RETROSPECTIVE COURSE EVALUATION BY ANALYSIS OF EXAMINATION RESULTS

Tjeerd Plomp and Adri Van Der Meer

Twente University of Technology, The Netherlands

1. Introduction

In higher education, course objectives are seldom explicitly stated. Consequently, it is almost impossible to indicate which objectives have been attained by the students at the end of a course. Thus, there is often a need to establish, post facto, which objectives were pursued by the teachers and attained by the students. This need is especially important where a new instructional method is weighed against the usual one or where the identification of critical impact points for improving an existing instructional method is the intent. An analysis of course examinations over several years can provide good insight about the objectives teachers have pursued, because the examination items can be considered operationalizations of the pursued objectives. And, by analyzing achievement on the examination problems, it is possible to determine which objectives the students have attained.

In #3 of this paper a method pertaining to objective identification and analysis will be discussed. Thus a framework will be developed by which post factor objectives can be determined and students' attainment of the objectives can be assessed. The method can also be used for examining the quality of instruction (see #5). Starting point for the analysis is to make a clear distinction between necessarily subjective prior judgments and following rational derivation of relevant conclusions which are consistent with these judgments and with the data. This subjective moment is located in the choice of values for three parameters (α , β and γ ; see #3.2 and 3.4).

Using this method, we are able to determine mathematics course objectives and attainment of the objectives by the group of students who passed the course (see #4). The results of this study are based on mathematics courses to first-year engineering students in the period 1970-1973, using the traditional format of lectures, discussion groups and "open question" examinations.

The explorative character of this research suggests that conclusions from the data have to be considered illustrations of the type of questions that can be analyzed within the proposed framework.

2. Research questions

In this paper, we discuss only a few of the research questions that could be discussed in the context of a retrospective analysis of implicit objectives.

The central question is: *What objectives were attained by the students who passed the course?* Several questions can be distinguished from this general

question:

1. How can the pursued objectives be made explicit so that they can be discussed as well on content levels as on behavioral levels?
2. What were the objectives (on both levels) pursued in the mathematics course?
3. What were the objectives (on both levels) the passed students attained?
4. Were in the examinations the number of items, which are operationalizations of the objectives attained by the group of passed students, such a proportion of the respective examination, that based on the scores on these items a passmark could be obtained? If not, is it possible to characterize the not-attained objectives, of which then must be concluded that the passed students have not or only partially attained them?

The relevance of these research questions comes from the situation that passed students (i.e. students with a pass mark) are considered to have attained course objectives, while often in practice evidence about this assumption is not considered.

Ideally, questions pertaining to the attainment of objectives should be answered within a research design in which all the items from different examinations are connected by calibrating them on the same scale. For this it is necessary either to present some items to persons from different groups or to calibrate afterwards all the examination items on one group. In practice this requirement can not be met, so a retrospective analysis, in which plausible assumptions are made, is needed to answer the above questions.

3. Method of Research

In this section we introduce the method of research which will be applied in the next section as a procedure by which the research questions can be answered.

Considering the first question, we propose a two-dimensional grid fitting the situation of examinations based on implicit objectives (#3.1). After having classified the examination items within the grid, question 2 can be answered (#4.3). To address questions 3 and 4, it is necessary to introduce in #3.2 such constructs as: (1) item in an "open question" examination, (2) the mastery of an item by a student, and (3) measure for attainment of an objective. The method for conducting the retrospective analysis of course objectives is based upon the analysis of examination results which may be influenced by several sources of variation. These sources and some assumptions are discussed in #3.3, as well as the method of analysis in #3.4.

3.1 Grid of objectives

The items or problems in course examinations can be considered operationalizations of the pursued objectives. Each item is the operationalization of a particular objective that may be described on two dimensions: (a) content and (b) behavior. Content refers to specific substantive elements in the body of knowledge. If the content levels are formulated too specifically, items may be a composite of several objectives. Besides, then a too large number of cells should be generated so that per cell not a meaningful number of items will be obtained. Therefore, the content levels must be formulated broadly enough to enable an unambiguous classification and a meaningful interpretation. This can be done by using general terms, such as chapter titles from text books or main topics from the subject field. For the behavioral dimension, several classifications are available in the literature (e.g. the taxonomy of Bloom (1956) for the cognitive domain or its elaboration for mathematics by Wilson (1971). However, it is essential that the intellectual skill domain is as exact a classification

of the measured behaviors as possible. Since the analysis of objectives starts with examination problems, then the nature of these questions must be reflected in the behavioral levels. Therefore, we have chosen as a classification system the main categories of Polya (1957), who distinguishes between "problems to find" and "problems to prove." In each of the two categories some levels are defined, reflecting the different levels of difficulty of the examination problems (see #4.2).

