

# The Longitudinal Guttman Simplex: A New Methodology for Measurement of Dynamic Constructs in Longitudinal Panel Studies

Linda M. Collins, Norman Cliff, and Clyde W. Dent  
University of Southern California

Traditional psychometric procedures can be inadequate for the measurement of dynamic constructs in longitudinal panel studies. This paper introduces an alternative based on the longitudinal Guttman simplex (LGS) model, a measurement model developed especially for dynamic constructs measured longitudinally. The LGS is a model of cumulative, unitary development. It is cumulative in the sense that as persons acquire new skills (or abilities, or opinions), earlier obtained skills are retained; it is unitary in the sense that all persons progress through a sequence of skills in the same skill order. CL, a consistency index that gives the researcher a measure of the extent to which the LGS model axioms are obeyed in a given dataset, is introduced. By making use of this consistency index, the researcher can develop scales uniquely sensitive to cumulative, unitary development. LGSCUS, an exploratory procedure to find longitudinal Guttman scales in empirical datasets, is described. An artificial data study is reported, the purpose of which was to test the performance of LGSCUS under controlled conditions. The artificial data study showed that, in general, LGSCUS recovers longitudinal Guttman scales with a high degree of accuracy. There remains a need for measurement procedures for dynamic constructs exhibiting types of development other than cumulative and unitary. *Index terms: Dynamic constructs, Guttman simplex, Longitudinal panel studies, Mathematical models, Measurement theory, Scaling, Three-set data.*

---

APPLIED PSYCHOLOGICAL MEASUREMENT  
Vol. 12, No. 3, September 1988, pp. 217-230  
© Copyright 1988 Applied Psychological Measurement Inc.  
0146-6216/88/030217-14\$1.95

Recently a new perspective on longitudinal research has been emerging (Bryk & Raudenbush, 1987; Collins, 1983; Collins & Cliff, 1985, 1988; McArdle & Epstein, 1987; Rogosa, Brandt, & Zimowski, 1982; Rogosa & Willett, 1985). Central to this perspective is the recognition of a fundamental difference regarding the nature of the constructs of interest in longitudinal research as opposed to those in cross-sectional research. In cross-sectional research, constructs are generally static, at least for the purposes of a given study. In contrast, a major focus of longitudinal research is on constructs that change as a function of time, that is, *dynamic* constructs (Collins & Cliff, 1988; Singer & Spilerman, 1979).

Much research has been devoted to dynamic constructs, although the published literature has not always explicitly identified them as such. Examples of dynamic constructs include a number of theoretical and empirically-based conceptualizations of the development of substance use behavior in young people (e.g., Donovan & Jessor, 1983; Flay, d'Avernas, Best, Kersell, & Ryan, 1983; Yamaguchi & Kandel, 1984) suggesting a multistage process unfolding over time. Developmental psychology has contributed a number of theories involving dynamic constructs, including Piaget's theories about intellectual development, the fluid and crystallized intelligence constructs of Cattell (1963), and Kohlberg's (1969) stages of moral development.

Finding answers to research questions involving dynamic constructs often demands the rethinking of traditional methodological procedures. Although such rethinking is currently taking place with respect to the statistical treatment of dynamic constructs (e.g., Bryk & Raudenbush, 1987; Rogosa & Willett, 1985), there has been relatively little methodological research devoted to the psychometric treatment of these constructs. One exception is Collins and Cliff (1988), who suggested adopting a new approach for developing measures of dynamic constructs for longitudinal studies. These authors have shown that traditional methods of instrument development can be inadequate in the context of dynamic constructs. Collins and Cliff pointed out that traditional instrument development methods are based on test theories that assume an unchanging true score, an assumption obviously violated in the dynamic case. Instead of these methods, a measurement model designed especially for dynamic constructs measured longitudinally—one that incorporates a model of the change process—is needed.

#### The Longitudinal Guttman Simplex

Recently Collins and Cliff (1985; Collins, 1983) introduced the first measurement model for dynamic constructs measured in longitudinal panel designs. This model, the longitudinal Guttman simplex (LGS), is a special case of the three-set Guttman simplex model (Collins & Cliff, 1985). The LGS is an extension of the familiar Guttman simplex model (Guttman, 1950) to incorporate not only persons and items, but times as well.

The idea of a joint order is central to both the Guttman simplex and the LGS. The original Guttman simplex model is defined as a joint order of persons and items. This means that persons provide a difficulty order for items, and items provide an ability order for persons. This consistency allows persons and items to be ordered jointly in a Guttman scale (i.e., persons relative to items and items relative to persons). In contrast, the key feature of the LGS model is that for each person there is a joint items-times order, and that this order is con-

sistent across persons. That is, in a LGS a person's responses to a set of items at each of several times provide a difficulty order for the items and a temporal order for the times. This consistency allows items and times to be ordered jointly (i.e., items relative to times and times relative to items). In addition, this joint items-times order is consistent across persons. It follows from this (Collins & Cliff, 1985) that in a LGS the joint order of persons and times for each item is consistent, and that the joint order of persons and items for each time is consistent. However, even under these constraints a surprising degree of flexibility remains.

