ACHIEVEMENT EXPECTATIONS: A SURVEY OF PROFESSIONAL OPINIONS*

by

Paul A. Bradley

CSE Report No. 80
July 1972

Evaluation Technologies Program
Center for the Study of Evaluation
UCLA Graduate School of Education
Los Angeles, California

*Paper presented at the annual meeting of the American Educational Research Association, Chicago, April, 1972.

TABLE OF CONTENTS

Introduction

One of the great demands on schools today is to make changes and revisions in their curriculum, either to make education more responsive and relevant to student needs and interests or to improve student performance on an important educational goal. At any given time in a school there may be several educational goal areas in which changes in the instructional program are being considered. In most instances a school probably does not have the resources, in terms of money, time, or personnel, to plan and implement new programs in every educational goal area. Instead, the school has to establish its priorities for change; it has to decide which educational goals are most in need of curriculum change. The CSE Elementary School Evaluation KIT: Needs Assessment (Hoepfner, Bradley, Klein, and Alkin, 1973), a product of the Evaluation Technologies Program at the Center for the Study of Evaluation, contains a set of procedures for carrying out a needs assessment.

The goal of the KIT is to rank educational goals in terms of priorities for curriculum change. The educational goal that is ranked first in priority is the educational goal area for which a new program will be planned. The KIT does not specify what the new program will be; the selection of a program is the goal of Program Planning, and procedures for accomplishing this will be contained in the forthcoming CSE Elementary School Evaluation KIT: Program Planning. In the Needs Assessment KIT the ranking of the educational goals in term of priority for curriculum change is based on the application of a decision model. While there are other important aspects of the KIT the heart of it is the multivariate decision model.

There are four variables in this decision model:

(1) the rated importance of an educational goal area

(2) student achievement in an educational goal area

(3) the utility or value of improving student achievement in an educational goal area

(4) the probability of improving student achievement in an educational goal area

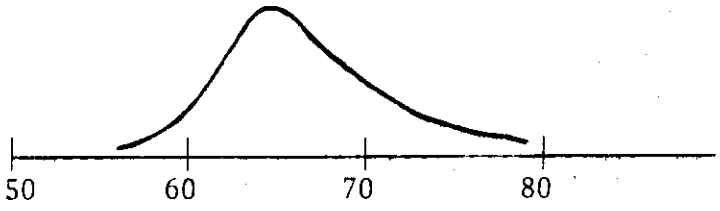
The theoretical foundations of this decision model are discussed in two CSE publications (Amor and Dyer, 1970; Dyer, 1972), and its implementation is described in the Needs Assessment KIT (Hoepfner, et al, 1973). Briefly, in order to implement the decision model it is necessary to obtain a value for each goal area on all four variables in the model. Variables three and four -- the utility of improving student performance and the probability of improving student performance -- are combined into a single variable that is called the probable increase in utility, and the probable increase in utility in a goal area is conditional on the level of student achievement in a goal area. In general, and not unexpectedly, the lower the student achievement, the higher the probable increase in utility. The probable increase in utility of a goal area is multiplied by the rated importance of a goal area, and this product is called the priority value of a goal area. When priority values have been computed for all goal areas, the goal areas are ranked in terms of priority value.

The strategy for implementing the decision model called for the user of the KIT (typically an elementary school principal) to determine the value of goal areas on the first two variables, but the value of goal areas in the last two variables would be included in the KIT itself. That is, the user of the KIT is required to obtain ratings of importance of educational goal areas and to determine student achievement in the goal areas, but he is provided with the probable increase in utility of goal areas. In order to provide these data it was necessary to obtain information separately for the two variables that probable increase in utility is based on -- probability and utility of improving student performance in goal areas. The research

reported here was undertaken to obtain data on the probability of improving student performance in various educational goal areas.

Method

The theory behind the decision model for ranking goal areas in terms of priority for curriculum change (Amor and Dyer, 1970; Dyer, 1972) calls for having conditional probability functions for each educational goal area that give the probability of improving student performance from its current level. It was arbitrarily decided that student achievement would be measured in terms of norm referenced percentile ranks. Therefore, in theory, for each educational goal area there should be 99 probability distributions. Each distribution would give the distribution of student achievement scores (percentiles) at a time -- t_1 , given that the mean student achievement score at a prior time -- t_o -- was at the *n*th percentile, and given that between time t_1 and t_o the students received an experimental treatment that was designed to improve student achievement. It is assumed that these probability distributions represent net improvement in student achievement, with any improvement due to maturation, the regular instructional program, or any other effect not associated with the experimental program being removed. Here is an example of what one of these probability distributions might be like, if mean student achievement in arithmetic operations at time t_o was at the 60th percentile.