The cells of the grid and the row and column totals should be studied as starting points of the analysis. Since examinations are samples of content and intellectual skills, it is much more likely that the cells of the grid will be representative of the course objectives if a number of examinations is aggregated. Then, not only a good impression of what objectives were pursued can be obtained from the grid, but also from the number of items per cell and per marginal total an impression of the relative importance of the respective objectives. A caution should be noted. Empty or sparingly filled cells do not necessarily mean that the objectives for the cells were not pursued; such cells may appear for example if there is some hierarchy within the columns or the rows, so that some objectives are implicitly included in other objectives.

3.2 Definitions

An *item* or *problem* is the smallest unit in an examination to which a separate score will be given by the corrector and which can be solved independently from the other problems in the examination. That means that so called 'pile'-questions (i.e. questions of which the solution is dependent on the results of one or more preceding questions) will be considered as one item or problem. The m items in a cell are noted as $v_i (i = 1, 2, \dots, m)$.

Student j has mastered item v_i is his score on v_i is more than a certain fraction of the maximum score S_i ; in other words if his score S_{ij} on v_i applies to: $S_{ij} > \alpha S_i$ with $0 < \alpha < 1$. In every application of the method, the value of the parameter α must be chosen. This choice will depend on 'local' factors like the character of the objectives of a particular course or judgments about the appropriate level of mastery.

Attainment of an objective will be defined with the help of the definition of mastery of an item by single students. By introducing the parameter α , the score of a student on an item is dichotomized. This can be made explicit by stating a 1 if $S_{ij}/S_i \geq \alpha$ and a 0 if the reverse holds.

A measure of attainment of an objective operationalized in the item v_i is the ratio:

$$p_i = \frac{\text{number of students with a 1 on } v_i}{\text{total number of students on } v_i}$$

3.3 Assumptions and sources of variation

Consider a cell of the grid with items v_i with $p_i (i = 1, 2, \dots, m)$. Generally, the $p_i (i = 1, 2, \dots, m)$ in a cell will differ. Apart from influences by chance, this should not be the case because all the items in the cell are operationalizations of one objective. Important questions are: (a) how do we explain this variation and (b) how do we combine the information from all the items in a cell and to interpret this on one scale as a measure for the attainment of the objective represented by that cell. Three sources of variation can be considered: (a) students, (b) instruction and (c) items. Students differ in ability, preparation, attitude, etc. The quality of instruction is influenced by different teachers personalities, teaching-styles, instructional procedures, etc. Since these two types of variation can be considerable, it is important to aggregate different course examinations where as little variation has occurred as possible. We assume that the combined influence of both sources is constant

over the considered period of time. This assumption is plausible in situations in which, over the period of time during which a course will be studied, nothing has been changed with respect to (a) the entry requirements, (b) the preceding instruction, (c) the group of executive teachers, (d) the instructional procedures (e) the basic textbook, etc.

In the third source of variation items may differ in difficulty and use of enabling skills, etc. Moreover, in the situation of the retrospective analysis, the items come from different examinations, taken by different groups of students (from the same population). Considering the practice of aggregating examinations, it follows from the first assumption that differences in p-values in a cell are due to differences between items. Thus, for a cell of the grid the p_i ($i = 1, 2, \dots, m$) not only pertain to a measure of attainment of the objectives as represented in the respective items v_i ($i = 1, 2, \dots, m$) but these values also reflect differences in the difficulty of the items. We will assume that all items in a cell discriminate in the same way between more able and less able students.