The above discussion about the psychometric definition of the LGS model can be expressed in more substantive terms. Simply put, the LGS model is an operational definition of a theory of cumulative, unitary (CU) development. Suppose items represent developmental stages. Then, according to the LGS model, each person provides an internally consistent Guttman scale made up of stages and times. Such a Guttman scale implies that persons retain the skills in each stage while advancing to new stages—that is, development is *cumulative*. The LGS model further specifies that this stages-times Guttman scale is consistent across persons. This implies that development is *unitary*, that is, that each person advances through the stages in the same order.

It is interesting to note several kinds of consistency that are not required by the LGS model. For example, the persons-items joint order, although consistent at each time, may vary across times due to differences in the persons order. Similarly, the persons-times joint order, while consistent for each item, may vary across items (see Collins & Cliff, 1985, Figure 2). But most important substantively, persons can advance through the stages at varying rates even though the data satisfy the stringent consistency requirements described above (Collins & Cliff, 1985). This makes the LGS model a realistic representation of development in many situations.

For the sake of clarity the above discussion has been in terms of ability measurement, but clearly the model generalizes to other types of constructs as well, such as psychosocial constructs.

The present paper introduces methodology enabling the LGS model to be applied in two areas. The first application is instrument development. In order for a researcher to select a group of items that forms a scale, there must be a method of determining how well the idealized measurement model describes empirical data. One approach to this problem that is quite familiar in psychometrics is the use of a consistency index, such as Cronbach's  $\alpha$  (Cronbach, 1951) or the numerous measures of fit for cross-sectional Guttman scales (Cliff, 1983).

A consistency index for longitudinal Guttman scales, CL, will be presented. CL gives the researcher an indication of the extent to which a set of empirical longitudinal data can be described by the LGS model, and helps identify poorly fitting items to be removed from an instrument. A researcher can thus construct improved measures of CU development in terms of measurement accuracy and construct validity.

The second application is that of empirical subscale definition. Sometimes a researcher working with a set of data has only a limited idea of its structure. Under these circumstances the researcher may wish to explore a database for evidence of internally consistent subscales. The subscales may be of use in further multivariate analysis, or they may be interesting in and of themselves. LGSCCLUS, an exploratory procedure for longitudinal data, will be presented. LGSCCLUS is an agglomerative, non-hierarchical clustering procedure based upon BINCLUS (Cliff, McCormick, Zarkin, Cudeck, & Collins, 1986). The aim of LGSCCLUS is to cluster items (or variables) into scales such that within scales the responses display the CU properties described above, at least to a close approximation.

#### CL: Assessing Consistency in Longitudinal Data Through Order Relations

The strong axiomatic foundation of the LGS model (Collins & Cliff, 1985) provides a basis for developing a consistency index by counting order relations, that is, actual instances of adherence to and deviations from model axioms. Based on the

model axioms, it is possible to identify and count patterns that are confirmatory of CU development, and instances of departures from it, within empirical data. This method is an adaptation of Cliff's (1979) approach to developing consistency indices for cross-sectional data.

Counting order relations begins with person response patterns, each of which is made up of a single person's responses to two items at two times. Most of these  $2 \times 2$  response patterns indicate an order for the times and/or items; that is, they contain order relations. Some of the order relations indicate CU development while others are inconsistent with CU development. There are three general categories of order relations: redundant, unique, and contradictory. CL is a weighted combination of these relations.

#### Redundant and Unique Order Relations and CU Development

According to the LGS model, development is cumulative and unitary; that is, there is an items-times joint order for each person, and this items-times joint order is consistent across persons. The purpose of identifying order relations is to reveal the extent to which this is true in observed data. *Redundant* order relations are person response patterns where either (1) both items agree on an order for the times, or (2) both times agree on an order for the items. *Unique* order relations are person response patterns where only one item of the pair provides an order for the times (and only one time of the pair provides an order for the items; one implies the other). *Contradictory* order relations are person response patterns where the two items disagree on the order for the times (and the two times disagree on the order for the items; again, one implies the other).

The extent to which the LGS model describes a particular dataset is assessed by comparing the persons' joint orders to see how consistent they are. Contradictory relations clearly conflict with the idea of an internally consistent joint order, while redundant and unique relations must be examined to see whether they are consistent across persons.

However, comparing each person's response patterns to every other person's response patterns is impractical for large numbers of persons. Fortunately, an alternative strategy exists, that of determining an a priori order for times and for items against which each order relation is compared. An order relation agreeing with these a priori orders is evidence suggesting that the LGS model describes the data; a disagreeing order relation is evidence suggesting that the LGS model does not describe the data. The a priori order for times is obviously the natural (i.e., temporal) order. The a priori order for items is equally obvious: The proportion failing (or not endorsing) each item provides the modal item order.

**Identifying Redundant, Unique, and Contradictory Relations**

*Redundant relations.* There are two types of redundant relations: redundant time relations and redundant item relations. Redundant time relations occur whenever a person's responses to a pair of items agree on an ordering for a pair of times; the term "redundant" is used because the time ordering provided by one item is redundant with that provided by the other. Redundant item relations do the reverse, providing an order for items. Consider the following responses of a person to two items at two times:

	Item A	Item B
Time 1	fail	fail
Time 2	pass	pass

These two items agree on an ordering for the times, with both items showing Time 1 to be the "more difficult" or "earlier" time, so this is an example of a redundant time relation. Similarly, the following pattern shows a redundant item relation because the responses at both times agree on an ordering for a pair of items:

	Item A	Item B
Time 1	pass	fail
Time 2	pass	fail

Redundant relations may either agree or disagree

with the a priori order. Redundant relations agreeing with the a priori order are denoted *R*, while those disagreeing with the a priori order are denoted *R'*.