This distribution presents a very optimistic picture, as it shows that the probability of improving student performance is quite high. Only about 5% of the students will not surpass the old average percentile rank of 60.

If we were to take the expected value of this probability distribution, (or, equivalently, get the average percentile rank at time t_1), we would see that it might be about 67. This represents a net gain in student achievement of 7 percentile points, and this gain, in theory, is attributed solely to the effect of the experimental program.

Now, since, there are supposed to be 99 such probability distributions for each goal area, and there are many educational goal areas, it is easily seen that the required number of probability distributions is prohibitively large. More importantly, however, is the stark reality that it would be impossible to ever obtain such probability distributions. In lieu of having empirically based probability distributions it was decided to approximate them by means of a questionnaire. The questionnaire would solicit the opinions of professional researchers and educators in the field of curriculum with respect to the likelihood of improving student performance in various educational goal areas. The data obtained from this questionnaire survey would then be used to compute discrete probability distributions that would be taken as approximations to the continuous probability distributions required by the decision model.

The questionnaire that was used in the survey was the sixth version that was tried. The Appendix contains the directions to this questionnaire as well as a sample page. Without repeating the directions, the general idea of the questionnaire is that the performance of a group of students who receive an experimental instructional program is compared to the performance of a group of students who received the regular instructional program. Specifically, the respondent must indicate, "The percentage of students in the experimental group whose year end score will exceed that (the mean) of the control group by at least 1 (5, 10, 15) item is: _____." This procedure was arrived at after trying out several other means of obtaining the data we wanted. While

there were some reasons for discarding the earlier questionnaires, the choice of this procedure over its most recent predecessor was arbitrary. A small pilot study was undertaken to determine if either procedure was better than the other. The result was inconclusive, so the current procedure was chosen because we liked it better. The procedure not chosen was one that asked the respondent to indicate, "...the probability that improvement in performance will be at least 1 (5, 10, 15) item(s)." We thought that using the notion of proportion would be more meaningful to respondents than the notion of probability.

Up to this point reference to educational goal areas has been made several times. It is necessary at this point to indicate that we needed to obtain probabilities of improving student achievement for 40 specific educational goals. These specific goals are the ones that the Needs Assessment KIT presents as a comprehensive set of goals for an elementary school. These goals are listed in Table 1. Because there were four questions asked for each specific goal (the four different improvements in student achievement -- 1, 5, 10, and 15 items), it was decided for very practical reasons not to include all 40 goals in one questionnaire. Doing that would have made for a 20 page, 160 item questionnaire, and we felt that this was too long. Such a questionnaire would undoubtedly end up in too many circular files. Instead, the 40 goals were split into four groups of 10, which made for a more manageable questionnaire of 40 items. Thus, rather than being one questionnaire there are four questionnaires, and these are referred to in the results and discussion section as Questionnaires A, B, C, D. The 10 goals in each questionnaire form are listed in Tables 1a, b, c, and d.

There is one more variable included in the questionnaires. Each questionnaire gives the respondent information concerning the characteristic level of student achievement (low, medium, or high) and the current level of student

achievement in the specific goal areas (also low, medium, or high). This information was included because of the possibility that the probability of improving student performance might depend on the relationship of students' current level of achievement in a specific goal area to their characteristic level of achievement. For instance, the probability of improving the performance of students whose characteristic level of achievement is high but whose current level of achievement in a specific goal area is medium, might be different than the probability of improving the performance of students whose level of achievement is both characteristically low and currently low. Only six of the nine possible combinations of characteristic and current level of achievement were formed. The three combinations omitted were low characteristic-high current, medium characteristic-high current, and high characteristic-low current. This variable is referred to in the results and discussion section as SCHOOL TYPE. For a given questionnaire form, only one SCHOOL TYPE is specified.

When the six different school types are combined with each of the four questionnaire forms it results in there being 24 different questionnaires. To summarize, the 24 questionnaires differ from each other in two ways: (1) in terms of the four sets of educational goal areas, and (2) in terms of the six SCHOOL TYPES.