To interpret the ratio p_i , introduced at the end of the preceding section, as a measure for the attainment of the objective operationalized by item v_i it must be assumed that the relevant ability for answering this item correctly has a normal distribution over the group of students. (In effect not p_i itself but a monotonic transformation of its value must be used, for details see Plomp & van der Meer, 1979).

When a cell in the grid of objectives contains only one item this final assumption will be sufficient. But to combine the p-values of several items in a cell to one measure for the attainment of the objective operationalized by these items it is necessary to make also the other assumptions about the group of students and characteristics of the considered items. This will be mentioned in the discussion of the criteria for the attainment of objectives in the rest of this section.

3.4 Measure for the attainment of objectives

In this section, measures for the attainment of objectives by the students will be discussed. We will discuss the cases of cells with only one, respectively at least two items. In each case the criterion for attainment of the objective has to be chosen; this is a matter of rational judgment.

(a) *Cells with 1 item.* At the end of #3.2 a measure p of attainment for an objective operationalized in an item v was introduced. In this case, it is easy to define: the students have attained the course objective represented by item V if $p > \beta$ with $0 \leq \beta \leq 1$. For example we may judge that a given objective is attained if at least 75% of the students have reached the chosen level α of mastery of an item (thus $\beta = .75$).

(b) *Cells with more than 1 item.* Before discussing the two approaches, we shall define several notations with respect to a particular cell of the grid:

– items	$v_1, v_2, \dots, v_m$
– number of students (generally different for each item)	$N_1, N_2, \dots, N_m$
– number of students that master an item	$n_1, n_2, \dots, n_m$
– measure of attainment of an item; $p_i = n_i/N_i$	$p_1, p_2, \dots, p_m$

b1. Referring to the definition of p_i of each item v_i and to the case of a cell with 1 item, we can consider all the items which apply $p_i \geq \beta$. Let us denote by x the number of items from $v_1, v_2, \dots, v_m$ for which the students have attained the chosen level of mastery α for each item. We define as follows:

The group of students has attained the objective if $\frac{x}{m} \geq \gamma$, which $0 \leq \gamma \leq 1$. The choice of γ is again by judgment. By choosing this definition of attainment, we only have to assume that the combined influence of the sources of variation (a) students and (b) instruction is constant over the considered period of time.

Note that in this definition only the number of problems by which an objective is operationalized is considered. It is possible that the value of m could be very small, by which the application of this measure is not very meaningful.

Example: If $m = 1$, then $\frac{x}{m}$ can be only 0 or 1. If $m = 2$, then $\frac{x}{m}$ can be only 0, $\frac{1}{2}$ or 1.

By taking into account only the number of items, all items have equal weight in the procedure. Use is not made of the fact that different numbers of students are performing on the respective items. Thus, the score of a single student performing on an item taken by a small number of students has a relatively greater weight than in the case of an item taken by a greater number of students. This will not be a problem if (a) all the items are of almost the same difficulty (i.e. the values $p_1, p_2, \dots, p_m$ are almost equal) or (b) the numbers of students taking the respective items are almost equal. In educational practice those conditions are not always fulfilled. An example will help to illustrate the point. Let an objective be operationalized in three examinations in three different items and let the number of students in these examinations be 300, 400 and 40 respectively. This is a realistic situation in our university. Moreover, let us assume that only one item is attained ($p \geq \beta$) by the group of students. If the attained item is the one made by the group of 400 students and the two non-attained items are those in the examinations with the smallest number of students then, considering the whole population of students it may be reasonable to regard the item presented to the large group as a more representative operationalization of the objective than the two other items.

From this point of view, it is desirable to look for a procedure in which not only the difficulty of the items, but also the number of students taking the items will be considered.

b2. To arrive at a procedure for which the two conditions are fulfilled the assumptions of #3.3 are needed. If we use the monotomic transformation $p_i \rightarrow z_i$, defined by

$$p_i = \int_{z_i}^{+\infty} \phi(\theta) d\theta,$$

where $\phi(\theta)$ is the standard normal distribution, then the weighted mean of the z_i 's can be interpreted as a measure for the attainment of the objective in consideration. The z_i ($i = 1, 2, \dots, m$) can be found in a table of the standard normal distribution.