*Unique relations.* In some response patterns, only one of a pair provides an order. If only one of a pair of items orders the times, then it follows that only one of the times orders the items, as shown in the following example:

	Item A	Item B
Time 1	pass	pass
Time 2	fail	pass

Response patterns of this type are called unique relations because the ordering is unique to only one item and one time. Unique relations are of special interest in assessing the consistency of longitudinal Guttman scales because these relations directly reflect the joint order of items and times. For example, this response pattern indicates that Item A is more difficult than Item B, and also that Time 2 is more difficult than Time 1. Unique relations where *both* orders agree with the a priori orders are denoted *U*. Unique relations where one or both orders disagree with the a priori orders are denoted *U'*.

*Contradictory relations.* Contradictory relations provide inconsistent ordering information about both items and times, as in the following example:

	Item A	Item B
Time 1	fail	pass
Time 2	pass	fail

A response pattern of this type offers conflicting information about the ordering of both items and times. Contradictory relations are denoted *C*.

**The Construction of CL**

The above discussion has shown how consistency of an empirical dataset with the axioms of the LGS model can be assessed by counting order relations in the data. The purpose of counting these relations is to combine them to form a consistency index. However, although redundant, unique, and contradictory relations are the building blocks of

CL, arriving at the correct combination of these relations to serve as a consistency index for longitudinal Guttman scales presents a difficulty. It seems reasonable to use a weighted composite of the individual relations, but the difficulty lies in determining that one weighted combination is better than another. Usually, new indices that are being evaluated are compared with other well-established indices for the same purpose. However, in this case no such indices exist, and available indices for cross-sectional scales are not readily adaptable to the LGS case.

In lieu of such a standard, three characteristics desirable in a consistency index were specified, with emphasis on practical considerations. That is, the goal was to develop a consistency index that will be of real value to the practitioner who is trying to develop or evaluate a longitudinal Guttman scale. First, the index should be conceptually simple and easily interpretable. Second, the index should adjust for the effects of differing item marginals so that a set of independent items cannot produce a high index value, as occurs with Guttman's reproducibility index (White & Saltz, 1957). This is an extremely important attribute because it affects construct validity. Dynamic constructs are most often examined in the context of overall development occurring in a number of different areas. This can cause a certain number of consistent redundant and unique relations, even between items that do not reflect the same construct. If left unadjusted, the consistency index could produce spuriously high values whenever there is a sufficient amount of global development on unrelated items, leading to the erroneous conclusion that the items form a LGS.

The third goal has to do with the choice of weights to apply to the relations before combining them to form an index. Obviously  $R$  and  $U$  should receive positive weights, while  $R'$ ,  $U'$ , and  $C$  should receive negative weights. The possibilities range from equal weights, which would mean that each of the various types of relations contribute equally to the consistency index, to weights that count some of the relations more heavily than others. Rather than arbitrarily selecting a set of weights, as was done in Collins (1983), consideration was given to se-

lecting weights that would produce a highly sensitive index.

Researchers are ordinarily interested in how good a certain scale is, or which of several scales is best, or whether removing a certain item will improve the consistency of a scale appreciably. This suggests that a consistency index should reflect even small differences in consistency. The coefficient of reproducibility and the unadjusted Rand statistic (Rand, 1971) are examples of indices that sometimes tend to be insensitive at the upper end of the range; that is, they may not reflect small differences among scales that are fairly consistent. Thus, it was decided that a third goal in developing the consistency index would be to select a combination of weights resulting in sharp discrimination among datasets of similar degrees of consistency, even when the overall level of consistency is high.

#### Computation of CL

CL, an index that meets these three criteria, is computed as follows:

$$CL = \frac{\frac{\text{Consistent}}{\text{Total}} - \frac{\text{Expected Consistent}}{\text{Expected Total}}}{1 - \frac{\text{Expected Consistent}}{\text{Expected Total}}}, \quad (1)$$

where  $\text{Consistent} = R + 4U$  and  $\text{Total} = R + R' + C + 4(U + U')$ . The quantity in Equation 1 is recognizable as a quality index (e.g., Cliff, 1979), interpretable as the proportion of consistent relations over and above those attributable to differences in item marginals, or alternatively as a measure of relative improvement over chance. Hence this definition for CL meets the first and second goals discussed above, because it has a straightforward interpretation and because it incorporates an adjustment for chance.

In general, the denominator of a quality index reflects maximum possible improvement over chance. This is evident in Equation 1. The right side of the denominator in Equation 1 is the weighted proportion of consistent relations expected if the items are independent (Collins, Dent, & Cliff,

1985a), while the left side, unity, is the largest possible improvement over this—that is, all consistent relations. The numerator represents the observed improvement over chance, subtracting the weighted expected proportion of consistent relations from the weighted observed proportion of consistent relations. An index of the form in Equation 1 ranges from 1 to  $-\infty$ , with 0 the value when there are no more consistent relations than would be expected due to chance. A negative value is unlikely, representing a scale whose consistency is worse than would be expected due to chance.

