American Educational Research Association

The Properties of Bloom's Taxonomy of Educational Objectives for the Cognitive Domain
Author(s): G. M. Seddon
Reviewed work(s):
Source: *Review of Educational Research*, Vol. 48, No. 2 (Spring, 1978), pp. 303-323
Published by: American Educational Research Association
Stable URL: http://www.jstor.org/stable/1170087
Accessed: 04/11/2012 15:22

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
http://www.jstor.org/page/info/about/policies/terms.jsp

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

American Educational Research Association is collaborating with JSTOR to digitize, preserve and extend access to *Review of Educational Research.*

http://www.jstor.org

Review of Educational Research
Spring 1978, Vol. 48, No. 2, Pp. 303–323

The Properties of Bloom's Taxonomy of Educational Objectives for the Cognitive Domain

G. M. Seddon

University of East Anglia

There is no doubt that Bloom's *Taxonomy of Educational Objectives* for the cognitive domain (Bloom, Englehart, Furst, Hill, & Krathwohl, 1956) has had a considerable impact on educational thought and practice all over the world. For example, the bibliographies by Cox and Unks (Note 1) and Cox and Wildeman (Note 2) list a large number of articles and reports describing applications of the taxonomy. The taxonomy has also been cited in dozens of papers concerned with educational research and development each year (ISI, 1970-1976). In this respect it has achieved one of Bloom's criteria for its success:

> The taxonomy must be accepted and used by workers in the field if it is to be regarded as a useful and effective tool. (p. 24)

In addition, the authors of the taxonomy made various claims about its properties that concern both educational and psychological issues. The educational issues are concerned with the categorization of educational objectives "to facilitate communication" (p. 10). The psychological issues are the ordering of the categories to be "consistent with relevant and accepted psychological principles and theories" (p. 6). Over the years, these claims have stimulated a number of investigations to examine their validity, and as the taxonomy is now entering its twenty-second year, this review attempts to survey these investigations and to attempt an overall appraisal of the different findings.

The complete validation of the properties of the taxonomy must go through two stages, the first concerned with the educational properties, the second with the psychological properties.

> A classification scheme may be validated with reference to the criteria of communicability, usefulness, and suggestiveness; while a

> taxonomy must be validated by demonstrating its consistency with the theoretical views in research findings of the field it attempts to order. (Bloom, p. 17)

It will therefore be convenient to consider the educational and psychological issues separately.

Educational Issues

> We are of the opinion that although the objectives and test materials and techniques may be specified in an almost unlimited number of ways, the student behaviors involved in these objectives can be represented by a relatively small number of classes. . . . It is assumed that essentially the same classes of behavior may be observed in the usual range of subject-matter content, at different levels of education (elementary, high school, college), and in different schools. (Bloom, p. 12)

The different categories (i.e., Knowledge, Comprehension, Application, Analysis, Synthesis, Evaluation) were chosen with a view to reflecting the distinctions that, it was felt, teachers already make among student behaviors (Bloom, p. 13). The taxonomy also attempts to clarify these distinctions as accurately as possible with examples of educational objectives and test items for each subclass. However, it is not claimed that the resulting method of classification permits complete and sharp distinctions among behaviors (Bloom, p. 15). Indeed it is suggested that, depending on the nature of the prior learning experiences, different students can solve the same question in different ways, and that one objective or test item can actually be placed in different categories (Bloom, p. 16). It is, therefore, considered essential to know or at least to make some assumptions about a student's prior learning experiences before attempting to assign items or objectives to the different categories (Bloom, p. 51). Nevertheless, within these constraints Bloom (p. 21) considered that the communicability of the taxonomy can be assessed in terms of the extent to which educators can agree in classifying on an individual and independent basis the educational objectives or test items according to the different categories.

Bloom (p. 20) reported that he and his co-workers actually carried out a series of such studies, but he provided no experimental or statistical details. He merely stated that "although we have little difficulty in determining the major class within which a behavior falls, we are still not satisfied that there are enough clearly defined subclassifications to provide adequately for the great variety of objectives we have attempted to classify" (p. 21).

The results of the investigations relating to the main categories are summarized in Table 1, where it is seen that the extent of perfect agreement between the judges varies from one investigation to the next. Unfortunately it is not possible to determine which factor or factors are directly related to these differences. However, it will be noted that in general the extent of the perfect agreement decreases as the number of judges increases. Since the number of possible disagreements between pairs of judges increases as the number of judges increases, it is very

TABLE 1
Summary of the Results of the Investigations into the Reliability of Judges Classifying Test Items to the Taxonomic Categories

	Number of judges	*Number of items*	*Subject matter*	*Percentages perfect agreement*
Scannell & Stellwagen (1960)	2	?	Chemistry	90[a]
Tyler (1966)	2	384	Geography	75
Cox (1965)	3	379	Natural Science	85
Stoker & Kropp (1964)	4	36	Chemistry	31
	4	36	Size relations	31
Herron (1966)	5	83	Chemistry	40
Poole (1971)	6	32	Social	16
Poole (1972)	7	44	studies	14
Fairbrother (1975)	22	40	Physics	7
	22	40	Physics	0

[a]After making allowances for the fact that some items effectively occurred more than once in the total item pool.

plausible that the number of judges is itself an important factor in determining the percentage of the items on which there is perfect agreement.

Fairbrother (1975) gave the number of judges who assigned each item to the different categories, and for each item reported the results of a Kolmogorov-Smirnov test that was carried out to determine whether the frequency distribution of the assignments differed significantly from that which would be expected if the items had been classified purely on a random basis—i.e., equal numbers of assignments to each category. In the two tests there were 17 and 7 items respectively for which the null hypothesis could be rejected at the 1 percent level; at the 10 percent level there were 24 and 9 items respectively.

The only investigation concerned with the ability of judges to assign items to the subcategories was carried out by Stanley and Bolton (1957), who chose their judges from a class of 46 graduate education students who had been studying the taxonomy over a period of four weeks. The top eight students on the basis of Terman's Concept Mastery Test classified 227 items covering a variety of subject-matter areas and educational levels. In a second experiment the top 11 students classified 44 specimen items from a Graduate Record Examination covering social science, natural science, verbal ability, and quantitative ability. In the first experiment the percentages of the items for which there were 8, 7, 6, 5, 4, 3, and 2 agreements respectively were 5, 13, 17, 17, 21, 19, and 8. When only the main categories were considered, the extent of the agreement was considerably better. However, Stanley and Bolton did not provide any statistical details; they merely reported that the agreement was "by no means perfect." For the second experiment the results for each subtest were presented as the average number of agreements per item—i.e., social science, 4.2; humanities, 5.5; natural science, 3.4; verbal ability, 8.0; and quantitative ability, 9.0.