By computing for each cell

$$z^* = \frac{\sum_{i=1}^m N_i z_i}{\sum_{i=1}^m N_i}$$

the different numbers of students per item are taken into account. Moreover, the weighted mean z^* of difficulties can be considered as representing the average

difficulty of the items about the objective students are expected to attain. By means of the inverse transformation at each z^* , a value p^* can be found. This p^* can be interpreted as a sort of 'mean' value of the p_i 's for the items in a cell. Note, that p^* , as the transformed of the weighted mean of z_i 's, is not the same as the weighted mean of the p_i 's. For each objective, corresponding to a cell of the grid, p^* can be interpreted as a measure of the attainment of the objective. Therefore, we define:

The group of students has attained the objective if $p^ \geq \beta$, with $0 \leq \beta < 1$.*

Consistent with other parameters, the choice of β is a matter of judgment. Note that the case of one item per cell is a special case (viz $m=1$) of this.

4. Application

4.1 Available data

We have analyzed the mathematics courses for first year engineering students at our university in the first three quarters of 1970/71, 1971/72 and 1972/73. The students treated in the courses are listed in #4.2. Of each course the examinations at the end of the quarter, immediately after the teaching period, and at the end of the course year are analyzed. In #2 we pointed out that in this analysis the central question is: *what objectives were in fact attained by the group of students which passed the examinations*. Thus the method to determine the attainment of objectives was applied to this specific group.

The data for this retrospective analysis were collected in 1976. It appeared that not every teacher of the mathematics classes had saved the list with scores of the students on the respective problems. So, the analysis could only be applied on a part of the students that took the examinations. Table 1 contains the number of available data per examination. Each examination was labeled with a number. The first digit indicates the quarter of the course year in which the mathematics course was given. The second digit indicates the several examinations of the particular course.

TABLE 1: Number of Students per Examination

Exam. Number	11*	12*	13*	14	15*	16	21*	22	23*	24	25*	26	31*	32	33	34	35*	36	37*
Number of Students	66	21	40	16	46	4	85	12	37	10	46	21	107	4	38	23	104	31	55
Passed Students	40	17	17	12	28	3	53	5	18	8	18	7	87	3	10	13	50	14	38

*These examinations are used in the analysis of attained objectives.

From Table 1 we see that the available data on some examinations were very small. Therefore, the analysis of what objectives were pursued is based upon all the examinations, but the analysis of what objectives were attained is only applied on these examinations with at least 15 passed students.

In #4.4, the measure for attainment of objectives discussed in #3.4 under b2 is applied. With respect to the assumptions mentioned in #3.3, the following can be said. The assumptions about the item characteristics are underlying the analysis, their fulfilment was presupposed. The assumption about the distribution of relevant abilities over the entire group of passed students is plausible for the considered mathematics courses in the period 1970-1973, because in this period there were no changes in the entry requirements, the preceding instruction

(secondary school), the group of executive teachers, the instructional procedures (lectures and practice sessions), the basic textbook, etc. Finally it was assumed for the group of passed students that the variable expressing the ability to solve mathematics problems has a standard normal distribution.

For these considerations we conclude that the results of the analysis will have no general validity. They must be considered mainly as illustrations of the possibilities of the method.

4.2 Grid of objectives

The problems in the examinations were distributed over the following *content categories*:

1. *Mathematical Induction*;
2. *Limits* (of sequences; functions of one variable: continuity, differentiability, computing of limits, theorem of l'Hopital);
3. *Differential calculus* (finding the derivative, linear approximation, mean value theorem, extrema, points of inflection);
4. *Integral calculus* (primitive function, Riemann sum, definite integral, length of arc, area and volume of surfaces of revolution);
5. *Functions of two variables* (continuity, differentiability, partial derivatives, linear approximation);
6. *Differential equations* (different types of first order equations);
7. *Series* (tests for convergence and divergence, power series, series expansions of functions, Taylor's formula, computation with series);
8. *Improper integrals* (tests for convergence and divergence).

These categories reflect the main topics in the three mathematics courses. The categories 1, 2 and 3 refer to the first quarter course: mathematics I, 4, 5 and 6 to mathematics II and 7 and 8 to mathematics III. The number of categories is such that each item could be categorized unambiguously according to content, while yet each category contains a meaningful number of items.