*Good discrimination.* The third goal was to develop a sensitive index. The quest for such an index was undertaken by constructing 29 indices composed of various weighted combinations of relations, and then examining the performance of these indices with artificial data of known consistency (Collins et al., 1985a, 1985b). Each of the 29 indices was evaluated by determining whether it could (1) distinguish between random and non-random data and (2) make fine distinctions among various levels of consistency. CL showed the best overall differentiation, and in addition was shown to be robust across a number of conditions, including low person-to-variable ratios (Collins et al., 1985b). For these reasons CL is suggested as a measure of longitudinal consistency and as the basis of the exploratory clustering procedure described below.

**An Example of the Use of CL  
 to Construct a Longitudinal  
 Guttman Scale**

The following hypothetical example illustrates

the use of the LGS model and CL to develop a measure of CU development in longitudinal research. The example involves eight hypothetical test items administered at three occasions.

Suppose the problem is to use a subset of these eight items to construct a scale measuring the development of math skill in children. As Table 1 shows, the substantial increases in passing probability suggest that each item is measuring some type of development; the problem is whether all of the items are measuring the development of the same dynamic math skill construct. The CL for all eight items together is .81. Table 1 shows that if Item 6 is deleted from the scale, CL increases substantially. If desired, CL-if-item-deleted could be recomputed with the new seven-item scale in order to achieve a still higher degree of consistency, but in this case it seems reasonable to stop here. The result is a scale made up of Items 1 through 5, 7, and 8 measuring CU development of math skill.

The item passing probabilities in Table 1 suggest that children master addition first, then subtraction, then multiplication, then division. It is interesting to note that Items 2 and 4 are constants at Grade 2, because none of these hypothetical children had mastered division by the second grade. Although items that are constant at one occasion present a problem for many procedures, these items are not a problem for LGS procedures if they reflect change across time.

**LGSCLUS: Exploring Longitudinal  
 Data for Subscales**

Relatively few alternatives are available to the

Table 1  
 Hypothetical Example of the Use of CL to Select Scale Items

Item	Skill	Passing Probability			CL If Item Deleted
		Grade 2	Grade 3	Grade 4	
1	Subtraction	.3	.8	.9	.82
2	Division	.0	.2	.4	.77
3	Multiplication	.2	.6	.9	.81
4	Division	.0	.1	.3	.77
5	Multiplication	.2	.4	.8	.79
6	Vocabulary	.2	.7	.7	.95
7	Addition	.4	.9	.9	.77
8	Subtraction	.3	.8	.9	.84

researcher who wishes to explore the structure of a longitudinal dataset for dynamic subscales. Either the data must be forced into a structure intended for use with two-set data, or a three-set procedure not developed for longitudinal data must be used, such as three-mode factor analysis (Tucker, 1964). LGSCLUS differs from these procedures in a number of important respects. First, LGSCLUS is designed explicitly for longitudinal data. Second, the purpose of LGSCLUS is to find subscales that fit the LGS model, so that when LGSCLUS searches for longitudinal Guttman scales, it locates sets of items each reflecting the CU development of a single dynamic construct. This incorporation of a specific model of change into the procedure makes LGSCLUS unique among exploratory methods; in fact, it is the only exploratory method based on the idea of dynamic hypothetical constructs.

### The LGSCLUS Algorithm

The algorithm used by LGSCLUS is essentially the same as the algorithm used by BINCLUS (Cliff et al., 1986); as such, it is largely a heuristic procedure, providing for user intervention at several points. The procedure is as follows: At the outset, each variable starts a separate cluster. The next variable to be added is the variable showing the largest value of CL with the starter variable, followed by the variable having the largest average CL with the two variables already in the cluster, and so on. The history of this process is recorded in the step-by-step Cluster History Matrix. This matrix records which variable is added to each cluster at each step, as well as the average CL value of each newly-added variable with the other variables in the cluster. The process is continued until all variables are added to all clusters, or until a user-specified minimum for CL is reached.

*Defining cluster boundaries.* The cluster boundary selected by the user separates those variables in the cluster from those outside the cluster. LGSCLUS gives the user three options for defining cluster boundaries.

The first option is known as the Largest Gap Rule. The idea behind the Largest Gap Rule is that

a naturally occurring cluster boundary should show itself by a large decrease in the average CL when the first variable outside the cluster is added. When the user selects the Largest Gap Rule, the program searches the Cluster History Matrices for the largest decrease in average CL when a new variable is added (excluding the addition of the first variable). The boundary is then established just before that variable. The user is also asked to specify a minimum CL so that if the largest gap occurs after this minimum, the boundary is established at the minimum. This avoids producing clusters that consist entirely of unrelated elements.

The second option, called the First- $k$  Rule, is simply to select the first  $k$  variables added to the cluster to include in cluster  $i$ . For example, if an eight-item cluster is desired, the cluster boundary is drawn after the eighth item to enter. The third option is to select a constant cutpoint, so that no variable producing an average CL less than this value will be allowed to enter a cluster.

Once the cluster boundaries have been established, the results are displayed in a binary cluster membership matrix, where a 1 indicates that the row variable is a member of the column cluster.