In appraising the results of their respective experiments, the various authors came to rather different conclusions about the reliability of the judge's assignments. Scannell and Stellwagen (1960) concluded that there is a "reasonable degree of accuracy" (p. 13). Stoker and Kropp (1964) stated that their judges "can assign items to the appropriate categories with some accuracy" (p. 40). Stanley and Bolton (1957) pointed to the considerably better agreement in the case of ability as opposed to achievement items but concluded that "there seems to be enough agreement among graduate students independently classifying test items to warrant the regular analysis of teacher-made and standardized tests" (p. 634). On the other hand, Fairbrother (1975) and Poole (1971, 1972) both emphasized the observed general lack of agreement.

As mentioned previously, an obvious and important difference between Fairbrother's experiment and the other experiments was the number of judges taking part in the classification exercise. Also, although not being drawn at random, Fairbrother's judges were almost certainly more representative of a meaningful population of educators (i.e., schoolteachers) than those used in the other experiments. In fact, from the details provided in the actual reports, it is most unlikely that the judges taking part in the other experiments were representative of any population of educators who might wish to communicate with each other about the nature of the educational objectives or items being considered. It is therefore likely that Fairbrother's results and conclusions are the most generalizable to real-life educational contexts.

Psychological Issues

There are two basic psychological issues relevant to the properties of the taxonomy. One concerns the relationship between objectives belonging to different categories within the same subject-matter area. The other concerns the relationship between objectives belonging to the same category in different subject-matter areas.

Within one subject-matter area the ordering of the categories was based on the notion of complexity, and Bloom considered that in these items the categories do have a definite relationship.

As the taxonomy is now organised, it contains six major classes:

1.00 Knowledge
2.00 Comprehension

> 3.00 Application
> 4.00 Analysis
> 5.00 Synthesis
> 6.00 Evaluation
>
> Although it is possible to conceive of these major classes in several different arrangements, the present one appears to us to represent something of the hierarchical order of the different classes of objectives. As we have defined them, the objectives in one class are likely to make use of and be built on the behaviors found in the preceding class in this list. (p. 18)

Bloom then stipulated that the hierarchy has cumulative properties.

> Our attempt to arrange educational behaviors from simple to complex was based on the idea that a particular simple behavior may become integrated with other equally simple behaviors to form a more complex behavior. Thus our classifications may be said to be in the form where behaviors of type A form one class, behaviors of type AB form another class, while behaviors of type ABC form still another class. (p. 18)

In this respect the postulated properties of categories of objectives within one narrow subject-matter area are those of a set of variables that constitute a simplex—i.e., a set of variables that differ only in terms of the degree of their complexity rather than in content (Guttman, 1953).

In describing the relationship between objectives of the same category in different subject-matter areas, Bloom postulated the existence of "arts and skills" which pervade all subject-matter areas.

> "Arts and skills" refer to modes of operation and generalized techniques for dealing with problems. (p. 38)

He also postulated the existence of abilities from the combination of arts or skills with knowledge.

> Arts or skills + knowledge = abilities. (p. 38)

Thus an educational objective would be classified as an ability on the one hand, or an art or skill on the other, according to wheher or not specialized or technical information is required. It also follows that two objectives corresponding to the same categories from different subject-matter areas involve the same set of arts or skills. The objectives would differ only because of differences in the specialized or technical information corresponding to the two subject-matter areas. Thus in Guttman's (1953) terminology objectives or test items from the same taxonomic category in different subject-matter areas constitute a circumplex; i.e., they have the same degree of complexity and differ among themselves only in terms of the content.

The overall structure, as conceived by Bloom, of the psychological relationship between objectives from different categories and different subject matter areas is one in which the objectives differ among themselves simultaneously in both degree of complexity and kind of subject-matter area; in Guttman's terminology they constitute a radex.

The existence of these hypothetical psychological relationships lends itself to direct experimental investigation, since the scores on variables comprising a simplex, circumplex, or radex have definite and characteristic statistical properties. For example, on moving progressively up a simplex from the least complex variable j to the most complex variable n through variables of intermediate complexity k, Guttman has demonstrated that the correlation coefficients r_{jk} etc. have the following relationship

$$r_{jn} = r_{jk} r_{kn} , \tag{1}$$

where $(j < k)$

thus, $$r_{jn} < r_{kn} . \tag{2}$$

Furthermore, if a multiple regression equation is constructed in order to predict the score on variable n from a linear combination of the scores on the lower variables in the simplex,

$$\hat{z}_n = \sum_{i}^{n-1} \beta_i z_i \tag{3}$$

Guttman has demonstrated mathematically that all the β-weights except β_{n-1} must be zero. In other words, none of the indirect relationships between nonadjacent categories should be significant. Finally, if the correlation matrix is subjected to a principal components analysis, the scores on the original variables 1, 2, 3 etc. can be related ultimately to the scores on the factors I, II, III etc. corresponding to intellectual processes and error component e as follows:

$$\begin{aligned} z_1 &= \lambda_{1I} z_I & &+ \lambda_{1e} z_e \\ z_2 &= \lambda_{2I} z_I + \lambda_{2II} z_{II} & &+ \lambda_{2e} z_e \\ z_3 &= \lambda_{3I} z_I + \lambda_{3II} z_{II} + \lambda_{3III} z_{III} & &+ \lambda_{3e} z_e \\ &\text{etc., etc.} \end{aligned} \tag{4}$$

As regards the statistical properties of a set of variables taken from the different simplexes within a radex, the results of a factor analysis will reflect the differences in content between the simplexes as well as the differences in complexity within each simplex. Thus there will be content factors α, β, etc. as well as process factors I, II, etc. For example, in the case of simplexes A and B,

$$\begin{aligned} z_{A1} &= \lambda_{A1I} z_I & &+ \lambda_{A1\alpha} z_\alpha + \lambda_{A1e} z_e \\ z_{A2} &= \lambda_{A2I} z_I + \lambda_{A2II} z_{II} & &+ \lambda_{A2\alpha} z_\alpha + \lambda_{A2e} z_e \\ &\text{etc., etc.} \end{aligned}$$

and (5)

$$\begin{aligned} z_{B1} &= \lambda_{B1I} z_I & &+ \lambda_{B1\beta} z_\beta + \lambda_{B1e} z_e \\ z_{B2} &= \lambda_{B2I} z_I + \lambda_{B2II} z_{II} & &+ \lambda_{B2\beta} z_\beta + \lambda_{B2e} z_e \\ &\text{etc., etc.} \end{aligned}$$

Process factors — Content factors — Error factor

Hence, evidence in favour of the validity of the claims of the taxonomy must demonstrate that the factor structure associated with items from each category in different subject-matter areas is of the form illustrated above.

