angeles, California 90089-1061.

Psychological Bulletin, 1990, Vol. 108, No. 1, 128–134

As a starting point, consider one commonly used and, on the surface, reasonable approach to evaluating an instrument to measure growth. The approach is to define an adequate measure of monotonic individual growth as an instrument that shows satisfactory reliability at each occasion and exhibits monotonic change in mean total score across occasions. Because this procedure lacks a model of individual growth, it does not distinguish between a monotonic increase in mean total score based on instances of monotonic growth exclusively and one based on a mixture of monotonic growth and decided departures from monotonicity. However, as Rogosa et al. (1982) pointed out, if the latent variable being measured is hypothesized to undergo monotonic growth, then only an instrument reflecting monotonic growth at the individual level ensures adequate construct validity.

Items With Zero Within-Time Interindividual Variability

The misplaced emphasis on within-time variability that occurs when traditional measurement procedures are applied to measures of monotonic growth can result in the inappropriate removal of items from a scale. This occurs when items that have a within-time interindividual variability of zero are branded "constants" and routinely removed, irrespective of whether they show across-time intraindividual variability. There are several reasons why researchers routinely discard items that have a variance of zero at one or more occasions. One reason is that because the correlation of a constant item with any other item is undefined, it does not make sense to include such items in any statistical procedures based on correlations, such as regression, factor analysis, or item-total correlations. Another reason is that if constant items are included in an instrument development procedure, they typically reduce measures of within-time scale consistency to some degree. For example, adding k constant items to a measure originally n items long reduces Cronbach's alpha or Kuder-Richardson 20 by a factor of $k/(n - 1) \times (n + k - 1)$.

However, any item that subjects are failing on one occasion and then passing on another occasion is reflecting growth over time irrespective of the amount of interindividual within-time variability it contributes. For example, suppose childhood growth occurring in a certain sequence is being measured—say, a very simple math skills sequence involving first addition skills, then subtraction, then multiplication, then division. Clearly, a test measuring development along this dimension would have to include items from all four domains in order to claim adequate content validity. Suppose that on one occasion when the test is administered most of the children can add, a substantial number can subtract, and a few can multiply, but none can divide. Thus, the division items are constants because no one has passed them. When the test is administered again a year later the division items are no longer constants, but the addition items are constants because everybody has passed them. To see the effect that the practice of routinely discarding such items can have on content validity, consider what the measure discussed in this example would be like with the addition item, the division item, or both removed. The very limited usefulness of traditional methods in this context is illustrated by their inability to assess whether the constant items belong in this measure of monotonic growth. The major measures of intellectual development, such as the Stanford-Binet Intelligence Scale, the Wechsler Intelligence Scale for Children, and the Bayley Scales of Infant Development, escape this criticism. They include items that are passed or failed by all children at a given age. This is one of the enduring strengths of these measures.

Reliability

Few researchers would dispute that under most circumstances scales should be constructed so as to maximize reliability, which is an indication of "imprecision and precision of tests" (Lord & Novick, 1968, p. 61). However, because the traditional definition of reliability is based solely on within-time interindividual variability, it has little to do with the precision of a measure of monotonic growth (Rogosa et al., 1982). The traditional definition of instrument reliability as the ratio of within-time interindividual true score variance to total within-time interindividual variance,

$$\rho_{XX} = \frac{\sigma_T^2}{\sigma_X^2} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_E^2}, \tag{1}$$

has little relevance to growth. Equation 1 shows that given the same nonzero error variance, a population with a large within-time interindividual true score variance will yield a higher reliability than a population with a small within-time interindividual true score variance. This means that where the subject population is relatively homogeneous with respect to the latent variable being measured, such as in a study following an ability through the school years in a cohort of children, a measure cannot show large reliability even if it is reflecting growth with perfect accuracy.