### Second-Order Solutions

When data have an exceptionally clear subscale structure, the procedure outlined above may be all that is necessary to reveal that structure. However, noise in empirical data often obscures the underlying structure, making further analysis desirable. Like BINCLUS, LGSCLUS can make use of a second-order solution to clarify subscale structure. The second-order solution is performed on the binary items  $\times$  clusters membership matrix, which is treated like a data matrix and clustered using BINCLUS. That is, rather than variables being grouped into clusters, in the second-order solution clusters are grouped into superclusters. Clustering is performed using a matrix of gamma correlations (Goodman & Kruskal, 1954) for the measure of association, and using the Largest Gap Rule to determine the cluster boundary. The second-order solution is displayed by a permuted clusters  $\times$

superclusters membership matrix, and items are related to superclusters by means of a matrix  $\mathbb{P}$  that gives the proportion of clusters in the column supercluster containing the row item.

#### Artificial Data Study

The purpose of the artificial data study was to investigate the performance of LGSCLUS under controlled conditions. A number of artificial datasets with known subscale structure were submitted to LGSCLUS analysis, and the ability of the procedure to recover the underlying cluster structure was assessed.

#### Method

*Data generation.* Data were generated by simulating a situation where sets of items (tests) are administered repeatedly to persons whose ability is increasing between test administrations. Specifically, data were generated by means of the two-parameter logistic model (Birnbaum, 1968):

$$P_i(\theta_m) = \frac{1}{1 + \exp[-1.7a_i(\theta_m - b_i)]} \quad (2)$$

where  $\theta_m$  is the person ability on scale  $m$ ,  
 $a_i$  is the item discrimination,  
 $b_i$  is the item difficulty, and  
 $P_i(\theta_m)$  is the probability of passing item  $i$  in scale  $m$ .

A scale was a set of items to which responses were generated using the same  $\theta_m$ , that is, the same person ability parameter. Each dataset contained two longitudinal Guttman scales as well as four singleton items (i.e., items which did not belong to the scales). Each of the two scales and the four singleton items was generated by a different  $\theta_m$ .

In order to simulate a longitudinal testing situation, the simulated persons were administered the scale items at three different times. Persons were initially sampled from a distribution of ability, and this ability was used for the first test administration. Then person ability was incremented twice, with the same items administered after each increment. Item discrimination and item difficulty were not

altered as person ability was incremented. Thus, the simulated situation was one where persons took the same test three times, with person ability increasing between test administrations. Person ability was increased by adding a random increment to each person's  $\theta_m$  at each time.

On each trial,  $P_i(\theta_m)$  was compared with a random number sampled from a uniform distribution. If  $P_i(\theta_m)$  exceeded the random number, the item was considered passed.

*Independent variables.* Four factors were varied in the artificial data study:

1. Number of persons was varied in order to examine the effect of sample size. There were two levels:  $N = 50$  and  $N = 200$ .
2. The effect of item discrimination was studied using  $a = 1$  or  $2$ .
3. Because Collins (1983) suggested that the rate of person ability increase affects scale consistency, this was varied in the present study. In the slow-increase condition, ability increments were sampled from a uniform distribution over the range (0,2). In the fast-increase condition, ability increments were sampled from a uniform distribution over the range (0,4). Within an increase rate condition, ability increase was uncorrelated with  $\theta_m$ .
4. The scale overlap factor contained three conditions: (1) uncorrelated, non-overlapping scales; (2)  $\theta_m$  correlated  $r = .2$  between scales; and (3)  $\theta_m$  uncorrelated between scales, but one item belonging in both scales generated using the mean of the two  $\theta$ s. In the first two conditions there were 16 items in each dataset, consisting of one eight-item cluster, one four-item cluster, and four singleton items. The third condition involved these items plus the overlap item, resulting in a total of 17 items.

All four independent variables were fully crossed, resulting in a 24-cell design. Ten replications were performed in each cell, resulting in a total of 240 datasets. The LGSCLUS procedure was applied to each of these. The analysis was performed twice for each dataset, once using the First- $k$  Rule and once using the Largest Gap Rule. The choice of cluster boundary rule determined which items be-

longed in which clusters for the first-order solution. In the case of the second-order solution, a rule applied to the  $P$  matrix determined which items made up the superclusters. If an item occurred in half or more of the clusters in a supercluster, it was considered part of the supercluster. The subscale structure revealed by LGCLUS was compared to the known subscale structure for both the first-order and second-order solutions.

## Results

*Evaluating the cluster solutions.* The LGCLUS solutions were evaluated by comparing the clusters found in each dataset to the known subscale structures. In order to quantify this comparison, an index of cluster recovery was needed. Because the familiar Rand index (Rand, 1971) is not suitable when clusters overlap, as they did in the present study, the Omega index (Collins & Dent, in press) was used to compare cluster structures. This index is a generalization of the Hubert and Arabie (1985) adjusted Rand index, applicable to both overlapping and non-overlapping cluster structures. It is important to note that the unadjusted Rand index tends to be large; in fact, it can be quite large even in situations where there is no agreement between two solutions beyond what would be expected due to chance. Because the Omega index makes use of the adjustment for chance agreement suggested by Hubert and Arabie (1985), it tends to be lower than the Rand index. Its advantages are increased discrimination among various levels of cluster recovery and the ability to compare clustering methods whether one, both, or neither yield overlapping clusters.

As noted above, LGCLUS returns two solutions, a first-order solution and a second-order solution. The Omega index was computed for both solutions.