In considering the observational conditions required to investigate the validity of the claimed properties, Kropp, Stoker, and Bashaw (1966) contended that "it is mandatory to equalize content knowledge over subjects (i.e., students)" (p. 71). In attempting to achieve this equalization they recommended that the experiment should incorporate a specified reading passage or self-instructional programme that the students take before answering the test. If this equalization is not done, Kropp et al. argue that there will be errors in interpreting the results, because "incorrect products [answers to questions] will result not from the use of wrong process but from a lack of subject matter knowledge on which the process should operate" (p. 71). However, the existence of variation per se in content knowledge will merely imply that the scores on the content factors have nonzero variance. It will not prevent the confirmatory cumulative, hierarchical factor pattern from being observed. Hence from this point of view the incorporation of a reading passage is not necessary—even if one could be produced that did equalize content knowledge successfully. Kropp et al. also argued that equality of content knowledge is necessary in order to reduce the probability that different students will use different processes to answer the same questions. This aim is certainly desirable from the point of the experiment, but it is not clear how or why equality of content knowledge across students will necessarily achieve it. For example, it is quite possible that students having equal scores on a particular content factor will use different processes to answer the same question. However, it does seem plausible that the orientating effects—as opposed to any equalizing effects—of a preliminary reading passage or self-instructional programme will predispose the students to attempt to answer a question in a particular way, regardless of whether or not it reduces the variation in content knowledge or not. In fact, only the investigations by Kropp and Stoker (1966) and Stedman (1973) have adopted this procedure. Thus Kropp and Stoker prefaced each test with a reading passage of 600-900 words, and Stedman used a self-instructional programme of 93 frames.

In practice, the various researches into the validity of the properties have not all investigated every claim. At one level the investigations have been concerned with the validity of the postulated cumulative hierarchical relationship between the categories. At another level the investigations have also examined the claim that the processes transcend all subject-matter areas. These different aspects will now be considered separately.

Order of Complexity

Of all the investigations that have studied this particular issue the most notable and comprehensive is that of Kropp and Stoker (1966), who carried out their experiment with students from grades 9, 10, 11, and 12 in ten Florida secondary schools. The experiment used four specially prepared tests entitled "Atomic Structure," "Glaciers," "Lisbon Earthquake," and "Economic Growth." Each test comprised a total of 95 items, of which 20 had been classified to each of the categories Knowledge, Comprehension, Application, and Analysis, 5 to the Synthesis category, and 10 to Evaluation. Within each grade the total number of students taking each test was between 1100 and 1500.

For each test Kropp and Stoker tested the validity of the hierarchy by using a simplex analysis on the correlations computed from composite scores for the subtests corresponding to Bloom's different categories. However, even when the correlations were corrected for attenuation, it was obvious by inspection that the postulated order of categories did not have the pattern of correlations for a pure simplex in any of the 16 matrices covering the different combinations of test and grade. On applying Kaiser's (1962) least squares procedure to determine the order of subtests that most closely fitted the properties of a simplex, in only one of the tests, "Lisbon Earthquake," did this order exactly match that predicted by the taxonomy for all grades. However, for all tests and grades this procedure gave the theoretically expected relative order for the categories Knowledge, Comprehension, Application, and Analysis, and departed from the theory only in terms of the positions of Synthesis and Evaluation. For example, the "Atomic Structure" test placed Synthesis and Evaluation between Knowledge and Comprehension for the ninth grade. The test "Stages of Economic Growth" placed Synthesis between Knowledge and Evaluation in grade 12. The "Glaciers" test gave consistent results at all grades, but placed Synthesis between Knowledge and Comprehension.

In considering the differences consistently found between the observed and theoretically expected positions of Synthesis and Evaluation in three of the tests, Kropp and Stoker (p. 89) pointed to doubts that they originally had about the quality of these particular items and suggested that the use of faulty items may have been the reason why the experimental results did not support the validity of the theory completely. The authors concluded that further work on item construction is needed before the hypothesis can be validly subjected to testing.

Madaus, Woods, and Nuttall (1973) used a causal model approach to analyze the data of Kropp and Stoker for a subsample of 1128 students who had taken the Kit of Reference Tests for Cognitive Factors (French, Ekstrom, and Price, 1963). In this approach the validity of any proposed hierarchy is assessed in terms of the significance of direct and indirect statistical relationships between the various categories, using stepwise multiple regression analysis. The analysis showed that over the four tests and grades the categories from the experiment of Kropp and Stoker failed to comply with the conditions of a cumulative hierarchy because of the significance of eight indirect relationships between nonadjacent categories. After removing the effects of a g factor of general ability, only one of these significant indirect relationships remained. At the same time the direct relationships between Analysis and Synthesis and between Synthesis and Evaluation also lost their significance, but a direct relationship between Analysis and Synthesis retained significance. On the basis of these effects Madaus et al. suggested that the taxonomy had a Y-shaped structure in which the stem of the Y, formed from Knowledge to Comprehension to Application, subsequently divided into one branch of Analysis and another branch from Synthesis to Evaluation. They also made the tentative suggestion that this structure and all its attendant effects imply that only the categories Knowledge, Comprehension, Application, and Analysis measure achievement dependent upon learning and experience, whereas Synthesis and Evaluation measure general ability.

There is also another possible interpretation of all the discrepant results that would still be consistent with the existence of cumulative hierarchical relationships within the taxonomy, even though the analyses performed in the various investigations failed to reveal them. This possibility arises because in any achievement test it is practically certain that all the items belong to different simplexes. In this case the pattern of correlations obtained for pairs of individual items at different levels in different simplexes does not have to follow equation (1). In fact all patterns are possible. Therefore the correlational properties of composite scores—as used by Kropp and Stoker and Madaus et al.—for subtests corresponding to the different categories will not necessarily fulfill the simplex condition summarized by equation (1), even if the items had been taken from the same radex.

