The Irish Journal of Education, 1968, II, 2, pp 127-135

FORMATIVE EVALUATION INSTRUMENTS

PETER W AIRASIAN
Boston College

Formative evaluation provides a means for using test procedures to guide and foster learning This use of evaluation techniques represents a departure from the typical use of evaluation to judge or grade teaching and learning The paper discusses the use of hierarchical structures of tasks, tested and scored in terms of item response patterns, to provide information to the teacher, learner, and curriculum constructor regarding inadequacies in the instructional context

Ralph Tyler's (19, 20) pioneer work in the areas of curriculum and instruction is the basis for most current evaluation theory For Tyler education is a process designed to result in changes in students There is much to be gained from explicitly stating these changes in terms of what students are expected to be able to do at the end of a course Teaching, then, is the process of fostering development in the direction of the desired changes The function of evaluation is to determine the extent to which students have changed in the desired manner

Perhaps because of its close link with Tyler's theory of curriculum planning and instruction, evaluation today is seen primarily as an aid for making judgments about students, about teachers and about the curriculum at the conclusion of a lengthy instructional period (3, 5) Tyler's formulation of the role of evaluation, however, is not limited to judgments which come at the end of a course End of course or end of semester evaluations do perform a number of functions which are central to the educative process They enable one to differentiate passing from failing students, good from bad teachers, and suitable from unsuitable curricula Such evaluation, however, coming at the conclusion of a course and used primarily to aid in judging or grading, seldom contributes to learning by providing the student, teacher, or curriculum constructor with specific information about inadequacies in teaching or learning

For several reasons, the usual evaluation instrument cannot provide the specific, usable information which would help students and teachers with their everyday work Because it evaluates learning over a lengthy period, it does not identify many of the learner's difficulties at a time when knowledge of such difficulties is most crucial The learner needs to be informed of his inadequacies at a time when he can still do something about them The timing of evaluation and the identification of

failures are especially important in those subjects in which content introduced in the early stages of instruction forms the basis for later learning Moreover an evaluation instrument constructed to test a lengthy instructional period can only sample the skills and abilities which form the objectives of the course Such instruments, because of time limitations examine the content taught in a global manner, and so they cannot reveal all the student's failings Because all relevant content and skills are not evaluated, specific remedies cannot be prescribed It would seem that evaluation can and should perform a more vital role in fostering desired student changes

SUMMATIVE AND FORMATIVE EVALUATION

Scriven (18) introduced the terms summative and formative evaluation in his discussion of curriculum evaluation Summative evaluation is that type of evaluation which is designed to yield terminal judgments about a curriculum as a whole A question such as 'Is this curriculum better than an alternative curriculum?' is the type of question which summative evaluation is intended to answer Formative evaluation on the other hand, is designed to generate information about the adequacy of specific subunits learning materials, and instructional sequences, while teaching is in progress Cronbach has stated

> To be influential in course improvement, evidence must become available midway in curriculum development not in the home stretch when the developer is naturally reluctant to tear open a supposedly finished body of materials and techniques Evaluation used to improve the course while it is still fluid contributes more to education than evaluation used to appraise a product already placed on the market (4, p 675)

The approach to curriculum evaluation described by Cronbach is formative in so far as it advocates frequent, on the spot, feedback which the curriculum constructor can employ to identify areas of weakness in his curriculum while it is still in the developmental stage Thus for Scriven and Cronbach it is the time differential the difference between evaluation used to aid development and evaluation used to judge a finished product, which differentiates formative from summative evaluation

Formative evaluation, if it is to be integrated into the teachmg-learning process should go beyond curriculum building to include appraisal of student learning and teaching effectiveness Instruments should be constructed to reveal where precisely a student failed Having

identified and differentiated what a student had mastered and what he had not, steps can be taken to correct each individual's deficiencies

If Scriven's definition of formative evaluation is broadened to include the learner and teacher, some consequences follow for the construction and scoring of formative evaluation instruments In order that formative evaluation devices should yield the detailed, usable information which can aid the student teacner and curriculum constructor, the item sampling approach used in the construction of summative evaluation instruments must be altered It is not enough to sample the relevant skills and abilities taught in a learning unit, the formative test must examine a student in each relevant skill and ability