Sample

The population chosen for the survey was the membership of Division B, Curriculum and Objectives, of AERA. Each member of AERA who indicated a primary affiliation with Division B was sent a questionnaire, a cover letter, and a return envelope. While the questionnaires were not sent out randomly, the system used was not biased in any systematic way. The questionnaires were arranged into sets of the 24 different questionnaires, the membership

list was typed onto labels in alphabetic order, and then each label was matched with one questionnaire. Over 2000 questionnaires were mailed, and 368 complete and usable questionnaires were received. Another 103 questionnaires were returned, but they were either blank or incomplete. Several of those who returned blank questionnaires expressed an opinion of the survey, and it may be that all of those who returned a blank questionnaire were expressing an opinion. No attempt was made to obtain any characteristics of those who did return questionnaires. On an a priori basis, there were no compelling reasons for obtaining any data on age, sex, geographical region, institutional affiliation, etc. In particular none of this information was necessary for the Needs Assessment KIT, and there were no research hypotheses that we wished to explore. The 368 usable questionnaires were the basis for the subsequent analyses. The distribution of these questionnaires over the four questionnaire forms is given in Tables 1a, b, c, and d.

Analysis

Because the four questionnaire forms contained different sets of goals, it was necessary to perform separate analyses for each questionnaire form. At first, the model chosen for performing the various analyses was the repeated measures analysis of variance model. Later, when a multivariate analysis of variance program was available, the multivariate model became the means of analyzing the data. Two major series of analyses were performed. The first series was based on the raw data from the questionnaires, and its main purpose was to see if there was an interaction between the variables GOALS and AMOUNT OF IMPROVEMENT. Figure 1 illustrates a way of looking at the results, and trend analyses were performed to determine if there were differences between the 10 goal areas in each questionnaire form in terms of these functions.

A second series of analyses was performed to test for differences between SCHOOL TYPES, GOALS, and questionnaire FORMS. In order to reduce the amount of data the original data were transformed so that the expected improvement in student achievement in each goal area was characterized by a single value rather than by the four proportions obtained in the questionnaire. The transformation corresponds to taking the expected value of a discrete probability distribution, and was defined as:

$$EV = 3(P_1 - P_2) + 7.5\ (P_2 - P_3) + 12.5\ (P_3 - P_4) + 15\ (P_4)$$

where

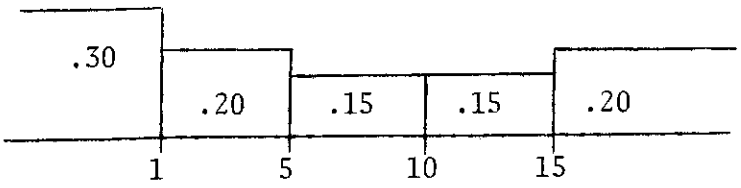
P_1 = the percentage of students in the E group whose year end score exceeds the mean of the C group by at least 1 item.

P_2 = the percentage of students in the E group whose year end score exceeds the mean of the C group by at least 5 items.

P_3 = the percentage of students in the E group whose year end score exceeds the mean of the C group by at least 10 items.

P_4 = the percentage of students in the E group whose year end score exceeds the mean of the C group by at least 15 items.

These four proportions (or percentages) can be used to construct a discrete probability distribution. For example, if $P_1 = .70$, $P_2 = .50$, $P_3 = .35$, and $P_4 = .20$ then the discrete probability distribution obtained from those proportions would look like



It can now be seen that the transformation corresponds to taking the sum of the products of the proportion of students in an interval and the mid-point of the interval. However, for the first and last intervals it was arbitrarily decided to set the mid-points at 0 and 15, respectively. With these transformations made, a one-way multivariate analysis of variance was performed for

each questionnaire form. The independent variable was school type, with 6 levels, and the dependent variables were the expected improvements in student achievement for 10 goal areas.

This design allowed for three basic hypotheses to be tested. Two of these hypotheses correspond to testing main effects while the third hypothesis corresponds to testing an interaction effect. Since a significant interaction effect would probably limit the meaningfulness of significant main effects, it is necessary to examine this hypothesis first. This interaction hypothesis is concerned with whether the mean expected improvements in achievement in the 10 goal areas are equal across school types. In the multivariate model this hypothesis is tested by the multivariate test on the independent variable. The main effects hypothesis that is of major interest is whether the mean expected improvements in the 10 goal areas are equal. The other main effects hypothesis is whether the mean expected improvements in the 6 school types are equal.