One suggestion for avoiding this problem might be to base instrument development on samples that are heterogeneous with respect to the latent variable being measured (assuming that such variability exists in the population) in order to ensure an amount of within-time interindividual true score variability that is sufficient to allow the instrument to show good reliability. Another suggestion (Rogosa et al., 1982; Saupe, 1966) involves identifying a parameter that represents the individual's true rate of linear growth over time and a statistic that estimates this parameter. Reliability is then redefined as the proportion of interindividual variability in the statistic attributable to interindividual variability in the parameter. However, both of these strategies lead to the evaluation of instruments in terms of criteria that are not directly relevant to the measurement of individual growth. Although in a particular growth situation there may be interindividual differences in either the latent variable or the rate at which the latent variable is changing, it is quite possible to have neither and still have across-time intraindividual variability (e.g., growth).

It is worth noting that although high test–retest reliability is often considered desirable in measures of behavioral development (e.g., Rushton, Brainerd, & Pressley, 1983), using this estimate of reliability does not avoid the problems we have listed. An instrument measuring growth may not produce a substantial between-occasions correlation even if the measure is a very accurate one. Given the same degree of measurement precision, a population whose members all develop at the same rate will produce larger test–retest correlations than a population whose members develop at different rates. Thus, except in the rare instances where there is complete between-subjects homogeneity of developmental rate, the test–retest correlation is not a good reflection of the precision with which a measure reflects growth.

One Approach to Measurement of Monotonic Growth: The Longitudinal Guttman Simplex

The preceding discussion demonstrated that a new approach to measurement of growth is needed, and that this new approach must have several characteristics. First, it must take as a starting point a model of the growth process. Second, criteria for evaluating instruments must be based on intraindividual differences (i.e., individual growth) rather than on interindividual differences. Third, because individual differences in growth rate are to be expected in almost every situation, the presence of interindividual or even intraindividual differences in growth rate should not lead to the conclusion that the scale in question is inconsistent. This section offers an alternative that has these characteristics. Discussed in this section is the longitudinal Guttman simplex (LGS; Collins & Cliff, 1985; Collins, Cliff, & Dent, 1988), the first measurement model especially for development of instruments for longitudinal measurement of monotonic change.

Cumulative, Unitary Development and the LGS Model

As discussed earlier, the first step in developing a measure of growth is an explicit theory about the nature of individual growth in a particular research situation. The theory of individual growth forming the basis of the LGS is called *cumulative, unitary* growth. By *cumulative*, it is meant that as new skills are acquired, previously acquired skills are retained. By *unitary*, it is meant that all subjects develop in the same direction. In other words, all skills are acquired in the same order, and once acquired, a skill is not lost.

The LGS is an extension of the familiar Guttman simplex model (Guttman, 1950). There are two important differences between the Guttman simplex and the LGS model. First, whereas the Guttman simplex is a two-set model, where the sets are persons and items, the LGS model incorporates a third set, time. Second, the Guttman simplex is a joint order of persons and items; that is, an ordering of the data matrix so that the items are ordered by difficulty, the persons are ordered by ability, and persons and items order each other. In contrast, the LGS model is a joint items–times order consistent across persons. The joint items–times order consistent across persons is a key feature of the LGS. It means that for each person there is an internally consistent Guttman scale of items and times, and that every person's Guttman scale provides the same ordering for items and times.

The LGS is an operationalization of cumulative, unitary development. To see how this is so consider Table 1, which shows a hypothetical example of data on a math skills acquisition measure. In the hypothetical example, there is a measure with four items, one for each of four skills, taken in Grades 1, 2, and 3 by Child A and Child B. Growth in math skill is cumulative here because as more skills are acquired, previously acquired skills are retained. For example, Child A, who has learned multiplication, still passes the subtraction and addition items. Growth is also unitary, because both Child A and Child B are acquiring skills in the same order. Thus, it is clear that Table 1 is an example of cumulative, unitary growth. Another way of looking at Table 1, however, is as a perfect LGS, with a joint items–times order consistent across the two children. For each of the two children, there is a joint order of items and times. This joint order is consistent across the two children, ordering the times such that math skill is clearly increasing over time and ordering the items such that they are passed in the following order: first addition, then subtraction, then multiplication, then division.