Yet another interpretation arises out of the possibility that the judges may have assigned the items incorrectly to the different categories. Unfortunately Kropp and Stoker give no detailed information on the extent to which the judges did actually agree in categorizing the items. However, the results of the earlier study (Stoker and Kropp, 1964) using preliminary versions of two of the tests used in the main experiment, suggested that the judges would have been unanimous in categorizing only 30 percent or so of the items. Moreover, the qualitative description given by Kropp and Stoker of the categorization exercise indicates that in the main study the extent of the agreement was least in assigning items to the Synthesis and Evaluation categories—i.e., those categories that were misplaced according to the conclusions reached in the subsequent investigations. It is therefore quite possible that the apparent deviations from the theoretical order observed by Kropp and Stoker and Madaus et al. could be ascribed to the effects of carrying out the analysis on composite scores for subtests that contain incorrectly assigned items.

Smith (1968, 1970) applied hierarchical syndrome analysis (McQuitty, 1960, 1966) to investigate the existence of the cumulative, hierarchical relationships within the Bloom taxonomy. In this method the analysis proceeds through stages in which correlation coefficients are used to select categories having a succession of common characteristics—the higher the correlation coefficient, the more similar these characteristics are considered to be. The sequence in which the categories of the taxonomy are grouped together is then regarded as reflecting the order within whatever hierarchical classification exists between the categories. For example, Figure 1 is a graphic representation of the results of a hierarchical syndrome analysis which Smith (1968) argued would support the taxonomy rationale, if variable 1 represents Knowledge, variable 2 represents Comprehension, and so forth.

Hierarchical syndrome analysis and simplex analysis can in fact lead to different conclusions when applied to the same correlation matrix.

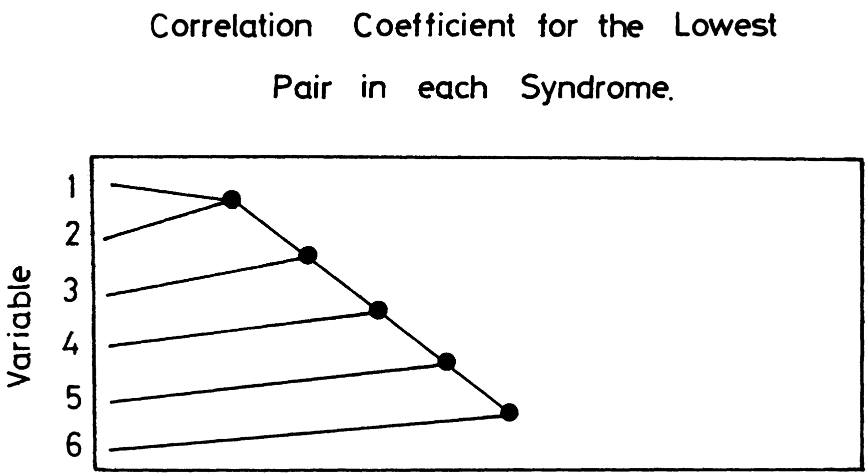

Figure 1
The results of a hierarchical syndrome analysis.

For example, when Smith (1968) applied hierarchical syndrome analysis to the correlation matrices of Kropp and Stoker (1966), the resulting order of complexity consistently placed the Knowledge category in a different position from that obtained in simplex analysis. Moreover, when simplex analysis would support the existence of a cumulative hierarchy, hierarchical syndrome analysis would not necessarily do so. In particular, if simplex analysis were applied to the correlation matrix in Table 2, it would support the existence of a cumulative hierarchical relationship between the variables 1 - 6, whereas Figure 2 shows that hierarchical syndrome analysis would not. Conversely, in Figure 1, which shows the results of a hierarchical syndrome analysis supporting a cumulative hierarchical relationship, the relationship between categories 1, 2, 3 is such that r_{12} is greater than r_{13}, and r_{12} is greater than r_{23}. However, this same graph would have been obtained no matter which of the two correlations r_{13} and r_{23} were the greater. On the other hand simplex analysis would demand that r_{13} be less than r_{23} before it could be concluded that the order given for variables 1, 2, 3 represented a cumulative hierarchy. Now since the cumulative hierarchical relationships postulated within the Bloom Taxonomy are couched unambiguously in terms of a simplex, any alternative method of statistical analysis that does not give rise to the same conclusions as simplex analysis cannot be appropriate for investigating the existence of these properties. Hence the results of Smith's (1968, 1970) investigations cannot provide evidence for or against the validity of these properties within the Taxonomy.

Several investigations (Bloom, Englehart, Furst, Hill, & Krathwohl, 1956, pp. 18-19; Stoker & Kropp, 1964; Kropp & Stoker, 1966; Stedman,

TABLE 2
The Correlational and Factorial Properties of a Typical Simplex

Correlation Matrix

Variable	*Variable* 1	2	3	4	5	6
1	1.00	0.76	0.61	0.50	0.41	0.33
2		1.00	0.80	0.65	0.53	0.43
3			1.00	0.82	0.67	0.53
4				1.00	0.82	0.65
5					1.00	0.81
6						1.00

Loadings of Principal Components

Variable	*Component* I	II	III	IV	V	VI
1	0.71	0.56	−0.35	−0.20	0.90	−0.03
2	0.84	0.43	0.02	0.23	−0.21	0.09
3	0.90	0.15	0.31	0.12	0.17	−0.18
4	0.90	−0.17	0.26	−0.20	0.06	0.21
5	0.85	−0.41	−0.04	−0.16	−0.21	−0.16
6	0.75	−0.52	−0.33	0.20	0.12	0.06

Loadings of Rotated Components

Variable	*Factor* I	II	III	IV	V	VI
1	1.00	0.00	0.00	0.00	0.00	0.00
2	0.76	0.63	0.00	0.00	0.00	0.00
3	0.61	0.55	0.64	0.00	0.00	0.00
4	0.49	0.42	0.49	0.58	0.00	0.00
5	0.40	0.33	0.39	0.48	0.59	0.00
6	0.32	0.27	0.28	0.40	0.47	0.60