Since there were 24 different questionnaires used in this survey there was some interest in determining if there were any response biases caused by the different contexts that affected the level of responses. In order to examine this issue it was necessary to take the original raw data and, for each person, convert the 40 proportions (10 goals by 4 improvements) to 4 proportions by averaging over goal areas. This transformation resulted in a design that had 2 independent variables -- questionnaire form (A, B, C, D) and school type (6 levels) -- and 4 dependent variables -- the proportion, averaged over 10 goal areas, of students in the experimental group whose year end score exceeded the mean of the control group by at least 1, 5, 10, or 15 items. This analysis, as well as all others, was performed using the program MULTIVARIANCE.[1]

[1]This program was written by Jeremy Finn, SUNY, Buffalo.

Results

Because this last analysis is the least interesting of the three, but potentially very annoying, the results for it are considered first. When a multivariate analyses of variance was performed with school type and questionnaire forms as independent variables and 4 improvements in performance as dependent variables, no effects were statistically significant ($p < .05$). This was encouraging as an earlier plot of the 24 mean proportions (averaged over the four dependent variables) suggested that there might be an interaction between school type and questionnaire form, as well as a significant main effect for questionnaire form. But the results of the MANOVA, as well as the results of a univariate ANOVA of the same data, lead to the conclusion that there are no significant differences in response level between the 4 questionnaires forms or the 6 school types.

Trend analysis

The basic purpose of the series of trend analyses was to determine if there were differences in trends over the four improvements in performance between goal areas. However, this was not the only trend analysis performed for each questionnaire form. There were a total of four different trend analyses that were performed for each questionnaire:

(1) Test for overall linear, quadratic, and cubic effects. This was done by collapsing the data over school types and goal areas. There is very little interest in this test.

(2) Test for differences in linear, quadratic, and cubic effects among the 6 school types. For this test the data are collapsed over goal areas, so there is less interest in this test than in the following two analyses.

(3) Test for differences in linear, quadratic, and cubic effects among the 10 goal areas. This is the analysis of most interest. The data are collapsed over school types.

(4) Test to see if any differences in (3) are the same for all 6 school types. There is not much interest in this hypothesis, but significant differences here would preclude making any generalizations about school types or goal areas.

There was a consistency of results for these four trend analyses across the four questionnaire forms. The results are grouped together corresponding to the above outline.

(1) In all four questionnaires the linear effect is very large and very significant ($p < .0001$). This is expected from an inspection of the plot of the overall means. What is not expected, except after closer examination of the plots, is that there are consistent quadratic and cubic effects. In only one questionnaire (form A) is the quadratic and cubic effect not significant at a p-level of .0005.

(2) In no questionnaire are there significant differences in trends between school types.

(3) In all four questionnaire forms the test for differences in linear effects between goal areas is highly significant ($p < .0004$), but of small magnitude. The F values range from 4.0 to 6.0, as compared to F values ranging from 600. to 800. for the overall linear effects in (1). If an alpha level of .01 is adopted for rejecting null hypotheses then no significant differences in quadratic or cubic effects among goal areas were found. Finding that there are significant differences in linear effects among goal areas was somewhat surprising, as an inspection of the plots of the trends for goal areas suggests that there would not be significant differences in linear effects. Figure 1 plots the trends for the 10 goal areas in questionnaire D (which had the largest F value for linear effects). In this plot only 1 goal area has a trend that is "perceptually" different from the others. If it is any consolation it is that the magnitude of the differences in linear trend among the goal areas is small.

(4) No significant differences were found, anywhere.

Expected improvement in student achievement

This series of analyses was of greatest interest as it involved looking for differences among goal areas and school types in terms of expected improvements in student achievement. The results of the MANOVA for each questionnaire form are given in Tables 2a, b, c, and d. As with the trend analyses there is substantial consistency in results across the four questionnaire forms. In all cases there are no significant differences among school types when there is only one dependent variable, there are small but significant differences among goal areas, and in three of the four questionnaire forms there is no interaction

between school types and goal areas -- the multivariate test of school types. In questionnaire form A there is a small but significant interaction between school type and goal area.