Table 1
Hypothetical Example of an LGS

Child	Addition	Subtraction	Multiplication	Division
Child A				
Grade 1	Pass	Fail	Fail	Fail
Grade 2	Pass	Pass	Pass	Fail
Grade 3	Pass	Pass	Pass	Fail
Child B				
Grade 1	Fail	Fail	Fail	Fail
Grade 2	Pass	Pass	Fail	Fail
Grade 3	Pass	Pass	Pass	Pass

Note. LGS = longitudinal Guttman simplex.

The LGS Model and Scale Development

In any instrument development work, it is important to know the degree of internal scale consistency. This is the purpose of such familiar consistency indices as Cronbach's alpha and the coefficient of scalability. Because the LGS is a logic-theoretic model with an axiomatic definition (Collins & Cliff, 1985), it is natural to evaluate consistency in terms of the LGS model axioms, that is, to record adherences to and departures from the model axioms in empirical data. This is the approach taken by CL, a consistency index for longitudinal Guttman scales developed by Collins et al. (1988).

CL provides a direct measure of the consistency of empirical data with the axioms of the LGS model by counting order relations (Cliff, 1977). The starting point for this is the 2×2 response pattern made up of an individual's responses to two items at two times. These response patterns contain order relations; that is, they provide an order for the items, for the times, or for both items and times. Those order relations that are consistent with cumulative, unitary development represent instances of adherence to the LGS model axioms in the data; any other order relations represent violations of the LGS model axioms.

The CL index is defined as follows:

$$CL = \frac{\dfrac{\text{Consistent}}{\text{Total}} - \dfrac{\text{Expected consistent}}{\text{Expected total}}}{1 - \dfrac{\text{Expected consistent}}{\text{Expected total}}}. \qquad (2)$$

In Equation 2, *consistent* refers to a weighted total of order relations in the data that are consistent with the LGS model. The weighted total includes the number of item-difficulty order relations that are in agreement with the model across times plus the time-progress order relations that are consistent across items. Added to these is the number of consistent "unique" relations; that is, the number of instances where a pair of items are unordered at one time (both are passed or both are failed) but are ordered in a way consistent with the model at the other time. *Total* refers to a weighted total number of item, time, or unique relations, whether consistent or inconsistent with the model. *Expected consistent* and *expected total* refer to the weighted total of relations of each kind that would be expected simply on the basis of the differences in marginal frequencies between items and across times.

Weights are applied differentially to the item orders, time orders, and unique orders so as to achieve a high degree of sensitivity to small differences in scale consistency. Note that CL incorporates an adjustment for consistency that would be expected due to chance, that is, the apparent consistency in Guttman scales that is attributable to differences in item marginals. In the case of the LGS, the adjustment is made both within times and across times. Thus, CL is an adjusted, weighted proportion of the total number of order relations consistent with the model axioms. CL can be used to indicate overall scale consistency, and CL-if-item-deleted can be used to help indicate items whose removal would improve scale consistency. (For additional details on the development of CL, see Collins et al., 1988.)

The metric of CL differs from more familiar consistency indices in that it is often lower than might be expected. Part of the reason for this is that CL is focused at the item level, and thus is analogous to an average interitem correlation. A similar effect is found in other item-level indices of cross-sectional consistency such as the one proposed by Loevinger (1947) and later by Mokken (1971) and Cliff (1977). We are currently working on consistency indices for measures of monotonic change that will apply at the total score level.

Artificial data illustration. To illustrate the use of the LGS methodology and to contrast it with traditional methods, we now analyze two artificial data sets. Suppose a researcher wishing to develop a measure of cumulative, unitary monotonic growth pilot tests a six-item instrument involving 20 subjects at two occasions. The resulting data appear in Table 2. Traditional methods seem to indicate that these six items make up a consistent measure of growth. First, all six items are sensitive to growth, as reflected in the increase in passing probabilities shown in Table 3. Second, Table 2 shows that the items and persons form a perfect Guttman scale at each occasion. Third, the difficulty order of the items is identical across times, suggesting that they form a growth sequence.