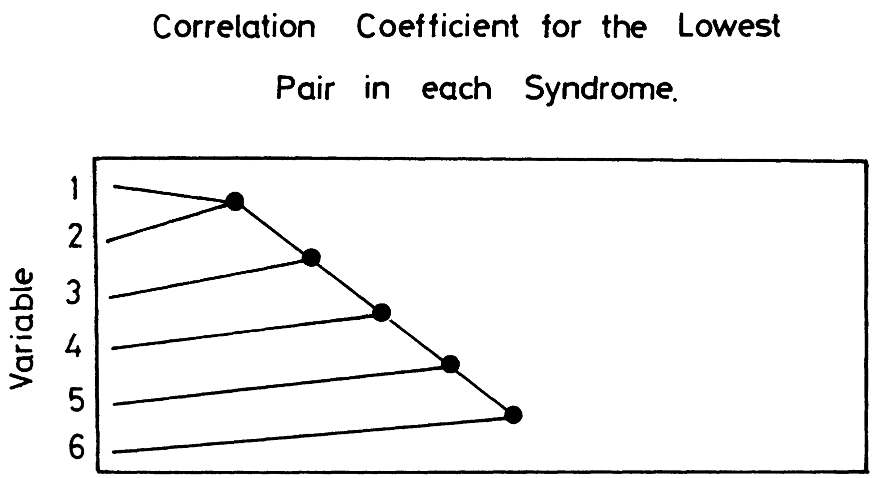

Figure 2
The results of a hierarchical syndrome analysis for the correlation matrix in Table 2.

1973; Fairbrother, 1975) have analyzed trends in means of items in the different categories in an attempt to observe evidence of the cumulative hierarchical relationships. Thus it is assumed, if the postulated hierarchical relationship between the categories is valid, that the facility (difficulty) of the test items corresponding to the different categories should decrease on progressing upwards from Knowledge through the intermediate categories to Evaluation. Unfortunately the rationale underlying this method is misconceived since it confuses the two distinct concepts facility and complexity. As described previously, complexity is reflected in trends in correlation coefficients, and it is facility, not complexity, that is reflected in trends in means. The distinction is important because, as Guttman (1953) demonstrated, facility and complexity are not necessarily correlated. Also there is empirical evidence to show that in practice there may be very little correlation complexity and facility. Thus Crawford (1966) investigated this relationship for eight medical examinations and demonstrated that the order of difficulty for various subtests within each examination was in every case except one statistically different from the order of complexity as determined by simplex analysis. Kropp and Stoker (1966) argued that in their experiment item complexity and item difficulty must be positively correlated as a result of their attempts to equalize knowledge of content for all students. However, even if this equalization were actually achieved, it would ensure that the variance of only the content factor scores were zero, and as Guttman (1953) demonstrated, in these circumstances it is still possible for the more complex tasks to be either more or less difficult than the less complex tasks. It must be concluded, therefore, that investigations that examine trends only in mean scores of items or subtests provide no evidence as to the existence of a cumulative hierarchical relationship between the categories.

Poole (1971, 1972) sought to distinguish between complexity and difficulty by measuring item complexity as the mean of the normal ogive fitted to the item-score-test-score regression data. However, as Lord and Novick (1968, p. 366) emphasized, the normal ogive model is intended for use with items having only one latent trait, and since a measure of complexity must reflect different numbers of latent traits, the mean of the normal ogive cannot be a measure of complexity. Poole was therefore attempting to investigate the validity of a theory that is fundamentally inconsistent with the principles behind the techniques he employed. Hence the use of this technique is not legitimate for the investigation.

Transcendence of the Processes

Investigations relevant to the postulated transcendence of the processes have been carried out by Poole (1971, 1972), Kropp and Stoker (1966) as well as by Stoker and Kropp (1971). Poole's investigations concerned the transcendence of the processes within single subject-matter areas such as social sciences, whereas Kropp and Stoker as well as Stoker and Kropp looked for transcendence across four recognizably different subject-matter areas.

Poole applied multiple group factor analysis (Harman, 1960) to the items within each of two tests entitled "Social Change" and "Freedom and Equality," respectively. If the claimed properties of the taxonomy were valid, Poole expected to find that each group of items would form its own factor and correlate most highly with it. While Poole did not provide any statistical details, he stated that the obtained factor structure did not fit this pattern.

Kropp and Stoker (1966) applied principal components analysis to the results of those students who had taken each of their four tests, previously described. Their aim was to see if it were possible to isolate process factors that characterized each category across all four tests in the cumulative hierarchical pattern depicted by equations (5). Within each grade, the analysis was based on a 24 x 24 correlation matrix in which each variable was one subtest corresponding to a different combination of Bloom category and type of test. In obtaining this matrix, Kropp and Stoker found that the number of examinees contributing to each correlation coefficient varied, but in all cases the number exceeded 750. In each case the final interpretation was carried out on the unrotated matrix for the first six principal components.

Over the four grades, the analysis produced a general factor having positive loadings on all the variables, together with bipolar factors that were interpreted as being pure process factors operating across content, pure content factors operating across processes, or mixed process-content factors operating within a particular combination of process and content. The factor structure differed from that which was expected in that the process factors did not have the cumulative hierarchical pattern illustrated in equations (5). Also, they were usually associated with more than one Bloom category. For example, in each of these factors, two or more of the subtests, Knowledge, Synthesis and Evaluation, were always in opposition to Comprehension, Application, or Analysis. Kropp and Stoker concluded that both the existence and consistent nature of factors, which are primarily determined by processes, offer some support for the transcendence hypothesis (p. 103). However, they also felt that the presence of the joint process-content factors tended to discredit it (p. 109).

In appraising the significance of these factor analytic studies, it must be borne in mind that in each case the computational algorithms were predetermined solely by mathematical considerations, and not by the nature of the psychological theory that was being investigated. In such cases unrotated factors can obscure the validity of a theory even when the raw data fit it. For example, Table 2 shows that the unrotated principal components of a simplex do not have the cumulative hierarchical pattern. However, as Table 2 also shows, when all possible pairs of principal components are suitably rotated the required factor pattern of a simplex can be obtained. Consequently, it would have been rather surprising if the unrotated factors produced by Poole (1971, 1972) and Kropp and Stoker (1966) had been of such a pattern as to provide immediate support for the theory. By the same token, the existence of a bipolar factor consistently associated with a particular set of categories across content does not, as Kropp and Stoker suggest, provide any support for the theory, unless these factors can be rotated to produce process factors each of which is associated with one category across all content. In conclusion, the results of the factor analyses described above provide neither confirmatory nor disconfirmatory evidence for the theory.