The mean expected improvements in achievement for the 40 goal areas are given in Tables 1a, b, c, and d. While the differences in these means are not great they do tend to be meaningful. The educational goals in which there seems to be the greatest improvement in achievement are in the areas of mathematics, reading, music, social studies, and arts and crafts. The educational goals in which there seems to be the least improvement in achievement are from the affective and cognitive domain. The particular goals that are perceived as being most difficult to effect improved student achievement are Personal Temperament, Social Temperament, Attitudes (including self-esteem), Reasoning, Creativity, Memory, Music Performance, and Physical Skills. It should be noted that this listing is based on the ranking of goals within a questionnaire form, not on absolute value of mean expected improvement in achievement. There are two interesting notes to make about these educational goals. First of all, they involve student characteristics that would be considered more innate, unlearned, and possibly heriditary than they are learned or determined by environmental influences. The second thing is that these goals typically are not taught in school. Rather they frequently are considered to be personal characteristics that are strongly related to student achievement. Consequently they are not considered explicitly as criterion variables but more as predictor variables. These goals may frequently be desirable outcomes of schools, but there probably are few instances where a curriculum actively attempts to affect student performance in these goal areas.

The particular goals that seem to be perceived as being less difficult to effect improved student achievement include Producing Arts and Crafts,

Reference Skills, Arithmetic Operations, Geometry, Measurement, Music Appreciation and Interest, Music Understanding, Oral-Aural Skills, Word Recognition, Reading Mechanics, Scientific Knowledge, and Sociology. With two exceptions (Producing Arts and Crafts and Music Appreciation and Interest) these educational goals pretty much involve acquired knowledge, which is what most elementary schools emphasize in their instructional programs. It is interesting to note that of all the goals in Mathematics and Reading there are only two that have been perceived as being more difficult to effect improved student achievement: Arithmetic Concepts and Reading Appreciation and Response. It is possible that Arithmetic Concepts is perceived to be more like an aptitude than an acquired skill. Mastery of arithmetic concepts is likely to be much more dependent on something like "quantitative aptitude" than is mastery of the other arithmetic skills. The low rank of Reading Appreciation and Response is, at first, puzzling and surprising, but a closer look at what the goal means reveals that one component involves attitude and behavior modification based on a students reading. It is likely that it is this component which has led to the goal's overall low rank.

Discussion

At this time it would be good to remember that the purpose of this survey was to obtain data on the probability of improving student achievement for the CSE Elementary School Evaluation KIT: Needs Assessment. The variable probability of improving student achievement is one variable of a decision model whose purpose is rank educational goals in terms of priority for curriculum change. While the ultimate outcome of using the decision model is a ranking of goals, the outcome is based, in theory, on having conditional probability distributions of student achievement which is very parametric, quantitative data. Since this kind of data was the goal, the questionnaire was constructed

to obtain a gross approximation to it. In retrospect, it seems that while the questionnaire may have provided data that was a reasonable approximation to the conditional probability distributions, it may not have been a good device (i.e., sensitive) for obtaining a ranking of educational goals in terms of the easiness or difficulty of improving student performance. A reason why the questionnaire may have been inadequate for ranking goals is that its structure would have placed greatest emphasis on the probable distribution of student achievement in one goal area. The questionnaire was not structured so as to emphasize comparisons between goal areas or between school types.

That is, it seems that there ought to be greater differences between goal areas and school types in terms of probable improvements in student achievement. While there were statistically significant differences among educational goals in terms of both linear trend and expected improvement in student achievement, the differences were quite generally of a small magnitude. Figure 1 is remarkable in terms of the "perceptual similarity" of the trends, and in Tables 1a, b, c, and d, the largest difference in mean expected improvement is the difference between Measurement (7.40) and Personal Temperament (4.85), a difference of 2.55, which hypothetically corresponds to a difference of 2.55 items on a 100 item achievement test. Given a most difficult area to work with, Personal Temperament, an area that even clinical psychologists, psychotherapists and analysts have difficulty effecting changes in, the questionnaire and procedures used to analyze the data from the questionnaire lead to the conclusion that there is little difference between this goal area and all the others in terms of probable improvement in student performance. It would seem that the sample used would have different opinions about differences between goals and school types than the results would indicate. In reality it may be that there are small differences among educational goals and school types in terms of changing student performance, but it seems that the opinions of professional researchers

and educators would reflect larger differences. Assuming that professionals do think there are larger differences among goals and school types than the questionnaires reveal, it would now be necessary to revise the questionnaires so that the emphasis is placed on comparisons among goals and school types rather than on the distribution of student achievement.