In fact, these data were generated so that five of the items form a perfectly consistent LGS and the remaining item does not belong in the scale. Table 3 shows that the scale CL improves dramatically when Item 2 is removed. This may seem paradoxical when traditional methods indicate that this item is a part of a perfect Guttman scale. However, although this item is part of a perfect cross-sectional Guttman scale, it is not part of a cumulative, unitary developmental sequence. This can be seen by examining the data in Table 2. The footnoted response patterns are those that are inconsistent with the LGS model axioms, because although the rest of the items reflect monotonic growth, Item 2 reflects nonmonotonicity. That is, if Item 2 is a part of the scale, this indicates that some individuals engage in nonmonotonic development by gaining Item 1 and Item 2, and then losing Item 2.

Table 2
Hypothetical Data Set 1

Occasion 1 items						Occasion 2 items					
1	2	3	4	5	6	1	2	3	4	5	6
0	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	1	1	1	0	0	0
0	0	0	0	0	0	1	1	1	1	0	0
0	0	0	0	0	0	1	1	0	0	0	0
0	0	0	0	0	0	1	1	1	1	1	0
0	0	0	0	0	0	1	1	1	1	1	1
1	0	0	0	0	0	1	1	0	0	0	0
1	0	0	0	0	0	1	1	0	0	0	0
1	0	0	0	0	0	1	1	1	1	0	0
1	1	0	0	0	0	1	0	0	0	0	0[a]
1	1	0	0	0	0	1	0	0	0	0	0[a]
1	1	0	0	0	0	1	0	0	0	0	0[a]
1	1	0	0	0	0	1	0	0	0	0	0[a]
1	1	1	0	0	0	1	1	1	0	0	0
1	1	1	0	0	0	1	1	1	1	0	0
1	1	1	0	0	0	1	1	1	1	1	0
1	1	1	0	0	0	1	1	1	1	1	1
1	1	1	1	0	0	1	1	1	1	1	0
1	1	1	1	0	0	1	1	1	1	1	1
1	1	1	1	1	0	1	1	1	1	1	1

[a] Inconsistent with a longitudinal Guttman simplex model.

There are two other aspects of this analysis that are worth noting. First, as Table 3 shows, at the first occasion Item 6 is a constant because no one has passed it; at the second occasion Item 1 is a constant because no one has failed it. These items both reflect growth over time and are perfectly consistent with the other items in the scale, although, as discussed earlier, many psychometric methodologies would routinely discard them. Second, although the removal of Item 2 produces a perfect LGS, the correlation between scale scores on the two occasions is only .49, reflecting the fact that different individuals grow at different rates. A researcher using test–retest reliability as a criterion would probably scrap this measure, even though it is an excellent measure of cumulative, unitary growth. Although individual differences in growth rate make a perfect test–retest correlation impossible, Table 3 shows that such individual differences do not affect CL.

Now suppose another researcher interested in developing a measure of growth collects data on a sample of individuals who are all of the same ability and who develop at the same rate. At the first occasion, all subjects give the response pattern (00000); at the second, all subjects respond (10000); at the third, (11000); at the fourth, (11100); at the fifth, (11110); and at the last, (11111). Unlike the data in Table 2, these data do not form an items–persons Guttman scale at each occasion. In fact, at any single occasion all of the items are constant. Thus, there is no within-time scale, and examination of the data panel by panel will not shed any light on sequencing of the items. Yet there is

Table 3
Item Passing Probabilities and CL if Item Was Deleted for Artificial Data

Item	Time 1	Time 2	CL if item deleted
1	0.70	1.00	0.47
2	0.55	0.75	1.00
3	0.35	0.60	0.38
4	0.15	0.50	0.38
5	0.05	0.35	0.38
6	0.00	0.20	0.40

Note. CL = a consistency index for longitudinal Guttman scales (Collins, Cliff, & Dent, 1988).

a great deal of individual growth, that is, a great many intraindividual across-time differences. The data show quite clearly that there is an items–times joint order, in other words, that individuals pass the items in an ordered sequence and that this sequence is the same for all individuals. The CL for these data is unity, indicating that the five items form a perfectly consistent measure of cumulative, unitary development.