*Overall cluster recovery.* Tables 2 and 3 contain information on agreement between the true cluster structure and the obtained cluster structure. As the first-order solution is an intermediate step in the LGCLUS procedure, the results will be discussed only briefly. Table 2 shows the mean Omega for the first-order solutions and Table 3 shows the

number of datasets in each condition (out of a possible 10) where recovery was perfect. Recovery is shown for the 24 experimental conditions and for both cluster boundary rules. Table 2 shows that mean recovery was  $\Omega = .90$  or above in 12 of the 24 conditions when the First- $k$  Rule was used, and Table 3 shows perfect recovery in half or more of the datasets in each of these conditions. There was perfect recovery in every case in the conditions involving No Overlap and  $N = 200$  when the First- $k$  Rule was used. Recovery was much poorer in general for the Largest Gap Rule, particularly in the One Item Overlap and Correlated Scales conditions.

Tables 2 and 3 also contain the mean Omegas and number of perfect solutions, respectively, in each condition for the second-order solutions. Comparison of the two solutions shows that the second-order solutions performed their assigned task of clarifying the first-order solutions. There was improvement in mean recovery by the First- $k$  Rule in every condition where improvement was possible, and there was improvement in mean recovery in every condition when the Largest Gap Rule was used. Using the First- $k$  Rule, the second-order solution produced perfect recovery in all 10 replications for 9 out of the 24 conditions, and mean Omega above .90, with medians of 1.0, in 19 conditions.

When the Largest Gap Rule was used, the second-order solutions effected a dramatic improvement, achieving recovery comparable to that of the First- $k$  Rule in all No Overlap,  $N = 200$  conditions. Despite this, overall recovery was still not as good as with the First- $k$  Rule, particularly in the One Item Overlap and Correlated Scales conditions.

*Effects of independent variables.* Recovery was noticeably better with  $N = 200$ , and this effect was particularly strong in the more difficult situations of overlapping or correlated clusters. The presence of a correlated scale structure had a particularly large effect, although even there the First- $k$  Rule was quite successful when  $N = 200$ . The First- $k$  Rule outperformed the Largest Gap Rule overall, except in one condition where both were perfect.

Table 2  
 Mean Agreement ( $\Omega$ ) Between True Cluster  
 Structure and LGSCLUS Solution

Solution	N	Fast Development		Slow Development	
		a=2	a=1	a=2	a=1
<b>First-Order Solutions</b>					
<b>First-k Rule</b>					
No Scale Overlap	200	1.000	1.000	1.000	1.000
	50	.888	.968	.895	.936
One Item Overlap	200	.952	.847	.975	.931
	50	.850	.765	.675	.865
Correlated Scales	200	.983	.956	.908	.862
	50	.624	.852	.381	.521
<b>Largest Gap Rule</b>					
No Scale Overlap	200	.439	.587	.601	.869
	50	.287	.428	.416	.452
One Item Overlap	200	.223	.223	.269	.427
	50	.237	.140	.337	.113
Correlated Scales	200	.197	.271	.171	.379
	50	.117	.231	.096	.224
<b>Second-Order Solutions</b>					
<b>First-k Rule</b>					
No Scale Overlap	200	1.000	1.000	1.000	1.000
	50	.954	.988	.988	1.000
One Item Overlap	200	1.000	.987	1.000	1.000
	50	1.000	.848	.806	.961
Correlated Scales	200	.988	.967	.909	.910
	50	.742	.935	.642	.553
<b>Largest Gap Rule</b>					
No Scale Overlap	200	.985	.951	.998	1.000
	50	.887	.816	.941	.833
One Item Overlap	200	.801	.777	.808	.883
	50	.781	.882	.690	.730
Correlated Scales	200	.836	.661	.549	.683
	50	.551	.676	.455	.410

Table 3  
 Number of LGSCLUS Solutions Showing Perfect Recovery

Solution	N	First-Order				Second-Order			
		Fast		Slow		Fast		Slow	
		a=2	a=1	a=2	a=1	a=2	a=1	a=2	a=1
<b>First-k Rule</b>									
No Scale Overlap	200	10	10	10	10	10	10	10	10
	50	7	9	7	9	9	9	9	10
One Item Overlap	200	9	5	9	9	10	9	10	10
	50	5	4	2	5	10	7	7	8
Correlated Scales	200	9	8	6	6	9	8	6	6
	50	2	5	0	1	4	6	1	1
<b>Largest Gap Rule</b>									
No Scale Overlap	200	1	2	1	7	9	8	9	10
	50	0	0	0	0	6	5	6	4
One Item Overlap	200	0	0	0	0	2	2	2	3
	50	0	0	0	0	3	3	0	1
Correlated Scales	200	0	0	0	0	3	2	0	2
	50	0	0	0	0	2	1	0	0

There was a slight tendency for recovery to be better in the Fast Development condition, particularly with the  $N = 50$ , Correlated Scales conditions. Surprisingly, item discrimination had little overall effect.

**An Example of the Use of LGSCLUS to Find Longitudinal Subscales**

The following hypothetical example involves seven items administered at three occasions, and was constructed so that Items 1, 3, and 4 form one scale, Items 5, 6, and 7 form another scale, and Item 2 does not belong in either scale. The LGSCLUS procedure applied to these data yielded the Cluster History Matrix in Table 4. Clearly, LGSCLUS has recovered the true subscale structure in these artificial data. When Items 1, 3, and 4 begin a cluster they add each other first, maintaining a high CL value. As soon as other items are added to the cluster, the CL value drops to unacceptable levels. Items 4, 5, and 6 behave in a similar fashion. Table 4 gives abundant evidence that Item 2 does not belong in either of the two clusters. First, the cluster beginning with Item 2 shows a very low CL when the first item is added, indicating that Item 2 does not have a strong relationship to any item in the set. CL increases slightly when subsequent items are added to this cluster, but this is because of the consistency of the added items with each other. Second, examination of the matrix of average CL values shows that when Item 2 is added to a cluster, it always shows a low average CL with the previously added items.