Kropp and Stoker (1966, 109—110) were evidently well aware of the difficulties of using unrotated factors to investigate the validity of the transcendence hypothesis. Consequently, as an alternative method of analysis they examined the correlation matrices containing subtests corresponding to the same categories from all tests to see if they had the properties expected for an equally spaced perfect circumplex—a special form of circumplex in which the totals of the columns in the correlation matrix are equal (Guttman, 1953, pp. 328-329). Over each grade and category, none of the matrices conformed exactly to these properties although many of them approximated very closely. Kropp and Stoker implied that the observed deviations may be due to sampling errors. They also pointed out that in these circumstances it is difficult to provide convincing evidence as to the existence of an equally spaced perfect circumplex, since its sampling properties are not known. Nevertheless, Kropp and Stoker concluded that their results supported the transcendence hypothesis to some extent (p. 109).

Stoker and Kropp (1971) subsequently applied Guttman-Lingoes smallest space analysis (Guttman, 1968; Lingoes, 1965) to their own correlation matrices for all the subtests from the four main tests. The end product of these analyses was a set of graphs in which the subtests are represented as points in two-dimensional space, and the distance between any two points reflects the extent of the correlation between the two corresponding subtests.

Figure 3 illustrates the important features that Stoker and Kropp reported as being common to the graphs for all grades whether they are

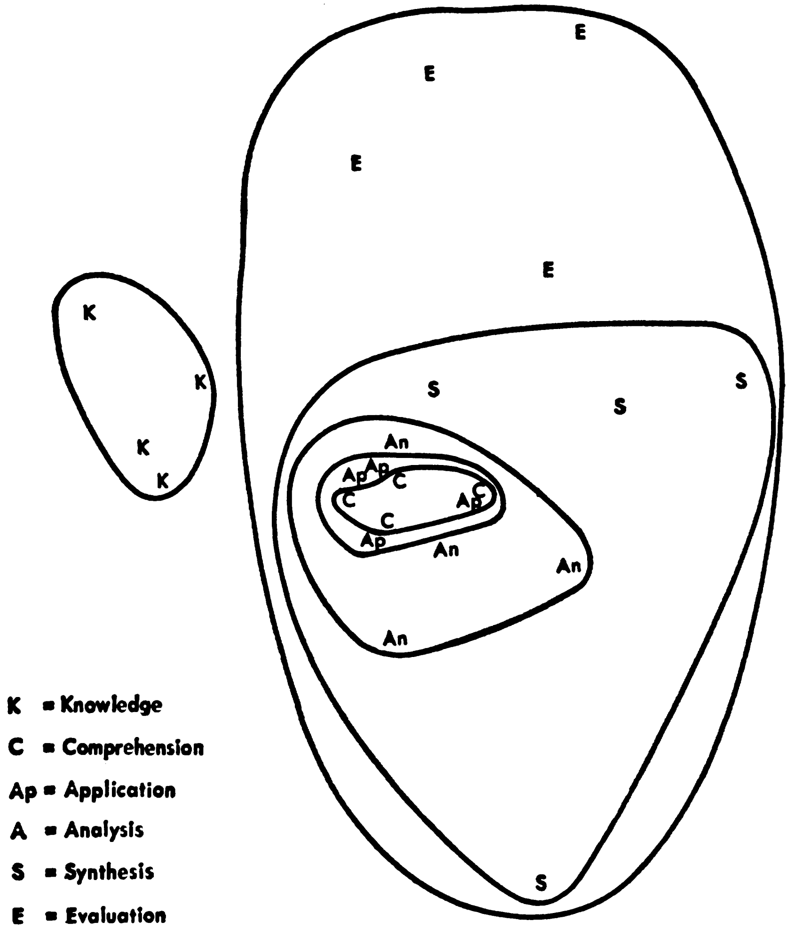

Figure 3

The results of a Guttman-Lingoes smallest space analysis—1 obtained by Stoker and Kropp (1971). (Reproduced by permission).

considered separately or in combination. It will be noted that the boundary lines are drawn by Stoker and Kropp emphasize the arrangement and clustering of points for the same and different categories. They show that, on moving from the centre of the graph, the order of the regions corresponding to Comprehension, Application, Analysis, Synthesis, and Evaluation follows the theoretical order of complexity. Stoker and Kropp concluded that these graphs provide some evidence for a radex, and in the case of Figure 3 rationalize the apparent misplacement of the Knowledge category in terms of the spurious correlational effects arising out of their attempts to devise items that

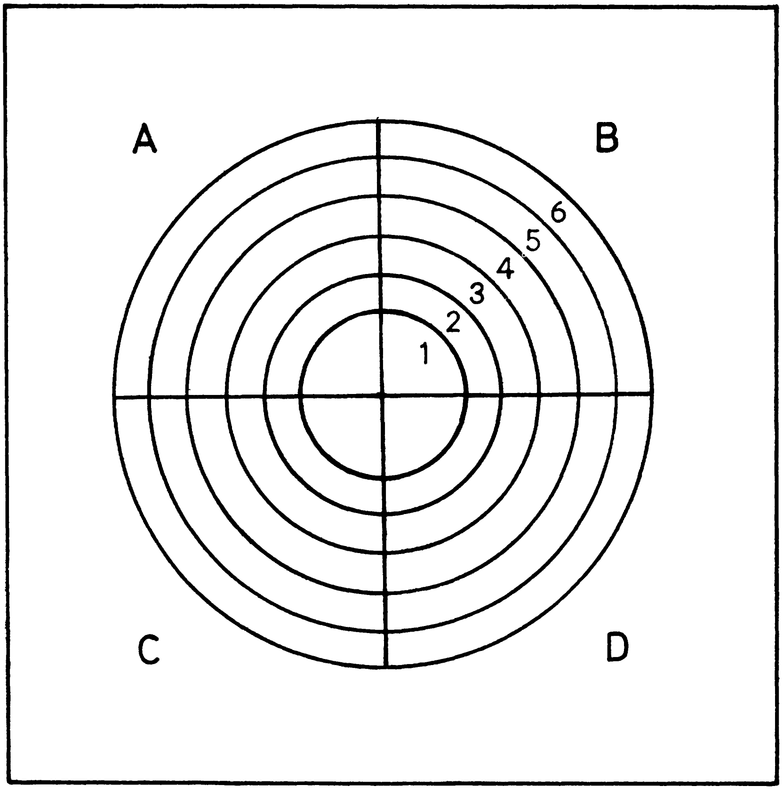

Figure 4
Schematic representation of the results of a Guttman-Lingoes smallest space analysis applied to a radex.

maximize the number of individuals receiving a perfect score on this category.