Empirical data illustration. The purpose of this illustration is to show that an instrument may appear to be a measure of cumulative, unitary monotonic growth by traditional criteria and yet fail to be consistent with the LGS model. The data presented here were originally analyzed by Rock and Pollack-Ohls (1987), who took a sample of 1,500 individuals from the High School and Beyond 1980 sophomore cohort data file. Each student had completed a 38-item math test in 10th grade and again in 12th grade. Rock and Pollack-Ohls's goal was to develop a measure based on some subset of the 38 items that would tap several levels of math skills development, where "ideally the higher levels would require all the skills at the lower levels plus some new skill unique to that level" (p. 4).

Rock and Pollack-Ohls (1987) began by constructing four 5-item "testlets," each of which reflected mastery of one of the following mathematics operations: (a) single operations on whole numbers; (b) powers and roots, decimals, and fractions; (c) low-level algebra; and (d) low-level geometry and moderate-level algebra. The testlets were then used as items. A mastery criterion was determined, with a subject demonstrating mastery given a *1* and a subject failing to demonstrate mastery given a *0*. The purpose of constructing testlets was to improve reliability at the item level, and this novel strategy appears to have worked quite well.

At first glance, these data appear to form a highly consistent LGS. Table 4 shows that within each time the difficulty order of the testlets is as would be expected, that is, with Testlet 1 the least difficult and Testlet 4 the most difficult. In fact, Rock and Pollack-Ohls (1987) reported that the testlets form a highly consistent Guttman scale within each occasion, with approximately 90% of respondents contributing perfectly consistent response patterns at each occasion. Table 4 also shows that each of the testlets is sensitive to change over time, reflected by an increase, although relatively modest, in passing probability between 10th and 12th grade. However, despite this apparent consistency the CL for these data is only .19. CL-if-item-deleted shows that this low CL is not caused by one problem item alone. CL increases to only .21 if Testlet 1 is deleted and to .20 if Testlet 2 is deleted, and decreases if Testlet 3 or Testlet 4 is deleted. Thus, the conclusion is that these four testlets do not form a consistent LGS.

Although CL is not based directly on response patterns, it is instructive to examine the response patterns to see why data that are apparently so consistent do not form an LGS. Table 5, which contains a cross-tabulation of the Time 1 and Time 2 response patterns in these data, shows clearly why the Rock and Pollack-Ohls (1987) data do not form a consistent LGS. The LGS model makes two implications for an individual's response pattern. The first implication is that each response pattern will be consistent with a perfect Guttman scale *at each time*. In this case, there are five such possible within-time response patterns. If the testlets are ordered numerically, the only response patterns should be (0000), (1000), (1100), (1110), and (1111). Table 5 shows that about 20% of the subjects in the Rock and Pollack-Ohls data provide non-Guttman response patterns at one or both occasions. The second implication made by the LGS model is that *across time* an individual's response pattern will show monotonic growth consistent with the direction of growth exhibited by all other individuals in the population. The diagonal of Table 5 shows that about 40% of the subjects did not undergo any change at all between the first and second occasion. The remaining subjects, who are in the upper and lower triangles of Table 5 (excluding the non-Guttman row and column), are subjects with consistent response patterns who exhibit change. Because in the Rock and Pollack-Ohls data growth is taking place in a positive direction overall, as Table 4 shows, if these data were perfectly consistent with the LGS model the lower triangle of Table 5 would be empty. In contrast, about two thirds of these subjects are above the diagonal, showing growth in ability between 10th and 12th grade, and about one third are below the diagonal, showing decline. Thus, these data do not form a consistent LGS because there is relatively little change over time, and what change does occur is not monotonic. Overall, only about one quarter of the subjects contributed responses completely consistent with the idea of cumulative, unitary development.

The conclusion from this analysis is that, although the skills form Guttman scales within times and show moderate growth across times, it is only moderately true that these math skills exhibit monotonic growth. Another way to describe it is that these four math skills do not constitute an LGS to more than a modest degree.

Discussion

The measurement procedure outlined here fits in well with the procedures for analysis of growth described by Rogosa et al.