For this example, any of the cluster boundary rules works. For example, the Largest Gap Rule

shows that there is a sharp decrease in average CL upon the addition of Item 2 to any of the clusters, and that the average CL is very low as soon as the first item is added to Cluster 2. Thus the boundary would be drawn before the addition of Item 2 to Clusters 1 and 3 through 5, and before the addition of any items to Cluster 2. In empirical data the first-order solution is unlikely to be as clear as this. The second-order solution would then be used to improve the first-order solution.

**Discussion**

**Using LGS Methodology for Instrument Development**

When measurement of a dynamic construct is of interest, the LGS methodology can be used as an alternative to traditional psychometric procedures. This new methodology allows the researcher to build instruments tailor-made to measure cumulative, unitary developmental processes. The researcher wishing to apply the LGS methodology to such a problem begins in the usual manner, by constructing a pool of test, questionnaire, or behavior items. Then the items are administered to a group of persons *on several occasions*. A minimum of two occasions of measurement in a longitudinal panel design is necessary. In addition, enough time must be allowed between occasions of measurement for some development to be exhibited.

The next step is to assess the consistency of the resulting scale. An index designed for that purpose, CL, has been presented here. If the CL index is large, then a reasonably consistent instrument has been developed. If the CL index is small, then the

Table 4  
 LGSCLUS Cluster History Matrices for Hypothetical Data

Step	Starting Item						
	1	2	3	4	5	6	7
1	4 1.00	6 .13	4 1.00	3 1.00	7 1.00	7 1.00	6 1.00
2	3 1.00	7 .25	1 1.00	1 1.00	6 1.00	5 1.00	5 1.00
3	2 -.08	5 .13	2 -.08	2 -.08	2 .07	2 .07	2 .07
4	7 .06	1 -.09	7 .06	7 .06	1 -.09	1 -.09	1 -.09
5	5 -.03	4 -.10	5 -.03	5 -.03	4 -.10	4 -.10	4 -.10
6	6 -.01	3 -.02	6 -.01	6 -.01	3 -.02	3 -.02	3 -.02

instrument may need improvement. (Users should be aware that the present authors' limited experience shows CL to have values considerably lower than traditional indices, even with good data.) Improvement can often be effected by removing one or two items that are causing a disproportionate amount of inconsistency. CL-if-item-deleted can be computed to help the researcher in identifying such items. This item selection procedure is conceptually the same procedure as is typically followed in cross-sectional data using Cronbach's  $\alpha$  or KR-20. Failure of data to support a LGS may be due to overall poor data quality, a pool of items that is largely insensitive to change, or an underlying construct that does not exhibit CU development as hypothesized.

#### Using LGSCLUS for Data Exploration

Sometimes a researcher is interested in post hoc discovery of constructs in data, rather than a priori measurement. If the data are dichotomous longitudinal panel data and it is thought that they might contain instances of cumulative, unitary development, LGSCLUS can be used to reveal sets of items that approximate longitudinal Guttman scales. LGSCLUS, like BINCLUS (Cliff et al., 1986), is an agglomerative, nonhierarchical clustering procedure involving both first- and second-order solutions. LGSCLUS takes as input a set of binary items (or variables) measured longitudinally. It then clusters the items, using as a similarity measure a variables  $\times$  variables matrix of CL. The user inputs cluster boundaries, either on a cluster-by-cluster basis or by using the First- $k$  Rule or Largest Gap Rule.

The artificial data study presented here tested the LGSCLUS procedure in a number of situations. Several findings useful to empirical researchers emerged from the results of this study. The excellent performance of the First- $k$  Rule suggests that the best use of LGSCLUS will be made by researchers who have specific a priori ideas about cluster size, or who study the cluster history matrices carefully in order to arrive at cluster boundaries that seem reasonable based on the observed data and substantive

theory. For situations where a priori ideas about how many variables to include in a cluster are not available, the artificial data study indicates that the Largest Gap Rule works well where correlated scales are not present and sample sizes are sufficient.

The finding that LGSCLUS is robust to variations in developmental rate (within the limits of variation used here) has implications for the scheduling of occasions of measurement in empirical studies. LGSCLUS can uncover longitudinal Guttman scales even if relatively little development has been shown between waves of measurement, as can occur either if development is occurring at a slow rate or if waves of measurement are close together. This suggests that under some circumstances a researcher interested in cumulative, unitary development need not collect years of data. Instead, it may be possible to collect longitudinal data over a fairly short time. Care should be taken, however, to ensure that there is opportunity for development on every item, and that there is reasonable opportunity for overall development.

The simulation study also showed that LGSCLUS is robust to variations in item quality, indicating that the procedure will perform well in situations involving empirical data, which of course are likely to contain considerable error. The results involving sample size suggest that, where possible, an  $N$  greater than 200 should be used.