Summary

A decision model for selecting the most critical goal area for an elementary school has as one of its components the improvement in student performance that can be expected when a new instructional program is implemented in the school. Questionnaires were designed to obtain the opinions of educators regarding the expected improvement in student performance for 40 goal areas and 6 different school types. The results indicated there were statistically significant differences between goal areas, but not between school types. However, the differences were so small in magnitude as to suggest that the questionnaires were inadequately structured to be sensitive to differences between goal areas and school types.

References

Amor, J.P. and Dyer, J. S. A decision model for evaluating potential change in instructional programs. CSE Report No. 52. Los Angeles: Center for the Study of Evaluation, University of California, 1970.

Dyer, J.S. A guide for the allocation of discretionary resources to educational subject areas. CSE Working Paper No. 20. Los Angeles: Center for the Study of Evaluation, University of California, 1972.

Hoepfner, R., Bradley, P.A., Klein, S.P., & Alkin, M.C. CSE elementary school evaluation KIT: Needs assessment. Boston: Allyn and Bacon, 1973, in press.

Table 1a

Mean Expected Improvements in Achievement

Questionnaire A (N = 92)

GOAL	$\bar{X}$
1. Music Performance	6.75
2. Reading Mechanics	7.33
3. Geography	7.61
4. Arithmetic Concepts	6.77
5. Reading Comprehension	6.88
6. Understanding Arts and Crafts	7.14
7. Needs and Interests	7.11
8. Reference Skills	7.70
9. Scientific Processes	6.86
10. Physical Skills	6.79

Table 1b

Mean Expected Improvements in Achievement

Questionnaire B (N = 88)

GOAL	$\bar{X}$
1. Music Understanding	6.65
2. History and Civics	6.22
3. Valuing Arts and Crafts	6.17
4. Oral-Aural Skills	6.33
5. Arithmetic Operations	6.41
6. Language Construction	5.97
7. Scientific Approach	5.79
8. Health and Safety	6.30
9. Attitudes	5.88
10. Memory	5.08

Table 1c

Mean Expected Improvements in Achievement

Questionnaire C (N = 93)

GOAL	$\bar{X}$
1. Music Appreciation and Interest	7.02
2. Application of Social Studies	6.03
3. Foreign Language Assimilation	6.09
4. Geometry	7.01
5. Producing Arts and Crafts	6.87
6. Social Temperament	5.28
7. Religious Knowledge and Belief	4.90
8. Sportsmanship	6.10
9. Reasoning	5.99
10. Reading Interpretation	6.53

Table 1d

Mean Expected Improvements in Achievement

Questionnaire D (N = 95)

GOAL	$\bar{X}$
1. Mathematical Applications	6.63
2. Creativity	6.50
3. Personal Temperament	4.85
4. Measurement	7.40
5. Scientific Knowledge	7.37
6. Reading Appreciation and Response	6.15
7. Physical Education	6.95
8. Foreign Language Skills	6.58
9. Word Recognition	7.39
10. Sociology	7.06

Table 2a

Multivariate Analysis of Variance - Questionnaire A

Source	F[a]	df Hypothesis	df Error	P
School Types - univariate	2.0542	5	86	ns
Goals	3.1850	9	83	<.0024
School Types - multivariate	1.5905	50	354.54	<.0094

a Rao's formula applied to Wilks' Lambda criterion

Table 2b

Multivariate Analysis of Variance - Questionnaire B

Source	F	df Hypothesis	df Error	P
School Types - univariate	0.8234	5	82	ns
Goals	5.3377	9	79	<.0001
School Types - multivariate	0.8567	50	336.3	ns

Table 2c

Multivariate Analysis of Variance - Questionnaire C

Source	F	df Hypothesis	df Error	P
School Types - univariate	0.9381	5	87	ns
Goals	6.8741	9	84	<.0001
School Types - multivariate	1.2624	50	359.10	ns

Table 2d

Multivariate Analysis of Variance - Questionnaire D

Source	F	df Hypothesis	df Error	P
School Types - univariate	0.4273	5	89	ns
Goals	8.0556	9	86	<.0001
School Types - multivariate	1.0937	50	368.22	ns