Table 4
Item Passing Probabilities for Rock and Pollack-Ohls (1987) Data

Item	Time 1	Time 2
Single operations	.67	.72
Powers, roots	.40	.47
Low-level algebra	.23	.33
Geometry	.06	.09

Table 5
Cross-Tabulation of Response Patterns for Rock and Pollack-Ohls (1987) Data

Response pattern	Twelfth grade 0000	1000	1100	1110	1111	Non-Guttman	Marginal
Tenth grade							
0000	217	133	24	6	0	28	408
1000	100	164	63	33	7	50	417
1100	10	40	73	84	12	29	248
1110	5	7	32	90	52	18	204
1111	0	0	3	17	44	3	67
Non-Guttman	18	30	30	52	6	20	156
Marginal	350	374	225	282	121	148	1500

(1982) and Bryk and Raudenbush (1987), involving fitting a regression equation for each individual's growth trajectory and then examining group differences in the regression coefficients. Use of the LGS methodology will result in improved measurement of monotonic growth, enhancing these analyses. There are four features of the LGS methodology that make it particularly suitable in this context. First, as discussed earlier the LGS methodology does not penalize for individual differences in growth rate. Thus, measures can be developed over an entire sample even if large individual or group differences in developmental rate are expected. Second, Rogosa et al. (1982) and Rogosa and Willett (1985) have called for use of more occasions of measurement in longitudinal studies. The LGS is unique among measurement models in that it can easily handle any number of occasions of measurement. For example, the second artificial data set discussed in the previous section involved six occasions of measurement. The LGS model places no upper limit on the number of occasions of measurement. Third, the LGS model is a model of monotonic growth, but not necessarily of linear growth, so the model is well suited to various kinds of nonlinear but monotonic growth. Fourth, CL is an important adjunct to research involving constructs that are believed to undergo cumulative, unitary development because it reveals the extent to which this is or is not the case. For example, CL showed that although math skills may form a hierarchical structure within times in the Rock and Pollack-Ohls (1987) data, these data offer little evidence for a hierarchy of monotonic skill acquisition across time.

The ordinal approach taken here may seem less desirable in some ways than positing continuous functions to describe change, as for example, Rogosa and Willett (1985) have done. However, it can be difficult to substantiate the idea that either observed scores produced by psychometric instruments or the latent variables estimated by them are defined more than ordinally (Cliff, 1989, in press). Thus, looking for monotonic change on such variables seems reasonable. Models more specific than the highly general one of monotonicity remain, of course, desirable. Collins (in press) suggested stochastic movement among latent classes as a more specific model, and presented ways of analyzing data to test the models and identify change.

Post Hoc Scale Development

We have emphasized the a priori development of instruments. However, scale development is not always a priori. Often a researcher is confronted with a large array of variables or items and wishes to explore the data for subscales. The aim of the exploration may simply be to reveal some substantively interesting subscale structure, or there may be a desire to go further and use scores on the subscales for subsequent analyses. LGSCLUS (Collins et al., 1988), a procedure for the exploration of data for longitudinal Guttman subscales, is an agglomerative, nonhierarchical clustering procedure using CL as a cluster homogeneity index. This procedure is one of the few exploratory procedures developed particularly for longitudinal research. (For more information about LGSCLUS, refer to Collins et al., 1988.)[1]

Measurement of Growth and Construct Validity: A More General Perspective

We have advanced an approach arguing that measurement of growth cannot proceed without an explicit a priori model of growth, and have demonstrated how a model of cumulative, unitary monotonic development has led to a measurement model. Our view that measurement theory is inextricably bound to substantive theory is rooted in some of psychology's most well-established measurement traditions. As Coombs (1964, p. 5) has remarked, "a measurement or scaling model is a theory of behavior, admittedly on a miniature level, but nevertheless theory. . . ." A longitudinal Guttman scale is an operational definition of a theory of cumulative, unitary development.

Such a close connection between substantive theory and measurement theory helps maximize construct validity. To take an example, consider a very simple theory of infant motor development. This theory predicts not only that motor skill increases from birth to 18 months, but that children learn first to grasp an object, then to sit up, then to stand, then to walk. Clearly a process like this is more than just an overall increase in motor skill. If it were merely this, the order in which infants attained the various skills would be irrelevant: Some babies would walk before sitting up, some would stand before being able to grasp an object, and so forth. Because this theory of infant motor development specifies that babies acquire skills in a particular order, any motor development measure motivated by the theory will reflect this order if it has good construct validity. To do otherwise is to ignore a theoretically central aspect of the growth process.