Formative evaluation procedures might be made more valuable if they were to reveal direct relationships between the tasks to be learned Instead of simply showing whether or not students had mastered what they had been required to learn, formative evaluation could be designed and used to establish psychological hierarchies among tasks If, for example, mastery of A is prerequisite to mastery of B, which in turn is prerequisite to mastery of C, such a hierarchy could greatly assist the curriculum constructor and also guide the teacher by identifying particular learning difficulties However, if a formative test is to yield such results it must be constructed and marked in a manner different than the usual evaluation instrument All tasks in the hierarchy must be tested Performance must be assessed in terms of item response patterns so that the relationships betwen items in the hierarchy will be evident

INSTRUMENT CONSTRUCTION

To construct formative evaluation instruments, a number of components are needed First, all of the tasks involved in the learning unit under study must be identified and specified The tasks identified should be exhaustive of the unit to be formatively evaluated since all pertinent tasks in the unit are to be tested A set of rules must be developed which will arrange the tasks in a hypothesized hierarchical order That is, a set of rules must be found which will permit independent judges to assign given tasks to the same hierarchical level Further, a model must be developed to analyze the results of formative evaluation instruments The model should be flexible enough to permit investigators to examine the response patterns of both individual students and groups of students drawn up taxonomies which serve to characterize categories of intellectual functioning In both taxonomies, the categories are arranged in hierarchies Gagne's taxonomy is constructed from the point of view

Specification of components in a learning task

The first requisite for the construction of formative evaluation instruments involves the determination of a set of procedures which permit elaboration of the instructional specifications These specifications serve as a basis for item construction Proponents of task description for programmed instruction provide useful hints about a set of procedures which might produce educational specifications with the desired characteristics The programmer, who must prepare numerous frames to teach a single terminal objective, must analyze content and instruction into finer detail than that recommended by Tyler and other evaluation theorists Gagne (7) lists three occasions when a detailed analysis of a terminal objective is necessary (i) when one is designing a curriculum, (ii) when one is attempting to assess individual student progress, and (iii) when one wishes to understand the conditions under which learning does or does not occur Several other authors have also stressed the necessity of analyzing globally-stated objectives into a series of more specific component objectives if one is adequately to identify strengths and weaknesses in learning teaching and the curriculum (13, 14, 15, 21)

Relationship between components in a learning task

It is, however, one matter to identify component tasks as disparate elements, and quite another to specify relationships between them Formative evaluation procedures could be made more powerful if, in addition to task analysis of terminal objectives into a series of component tasks, some scheme could be devised for categorizing the relationships between the component tasks Numerous researchers have proposed sets of descriptive categories which can serve to group tasks or behaviours manifesting similar characteristics into classes (9, 12, 15, 16) 'But what various authors have attempted to show, is that there seem to be classes of behaviour, the members of which have a *formal identity*, irrespective of their particular content These classes of behaviours can be defined as performances (that is, as objectives) and distinguished from each other' (7, p 41)

Lindvall (14, p 39) added another dimension to the issue of task categories when he wrote 'When one sets out to identify capabilities, the suggestion made by the evidence, early in the game, is that these capabilities are arranged in a hierarchy ' Bloom (2) and Gagne (7) have of the conditions of learning which produce terminal task mastery, Bloom's taxonomy is based on the complexity of intellectual functioning itself

Formative evaluation procedures could be further enhanced if the test constructor could hypothesize direct relationships and dependencies

between a specific task to be learned at one hierarchical level and another task to be learned at a different hierarchical level Formative evaluation under such conditions could do more than simply show which component tasks were unmastered by a given student It could begin to provide information about the relationship between the unmastered and mastered tasks The pattern of item responses could show how failure on higher level tasks was related to performance on lower level prerequisite tasks

The structure of a learning task

To describe the component tasks in a unit of learning, to organize the tasks into a hierarchy (either in terms of the conditions of learning or in terms of intellectual complexity), and to specify relationships between the tasks at one hierarchical level and those at another hierarchical level would produce a structure for a unit of learning A schema of a simplified structure is presented in Figure 1 Task 1, at the lowest hierarchical level, might involve remembering a definition Task 1 can be considered

FIGURE 1
LINEAR HIERARCHICAL STRUCTURE

a necessary but not sufficient prerequisite for mastering 2, knowing a rule If the dependency of 2 upon 1, indicated by the connecting line, is justified, then a student could not master 2 without having mastered 1 Similarly, mastery of task 3, application of a rule, would be dependent upon mastery of tasks 1 and 2 In order to test whether the pattern of responses supported the hierarchy, an item testing each of the three tasks would have to be included in the formative test