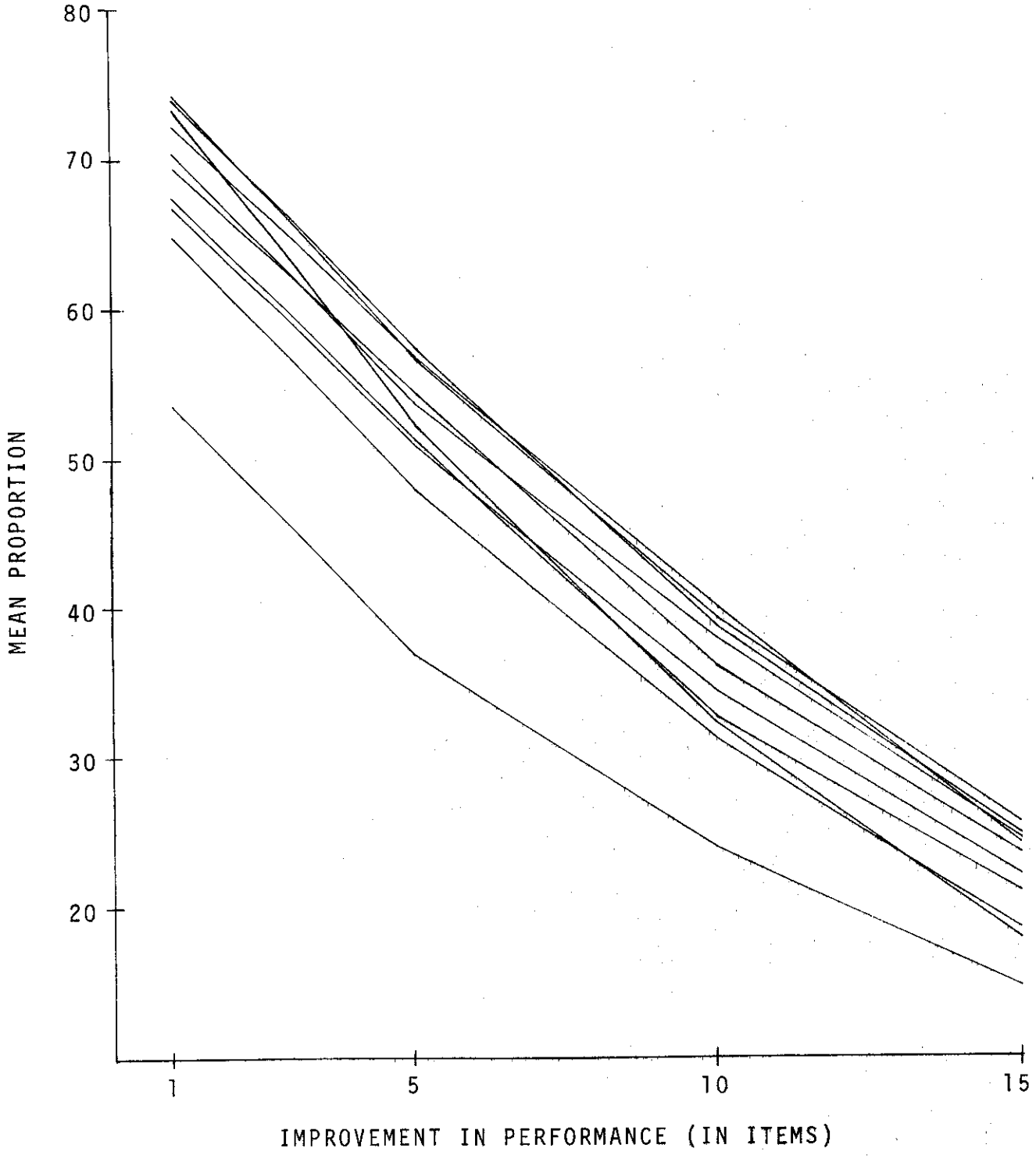


Figure 1. Trends for 10 Goal Areas in Questionnaire D

Appendix

Questionnaire Directions

CENTER FOR THE STUDY OF EVALUATION
UCLA
Subjective Probability Questionnaire

This questionnaire is designed to elicit your opinion regarding the likelihood of improving student performance in various educational goal areas at the elementary level. We want to obtain your judgments as to the chances of improving student achievement in such goal areas as reading comprehension, arithmetic operations, social studies, or self esteem. Because this is both a complex and a vague task, further definition of the situation is given. This is the situation you are to consider:

> *at the present time an elementary school is relying on a typical regular instructional program that involves teachers, students, and textbooks. In the interest of improving educational achievement, this school is going to implement a new instructional program and see if it improves student performance on a standardized achievement test. Accordingly, all students in a particular grade are divided randomly into two groups. One group continues to receive the regular instructional program (call this the control group), but the other group receives the new instructional program (call this the experimental group).*

The effect of the new instructional program will be assessed by comparing the year-end performance of the two groups on a 100 item standardized achievement test. In this standardized test, assume that each additional item correct increases a student's percentile score by one (1) point. For example, a raw score of 60 items correct corresponds to a percentile score of 60, and a raw score of 70 items correct corresponds to a percentile score of 70. Assume that the two groups had the same average at the beginning of the year.

We want to know your subjective estimate of the effect that a new instructional program will have on student achievement in a particular educational goal area. For instance, if a new instructional program in Scientific Knowledge is implemented, we might ask you to estimate the percentage of the students in the experimental group whose performance on the achievement test exceeded the average score of the control group by at least 10 items. That is, if the average score of the control group was 45 items correct, what percentage of the experimental group would get at least 55 items correct?

Now, we are well aware that such a proposition may depend on what the new instructional program is and how much it costs. At a later time we will investigate the differential effect of various instructional programs within one educational goal area; right now we are interested in the differences between educational goal areas. It is possibly the case that the skills in some goal areas are easier to improve than are the skills in other goal areas. If you think that it is relatively easy to improve student skills and achievement in a certain educational goal area, then you would indicate this by saying that a large percentage of the experimental group would get scores that exceed the average score of the control group by at least 10 items. On the other hand, if you think that it is relatively difficult to improve student

skills and achievement in an educational goal area, then you would indicate this by saying that a small percentage of the experimental group would get scores that exceed the average score of the control group by at least 10 items.

In addition to depending on the nature of the new instructional program, the chances of improving student performance may also depend on the relationship of the students' current level of achievement in the particular goal area to their general level of achievement in all other goal areas. For instance, suppose that the students' current performance in a goal area is at a medium level. It may make a difference to you whether these same students are characterized by a generally high, medium, or low level of achievement in all other goal areas. Accordingly, for each goal area in this questionnaire you are to assume that the students in both instructional groups (control and experimental)

are characterized by a generally low level of achievement, and

their current performance in the specific goal area is at a low level of achievement.

This is how the questionnaire is structured: each item specifies an educational goal area and a possible outcome in the year end performance of the experimental group. Look at a sample item.

Educational Goal - Reading Comprehension

Recognition of word meanings, understanding of complex ideas, and remembering information read.

The percentage of students in the experimental group whose year end score will exceed that of the control group by at least 5 items is: _____%

Now, look at the examples on the following page to see how one page of the questionnaire is structured. There are four items on a page. The educational goal area is the same for all four items and is specified at the top of the page. The four items differ from one another in the amount of improvement that is indicated. The response to the first item indicates that the rater thinks 90% of the experimental group will have a year end score in Reading Comprehension that exceeds the average year end score of the control group by at least 1 item. This implies that the remaining 10% of the experimental group will have year end scores equal to or less than the average score of the control group. The response to the second item indicates that the rater thinks 75% of the students in the experimental group will have year end scores that exceed the score of the control group by at least 5 items. Similar interpretations and elaborations can be made for the responses to the last two sample items.

There are ten pages to this questionnaire. Keep the following points in mind as you respond to the items.

we are not testing your accuracy; there are no known correct answers
we want your best estimate, based on whatever information you have
be sure to respond to every item, even if you have to "guess-timate"
remember the assumptions that were underlined in the directions

Sample Items

Educational Goal - Reading Comprehension

Recognition of word meanings, understanding of complex ideas, and remembering information read.

The percentage of students in the experimental group whose year end score will exceed that of the control group by <u>at least 1 item</u> is: 90%

The percentage of students in the experimental group whose year end score will exceed that of the control group by <u>at least 5 items</u> is: 75%

The percentage of students in the experimental group whose year end score will exceed that of the control group by <u>at least 10 items</u> is: 40%

The percentage of students in the experimental group whose year end score will exceed that of the control group by <u>at least 15 items</u> is: 15%

1	2	3	4	5