The assumption implicit in making this interpretation is that the differences in process and differences in content are represented respectively by the radial and angular coordinates that describe a point's position in relation to the centre of the graph. Hence, to be consistent with this assumption, a pure radex should result in a graph that has the features of the schematic diagram illustrated in Figure 4. The points corresponding to different categories, 1, 2, 3, etc., are situated in annular-shaped regions arranged in an order that reflects the relative complexities of the categories, and the points corresponding to different subject-matter areas A, B, C, etc. are in different positions around each annulus.

Figure 3 shows that in the study by Stoker and Kropp the categories Evaluation and Synthesis, as well as the categories Synthesis and

Analysis, are clearly separated by this analysis. However, despite the manner in which the boundary lines are drawn in Figure 3, it is obvious that in the case of the subtests corresponding to Comprehension, Application, and Analysis, there are a significant number of deviations from the arrangement expected for a perfect radex. Certainly the radex pattern is not nearly as clear as that which has been observed in other investigations (Guttman, 1965; Schlesinger & Guttman, 1969). For example, any perfect circle that is drawn to encompass all the points corresponding to Comprehension subtests will simultaneously enclose all the points corresponding to the Application subtests and two out of the four points corresponding to the Analysis subtests. Hence, even if Stoker and Kropp's rationalization about the misplacement of the Knowledge category is accepted, there are still too many deviations from the pure radex pattern to give strong support to the theory of cumulative hierarchical relationships and transcendence. Even the rationalization itself is most questionable, since the results of the simplex and multiple regression analyses described previously indicated that the Knowledge category was always correctly placed within the different simplexes.

The results of the Guttman-Lingoes smallest space analysis also differ from those of the simplex and multiple regression analyses as regards determining the relative complexities of the categories Comprehension, Application, and Analysis. Whereas the former method is unable to make a very clear separation among these categories, the studies by Kropp and Stoker (1966) and Madaus et al. (1973) placed these categories within the simplexes in the relative order of complexity predicted by the theory. Since there is no doubt that the use of simplex and multiple regression analyses is theoretically correct for the investigation, the Guttman-Lingoes smallest space analysis should reveal the same order of complexity for these categories. It may therefore be that the discrepant results obtained with the Guttman-Lingoes smallest space analysis imply that there are some problems associated with its use in this case.

Conclusions

In making an overall assessment of the validity of Bloom's claims, it is important to remember that the claims concerning the psychological properties of the different categories presuppose that objectives and test items can be classified correctly in the first place. Thus the psychological issues can be assessed only after a consideration of the educational issues.

Educational Issues

If, as was suggested previously, the results of Fairbrother (1975) are those that relate most closely to a practical educational setting, they are a most disturbing reflection of the extent to which the taxonomy facilitates communication. The rather sanguine conclusions reached by some of the other workers are therefore most misleading.

However, it is probable that experiments that set out to evaluate the usefulness of the taxonomy in such absolute terms are not those which are going to be of most practical use. It seems unlikely that any single taxonomy will ever provide a means of universal perfect understanding with any population of educators. Hence it would seem more relevant to try to find out which taxonomy is the best, although imperfect, means of communication. It is therefore suggested that further experiments should compare the extent of the agreement between judges using Bloom's taxonomy with that obtained using other taxonomies, e.g., Gagné (1965), Ebel (1965), Walbesser (1965), Merrill (1971).

Psychological Issues

In considering the experiments investigating the psychological issues, much of the discussion concerned problems in the choice or application of various statistical procedures. On the question of the existence of a cumulative hierarchy, several of the techniques used were found to be inappropriate, and the conclusions that had been based on these techniques were therefore considered to be invalid. On the question of the transcendence of the processes, all the statistical techniques used were considered to be appropriate in theory, but in each case it appears that not enough is known about how these techniques should be applied in practice. Hence none of the experiments investigating this issue provides a definite answer.

The most meaningful conclusions concerning the cumulative hierarchy are to be derived from the simplex analysis of Kropp and Stoker (1966) and the multiple regression analyses of Madaus et al. (1973). Both investigations concluded that in Bloom's theory the categories Knowledge, Comprehension, Application, and Analysis were in the correct relative order, but that Synthesis and Evaluation were not correctly placed. In the study by Kropp and Stoker, the conclusions concerning the relative order of complexity for the categories Knowledge, Comprehension, Application, and Analysis were based on statistical tests designed to determine the order of the categories that best fitted the characteristics of a simplex. Since in practice these tests will actually produce a "best" order from correlation matrices that have no simplex properties, the tests cannot be used to prove that a simplex actually exists. Such proof can be obtained only from a visual inspection of the correlation matrices. For the matrices of Kropp and Stoker, it is clear that in 5 out of 16 cases the correlation coefficients for the categories Knowledge, Comprehension, Application, and Analysis do not have the required properties. The observed deviations may be due to sampling fluctuations, but in the absence of suitable significance tests it is necessary to exercise due caution in drawing conclusions regarding the simplex properties of these four categories. Hence the conclusion of Kropp and Stoker must be seen as an approximate overall generalization that contains exceptions. Similarly, the conclusions of Madaus et al. relating to the same four categories were based on average rather than consistent effects observed for all tests and all grades, and if these important qualifications are not stressed, the conclusions once again could be misleading. Notwithstanding these points concerning the order of the first four categories of Bloom's taxonomy, the suggestion by Madaus et al. that the taxonomy has a Y-shaped structure is a very interesting interpretation of the anomalous pattern of correlation coefficients generally observed for Synthesis and Evaluation.

As a final assessment of the validity of the claims concerning the psychological properties of the taxonomy, it is perhaps fairest to say that the picture is uncertain. No one has been able to demonstrate that these properties do not exist. Conversely no one has been able to demonstrate that they do. The strongest supportive evidence concerns the cumulative hierarchical relationship between the categories Knowledge, Comprehension, Application, and Analysis. However, the evidence here is by no means conclusive, and it should be used merely as an encouragement of or justification for the carrying out of further experiments that are designed to avoid all the mistakes and difficulties encountered in the previous experiments. In particular, in view of the problems in interpreting the results of analyses performed on composite scores for various subtests, it is suggested that they could all be avoided by applying techniques such as factor analysis and Guttman-Lingoes smallest space analysis to the correlation matrices of individual items. The items should still give rise to the effects indicative of a cumulative hierarchy even though they may not have been drawn from the same simplex. In this way it should also be possible to overcome the effects of judges having misclassified the items, because the analysis should be able to reveal those items that have been misclassified. The use of this approach may well give rise to its own attendant difficulties, but there seems to be no reason why they cannot be overcome. It can be hoped that the solutions will be found before another twenty-two years elapse!