Although LGSCLUS works best when there is no overlap between clusters, it is also quite successful when there is item overlap between scales. However, caution is indicated where there are correlations between scales. In fact, the between-scale correlation used in the present study, .2, was chosen because in pilot studies it was found that higher correlations resulted in uniformly poor scale recovery.

#### The LGS Perspective on Measurement

The perspective on the measurement of dynamic constructs taken in this article is considerably different from that taken in past literature. First, the distinction between static and dynamic constructs, although seldom made explicitly, is seen as an im-

portant one. Second, most psychometric or statistical approaches to longitudinal data start from a persons  $\times$  items data matrix; in contrast, the present approach focuses on intra-person change over time, starting with an items  $\times$  times matrix for each person.

As an example of how central this difference is, consider the role that constant items play in most psychometric and statistical procedures. An item that is zero-variance at one wave of measurement in a longitudinal study either plays no role in methodological procedures, or must be removed before the procedures can be performed (as in factor analysis). However, an item that is constant at one or more occasions may play a very important role in a longitudinal Guttman scale, if it reflects intra-person change across occasions.

But the most important difference is the vital role that a model of the change process plays in the present authors' perspective on measuring dynamic constructs. In the case of the LGS methodology presented here, the model is one of cumulative, unitary development. However, although this model is one with wide applicability, it certainly does not characterize every dynamic construct of potential interest to psychologists. The LGS methodology is limited to one specific type of dynamic construct, but the issues raised by Collins and Cliff (1988) about the shortcomings of traditional measurement theory apply to other dynamic constructs as well. Further methodological work is needed for measurement of dynamic constructs involving stage-skipping, regression to previous stages, multiple paths, and other types of development that are not a part of the LGS model.

## References

- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick, *Statistical theories of mental test scores* (pp. 397-472). Reading MA: Addison-Wesley.
- Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, *101*, 147-158.
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*, 1-22.
- Cliff, N. (1979). Test theory without true scores? *Psychometrika*, *44*, 373-393.
- Cliff, N. (1983). Evaluating Guttman scales: Some old and new thoughts. In H. Wainer & S. Messick (Eds.), *Principals of modern psychological measurement*. Hillsdale NJ: Erlbaum.
- Cliff, N., McCormick, D. J., Zarkin, J. L., Cudeck, R. A., & Collins, L. M. (1986). BINCLUS: Nonhierarchical clustering of binary data. *Multivariate Behavioral Research*, *21*, 201-227.
- Collins, L. M. (1983). *The longitudinal Guttman simplex: A new scaling model for longitudinal data*. Unpublished doctoral dissertation, University of Southern California.
- 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. (1988). *Dynamic constructs and the measurement of individual change*. Manuscript submitted for publication.
- Collins, L. M., & Dent, C. W. (in press). The Omega index: A general approach to assessing cluster recovery suitable for hierarchical and overlapping solutions. *Multivariate Behavioral Research*.
- Collins, L. M., Dent, C. W., & Cliff, N. (1985a). *Development of a consistency index for the LGS model* (Technical Report 85-6). Los Angeles: University of Southern California, Health Behavior Research Institute.
- Collins, L. M., Dent, C. W., & Cliff, N. (1985b). *Robustness of CL, a consistency index for longitudinal Guttman scales* (Technical Report 85-7). Los Angeles: University of Southern California, Health Behavior Research Institute.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 294-334.
- Donovan, J. G., & Jessor, R. (1983). Problem drinking and the dimension of involvement with drugs: A Guttman scalogram analysis of adolescent drug use. *American Journal of Public Health*, *73*, 543-552.
- Flay, B. R., d'Avernas, J. R., Best, J. A., Kersell, M. W., & Ryan, K. B. (1983). Cigarette smoking: Why young people do it and ways of preventing it. In P. McGrath & P. Firestone (Eds.), *Pediatric and adolescent behavioral medicine*. New York: Springer-Verlag.
- Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross-classification tables. *Journal of the American Statistical Association*, *49*, 704-732.
- Guttman, L. (1950). The basis for scalogram analysis. In S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A. Star, & J. A. Clausen, *Mea-*

- surement and prediction (*Studies in social psychology in World War II, Vol. 4*). Princeton NJ: Princeton University Press.
- Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, 2, 193–218.
- Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 347–479). Chicago: Rand-McNally.
- McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development*, 58, 110–133.
- Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66, 846–850.
- Rogosa, D., 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.
- Singer, B., & Spilerman, S. (1979). Mathematical representations of development theories. In J. R. Nesleiroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development*. New York: Academic Press.
- Tucker, L. R. (1964). The extension of factor analysis to three-dimensional matrices. In N. Frederiksen & H. Gulliksen (Eds.), *Contributions to mathematical psychology*. New York: Holt, Rinehart and Winston.
- White, B. W., & Saltz, E. (1957). The measurement of reproducibility. *Psychological Bulletin*, 54, 81–90.
- Yamaguchi, K., & Kandel, D. B. (1984). Patterns of drug use from adolescence to adulthood: II. Sequences of progression. *American Journal of Public Health*, 74, 668–672.

#### Acknowledgments

*This research was funded by the National Institute on Drug Abuse, grant R23DA3673, and the National Science Foundation, grant BNS8403126.*

#### Author's Address

Send requests for reprints or further information to Linda M. Collins, Department of Psychology, University of Southern California, Los Angeles CA 90089-1061, U.S.A.