TABLE 1
POSSIBLE ITEM RESPONSE PATTERNS TO THREE ITEM HIERARCHY

Task 1	Task 2	Task 3
0	0	0
1	0	0
0	1	0
0	0	1
1	1	0
1	0	1
0	1	1
1	1	1

For any three dichotomously scored items, there are $2^3 = 8$ possible response patterns With 1 corresponding to task mastery and 0 to lack of task mastery, the eight response patterns for the tasks in Figure 1 are set out in Table 1

On the assumption that satisfactory performance at each of the lower levels is a necessary but not sufficient condition for satisfactory performance at the higher ones, the only patterns possible are 000, 100 110, and 111

Several investigators (6, 10 11, 17) analyzed such units as finding the sum of n integers and solving linear algebraic equations into hierarchies of subordinate tasks, and indicated the sequence in which the subordinate tasks were to be taught To control the teacher variable they arranged for the sequences to be learned by means of programmed instruction devices The findings showed that instances of students passing a later task in the sequence after failing an earlier task were rare

Summing up the studies, Gagne and Bassler concluded

> In these studies, the factor of *topic sequence* appears prominently as one which exerts considerable control over the process of learning In simple terms, it has been verified in these studies that the capability of performing certain identifiable units of subordinate knowledge creates a high probability of acquiring a new item or knowledge, whereas the lack of the capability of performing any one of these same subordinate units reduces the probability to very low values (8)

Airasian (1) analyzed existing learning units in algebra and chemistry into hierarchies of tasks according to Bloom's categories and postulated necessary but not sufficient conditions for mastery between lower and higher level tasks The conditions of learning were not controlled by means of programmed instruction The individual classroom teacher structured the instruction as he pleased On formative evaluation tests which included items at different hierarchical levels, 85 to 100 per cent of the student response patterns were of the predicted type lower level responses were prerequisite to higher level ones This finding was generalizable across subject areas and teachers

Research indicates, therefore, that there are important hierarchical relationships among the elements into which a unit of learning can be resolved, and that these relationships generalize across most students This is so whether the hierarchies are specified in Bloom's or Gagne's terms

THE USE OF FORMATIVE EVALUATION INSTRUMENTS

What kind of information then can formative evaluation instruments, incorporating a hierarchical structure and testing each task in that structure, provide? The curriculum constructor can study small segments of his product from the point of view of posited relationships between the tasks to be taught He can determine the extent to which he has incorporated appropriate transfer devices between related tasks He can if he is operating in some mode of programmed instruction determine the logical sequence for presentation If his concern is with a textbook for classroom use he can determine the extent to which higher level activities, that is activities other than recall, have been incorporated into his curriculum Finally, he can compare the structure of his curriculum with student achievement in order to determine where alternative or additional material is required

The teacher finds in the structure for a learning unit a set of specifications which reveal the content and the aims of the unit in detail He is provided with a basis for instruction, both in terms of appropriate sequencing and in terms of transfer devices between related tasks With this picture of the learning unit he can guide students to maximum learning

Having administered and marked the formative evaluation test the teacher can by inspecting conditional item probabilities, identify specific aspects of his work which were ineffective For a three-level structure such as that illustrated in Figure 1 Airasian (1) obtained the results for a chemistry class set out in Table 2

TABLE 2
CONDITIONAL AND UNCONDITIONAL PROBABILITIES FOR THREE ITEM HIERARCHY

Item	*Conditional* Probability	*Unconditional* Probability
1	750	750
2	333	250
3	800	200

Seventy-five per cent of the students in the class were able to answer correctly item 1 which tested the lowest level task in the structure The probability of answering item 2, given a correct response to item 1 was 333 Most of the students who answered both item 1 and item 2 correctly could answer item 3 correctly The unconditional probabilities show that very few of the students answered items 2 and 3 correctly regardless of their performance on lower level prerequisite items

A curriculum constructor or a teacher viewing such group results would notice that for most students the stumbling block was task 2 Somehow the teaching failed to convey to most students the information or the understanding demanded by task 2 Those students who did master task 2 did well on task 3 as evidenced by the high conditional probability for item 3 Further, knowing that steps 1 and 2 are prerequisites for steps 3, the logical place for the teacher to concentrate his corrective efforts is at level 2 Formative evaluation procedures, therefore, provide the teacher not only with an indepth picture of what learning has or has not occurred, but also with a built-in strategy for sequencing remedial activities If such formative evaluation instruments are administered at frequent points in instruction the teacher can locate and correct major areas of student difficulty at a time when such correction is most needed