Reference Notes

1. Cox, R. C., & Unks, N.J. *A selected and annotated bibliography of studies concerning the "Taxonomy of educational objectives: Cognitive domain"* (Working Paper 13). Pittsburgh: University of Pittsburgh, Learning Research and Development Center, June 1967.
2. Cox, R. C., & Wildemann, C. E. *Taxonomy of educational objectives: Cognitive domain. An annotated bibliography* (Monograph 1). Pittsburgh: University of Pittsburgh, Learning Research and Development Center, March 1970.

References

Bloom, B. S., Englehart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. *Taxonomy of educational objectives. Handbook 1: Cognitive domain.* New York: Longmans, Green, 1956.

Cox, R. C. Item selection techniques and evaluation of instructional objectives. *Journal of Educational Measurement*, 1965, *2*, 181-185.

Crawford, W. R. Item difficulty as related to the complexity of intellectual processes. *Journal of Educational Measurement*, 1968, *5*, 103-107.

Ebel, R. L. *Measuring educational achievement.* Englewood Cliffs, N. J.: Prentice Hall, 1965.

Fairbrother, R. W. The reliability of teachers' judgements of the abilities being tested by multiple choice items. *Educational Research*, 1975, *17*, 202-210.

French, J. W., Ekstrom, R. B., & Price, L. A. *Manual: Kit of reference tests for cognitive factors.* Princeton: Educational Testing Service, 1963.

Gagné, R. M. *The conditions of learning.* New York: Holt, Rinehart Winston, 1965.

Guttman, L. Image theory for the structure of quantitative variates. *Psychometrica*, 1953, *18*(4), 277-296.

Guttman, L. A new approach to factor analysis: The radex. In P. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences.* Glencoe, Ill.: The Free Press, 1964.

Guttman, L. The structure of interrelations among intelligence tests. *Proceedings of the 1964 invitational conference on testing problems.* Princeton, N. J.: Educational Testing Service, 1965.

Guttman, L. A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrica*, 1968, *33*, 469-506.

Harman, H. H. *Modern factor analysis.* Chicago: University of Chicago Press, 1960.

Herron, J. D. Evaluation and the new curricula. *Journal of Research in Science Teaching*, 1966, *4*, 159-170.

Institute for Scientific Information. *Social sciences citation index.* Philadelphia: Author, 1970-1976.

Kaiser, H. F. Scaling a simplex. *Psychometrica*, 1962, *27*, 155-162.

Kropp, R. P., & Stoker, H. W. *The construction and validation of tests of the cognitive processes described in the "Taxonomy of educational objectives"* (Cooperative Research Project No. 2117). U.S. Office of Education, 1966.

Kropp, R. P., Stoker, H. W., & Bashaw, W. L. The validation of the "Taxonomy of educational objectives." *Journal of Experimental Education*, 1966, *34*, 69-76.

Lingoes, J. C. An IBM 7090 program for Guttman-Lingoes smallest space analysis—1. *Behavioral Science*, 1965, *10*, 183-184.

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

Madaus, G. F., Woods, E. M., & Nuttall, R. L. A causal model analysis of Bloom's taxonomy. *American Educational Research Journal*, 1973, *10*, 253-262.

McQuitty, L. L. Hierarchical syndrome analysis. *Educational and Psychological Measurement*, 1960, *20*, 293-304.

McQuitty, L. L. Improved hierarchical syndrome analysis of discrete and continuous data. *Educational and Psychological Measurement*, 1966, *26*, 577-582.

Merrill, M. D. Necessary psychological conditions for defining instructional outcomes. *Educational Technology*, 1971, *11*(8), 34-39.

Poole, R. L. Characteristics of the "Taxonomy of educational objectives: Cognitive domain." *Psychology in the Schools*, 1971, *8*, 379-385.

Poole, R. L. Characteristics of the "Taxonomy of educational objectives: Cognitive domain"—A replication. *Psychology in the Schools*, 1972, *9*, 83-88.

Scannell, D. P., & Stellwagen, W. R. Teaching and testing for degrees of understanding. *California Journal of Instructional Improvement*, 1960, *3*(1), 8-14.

Schlesinger, I. M., & Guttman, L. Smallest space analysis of intelligence and achievement tests. *Psychological Bulletin*, 1969, *71*, 95-100.

Smith, R. B. An empirical examination of the assumptions underlying the "Taxonomy of educational objectives: Cognitive domain." *Journal of Educational Measurement*, 1968, *5*, 125-128.

Smith, R. B. An empirical investigation of complexity and process in multiple choice items. *Journal of Educational Measurement*, 1970, *7*, 33-41.

Stanley, J. C., & Bolton, D. T. A review of Bloom's "Taxonomy of educational objectives" and J. R. Gerberich's "Specimen objective test items, A guide to achievement test construction." *Educational and Psychological Measurement*, 1957, *17*, 631-634.

Stedman, C. H. An analysis of the assumptions underlying the "Taxonomy of educational objectives: Cognitive domain." *Journal of Research in Science Teaching*, 1973, *10*, 235-241.

Stoker, H. W., & Kropp, R. P. Measurement of cognitive processes. *Journal of Educational Measurement*, 1964, *1*, 39-42.

Stoker, H. W., & Kropp, R. P. An empirical validity study of the assumptions underlying the structure of cognitive processes using Guttman-Lingoes smallest space analysis. *Educational and Psychological Measurement*, 1971, *31*, 469-473.

Tyler, L. L. The taxonomy of educational objectives: Cognitive domain. Its use in evaluating programmed instruction. *California Journal of Educational Research*, 1966, *27*, 26-32.

Walbesser, H. H. *An evaluation model and its application.* Washington: American Association for the Advancement of Science, 1965.

AUTHOR

G. MALCOLM SEDDON *Address:* Chemical Education Sector, School of Chemical Sciences, University of East Anglia, Norwich, England. *Title:* Lecturer. *Degrees:* B.Sc., M.A., Oxford; Ph.D. Birmingham. *Specialization:* Educational measurement.