Collins et al. (1988) referred to latent variables that change systematically over time as *dynamic* and to latent variables that are not expected to change systematically as *static*. This terminology places the measurement of growth in the context of other measurement problems, and serves as a reminder that the concerns are somewhat different when growth is being studied.

[1] LGSCLUS software (Collins, Dent, & Cliff, 1986) and LGSINDEX software (Collins & Dent, 1986), which computes quantities helpful in instrument development, are available from the authors. These procedures are described in detail in Collins et al. (1988).

It also facilitates a broader conceptualization of growth. Dynamic latent variables may take many forms other than cumulative, unitary development, depending on substantive considerations. For example, some dynamic latent variables may be best represented as cumulative stage sequences where some individuals skip a particular stage; others may undergo development that is not cumulative; others are best represented as continuous autoregressive processes. Theory may predict gain during a certain period followed by loss in another period, or that individual change occurs at a certain average rate. In short, there is an almost infinite number of conceptualizations of the growth process.

This highlights an important limitation of the LGS model—it is suitable only for dynamic latent variables whose growth can be approximated by cumulative, unitary development. If the growth of interest in a particular research study is of some other form, then the LGS procedures would be a poor choice for instrument development. New methodologies are needed for the development of instruments measuring the many varieties of dynamic latent variables featured in psychological theory and research. These new methodologies will undoubtedly be quite different in many ways from the LGS, but those offering a high level of construct validity will start with a model of development.

References

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

Cliff, N. (1977). A theory of consistency of ordering generalizable to tailored testing. *Psychometrika, 42,* 375–399.

Cliff, N. (1989). Ordinal consistency and ordinal true scores. *Psychometrika, 54,* 75–91.

Cliff, N. (in press). Ordinal methods in the assessment of change. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change?* Washington, DC: American Psychological Association.

Collins, L. M. (in press). Measurement in longitudinal research. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change?* Washington, DC: American Psychological Association.

Collins, L. M., & Cliff, N. (1985). Axiomatic foundations of a three-set Guttman simplex model with applicability to longitudinal data. *Psychometrika, 50,* 147–158.

Collins, L. M., Cliff, N., & Dent, C. W. (1988). The longitudinal Guttman simplex: A new methodology for measurement of dynamic constructs in longitudinal panel studies. *Applied Psychological Measurement, 12,* 217–230.

Collins, L. M., & Dent, C. W. (1986). *LGSINDEX user's guide* (Health Behavior Research Institute Tech. Rep. No. 86-100). Los Angeles: University of Southern California.

Collins, L. M., Dent, C. W., & Cliff, N. (1986). *LGSCLUS user's guide* (Institute for Health Promotion and Disease Prevention Research Tech. Rep. No. 86-105). Los Angeles: University of Southern California.

Coombs, C. H. (1964). *A theory of data.* Ann Arbor, MI: Mathesis.

Guttman, L. (1950). The basis for scalogram analysis. In S. A. Stouffer, L. Guttman, E. A. Suchman, P. Lazarsfeld, S. A. Star, & J. A. Clausen (Eds.), *Measurement and prediction* (pp. 60–90). Princeton, NJ: Princeton University Press.

Loevinger, J. (1947). A systematic approach to the construction and evaluation of tests of ability. *Psychological Monographs, 61*(4, Whole No. 285).

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

Mokken, R. J. (1971). *A theory and procedure of scale analysis.* Hawthorne, NY: Mouton.

Rock, D. A., & Pollack-Ohls, J. (1987). *Measuring gains—a new look at an old problem.* Princeton, NJ: Educational Testing Service, Research Statistics Group.

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., & Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika, 50,* 203–228.

Rushton, J. P., Brainerd, C. J., & Pressley, M. (1983). Behavioral development and construct validity: The principle of aggregation. *Psychological Bulletin, 94,* 18–38.

Saupe, J. L. (1966). Selecting items to measure change. *Journal of Educational Measurement, 3,* 223–228.

Received February 9, 1987
Revision received April 3, 1989
Accepted October 6, 1989 ■