The student likewise can be given a very specific picture of his progress It is essential however that progress be shown in terms of the pattern of responses rather than as the total score Two students might easily gain the same total score, and yet reveal very different response patterns over the range of individual items If on the three-level structure of Figure 1, a student had mastered task 1 but not tasks 2 and 3, his revision should naturally concentrate on level 2 His revision should be quite different than that of a student who had mastered levels 1 and 2 but failed at level 3 In other words, the relationships between levels prescribe the strategies of revision It is no longer sufficient to tell the student simply to work harder, or to reread a whole chapter or the like We should be able to pinpoint exactly what he knows and doesn't know, and what the relationships are between what he knows and doesn't know

CONCLUSION

Formative evaluation procedures are yet in their infancy The research thus far seems to indicate that the theoretical underpinnings are sound What is required is additional study along the lines suggested in this paper in the hope that someday not too far in the future, we will be able to provide the curriculum constructor, the teacher, and the student with the type of information each needs at the time when he most needs it

REFERENCES

1 AIRASIAN, P W Formative evaluation instruments A construction and validation of tests to evaluate learning over short time periods Unpublished Ph D dissertation, University of Chicago 1969

2 BLOOM, B S (Ed) *Taxonomy of educational objectives Handbook I Cognitive domain* New York McKay, 1956

3 BLOOM, B S Testing cognitive ability and achievement In GAGE, N L (Ed), *Handbook of research on teaching* Chicago Rand McNally, 1963 Pp 379-397

4 CRONBACH, L J Course improvement through evaluation *Teachers' College Record* 1963, lxvi, 672-683

5 FURST, E J *Constructing evaluation instruments* New York McKay, 1958

6 GAGNE, R M The acquisition of knowledge *Psychological Review,* 1962, lxix, 355 365

7 GAGNE, R M *The conditions of learning* New York Holt, Rinehart and Winston, 1965

8 GAGNE, R M, and BASSLER, O C Study of retention of some topics of elementary nonmetric geometry *Journal of Educational Psychology,* 1963, liv, 121-131

9 GAGNE, R M, and BOLLES, R C A review of factors in learning efficiency In Galanter, E (Ed), *Automatic teaching The state of the art* New York Wiley, 1959 Pp 13 53

10 GAGNE, R M, MAYOR, J R, GARSTENS, H L, and PARADISE, N E Factors in acquiring knowledge of a mathematical task *Psychological Monographs,* 1962, lxxvi (7, Whole No 226)

11 GAGNE, R M, and PARADISE, N E Abilities and learning sets in knowledge acquisition *Psychological Monographs* 1961 (14, Whole No 218)

12 GILBERT, T F Mathetics The technology of education *Journal of Mathetics* 1962, 1, 7-73

13 GLASER, R Implications of training research for education In Hilgard, E R (Ed), *Theories of learning and instruction* The sixty third Yearbook of the National Society for the Study of Education, Part I Chicago NSSE, 1964 Pp 153-181

14 LINDVALL, C M (Ed) *Defining educational objectives* Pittsburgh University of Pittsburgh Press, 1964

15 MECHNER, F Behavioral analysis and instructional sequencing In Lange, P C (Ed), *Programed instruction* The sixty-sixth Yearbook of the National Society for the Study of Education, Part II Chicago NSSE, 1967 Pp 81-103

16 MILLER, R B Analysis and specification of behavior for training In Glaser R C (Ed) *Training research and education* New York Wiley, 1965 Pp 31-62

17 SCHUTZ, R E, BAKER, R L, and GERLACH, V S Measurement procedures in programed instruction Final Report, Title VIII, Project 909, NDEA, Grant No 7-12-0030-160, 1964

18 SCRIVEN, M The methodology of evaluation In Stake, R (Ed), *Perspectives of curriculum evaluation* Chicago Rand McNally, 1967 Pp 39-83

19 TYLER, R W *Constructing achievement tests* Columbus, Ohio Ohio State University, 1934

20 TYLER, R W *Principles of curriculum and instruction Syllabus for education 360* Chicago University of Chicago Press, 1950

21 VAUGHN, K W Planning the objective test In Lindquist, E F (Ed), *Educational measurement* Washington, D C American Council on Education, 1951 Pp 159-184