Indicated in #3.1, is the classification in levels of behavior using the main categories of Polya (1957), who distinguishes between "problems to find" (F) and "problems to prove" (P). Additionally a rest category, "reproduction" (R), was used.

F. Problems to find

The items in this category are characterized by "compute", determine, "find" and/or "approximate". Three levels have been distinguished:

- F1: Elementary problems, solved with only one standard method (i.e. a for the students well known method).
- F2: Problems solved by using several standard methods after each other or by using a standard method after a non-trivial re-formulation of the problem.
- F3: Problems solved with non-standard methods (i.e. new problems).

P. Problems to solve

These items are characterized by words such as: "prove" or "examine if...". We distinguish four levels:

- P1: Problems indicating which definition or theorem must be applied and/or which standard type of proof can be used.
- P2: Problems wherein the theorem to be used is not given and/or a non-trivial re-formulation of the problem is necessary as a first step.

- P3: Problems in which several definitions, theorems and/or standard types of proof must be used (i.e. new problems);
- P4: Proofs of generalizations or specializations of well-known theorems, of which it can be expected that the students have *not* learned them by heart.

R. Reproduction

This category is not further subdivided.

4.3 Pursued course objectives

The result of categorizing the examination items within the grid is summarized in Table 2.

TABLE 2: Number of Items per Objective

	Problems to find				Problems to prove					Re-prod.	Total
	F1	F2	F3	Total F	P1	P2	P3	P4	Total P		
1. Mathematical Induction	-	-	-	-	4	-	-	-	4	-	4
2. Limits	9	11	-	20	7	4	1	1	13	3	36
3. Differential Calculus	16	10	-	26	1	1	-	4	6	-	32
4. Integral Calculus	18	15	-	33	3	-	-	-	3	-	36
5. Functions of Two Variables	6	-	-	6	2	6	-	-	8	-	14
6. Differential Equations	7	7	-	14	1	-	-	-	1	-	15
7. Series	28	7	-	35	20	10	1	4	35	-	70
8. Improper Integrals	2	1	-	3	10	2	-	-	12	-	15
Total	86	51	-	137	48	23	2	9	82	3	222

By analyzing information in this table, some important conclusions pertaining to the pursued objectives can be derived:

- Apparently, in examinations the mathematics teachers laid stress on an operationalization of the objectives mainly on levels F1, F2 and P1 and, to less extent, P2. The remaining categories ("new problems") did not or seldom appeared.
- The grading of mathematics examinations uses compensatoric model, i.e. a student gets a pass mark if his examination score is at least 55% of the maximum score. From Table 2 we concluded that a student could receive a pass mark if he disregarded preparation for problems on the level of P2, P3 and P4. This means that while higher level problems were indeed discussed during the lectures and discussion groups, the department of mathematics accepted that students needed only take cognizance of the higher level objectives, but that they did not need to master them.
- Less attention was paid to "problems to prove" than to "problems to find". Exceptions were the content categories 7 "series" and 8 "improper integrals". But in Plomp and van der Meer (1979) we have pointed out the difficulty with the categorization of the problems on series and improper integrals: distinction between F- and P-problems can hardly be made, because, for instance, in many cases a *proof* for convergence or divergence can only be given by *computing* a limit.
- The row totals reflect the relative importance of the respective content categories. Two comments have to be made: (a) category 1: mathematical

was indeed a small topic, only treated on the level P1 and (b) the numbers in categories 7 and 8 were based upon four examinations, the remaining upon three examinations.

4.4 Attainment of course objectives

In Table 3 the computations according to the method b2 in #3.4 are summarized. The results are based on a portion of the examinations (see Table 1). The raw data are presented in Plomp and van der Meer (1979).

Before discussing conclusions from an analysis of Table 3, some parameters have to be discussed.

The choice of the parameter α for mastery of a problem (see 3.2) by a student was $\alpha = .66$. The rationale behind this choice is the following. Most of the essay problems in the examinations had a maximum score between 2 and 5 points. A student who made a minor error, e.g. in the computation or in the wording of the solution, would not get the maximum score. However, we could assume he/she understood the solution of the problem.

In this case the score on a problem was often $1\frac{1}{2}$ out of 2 points, 2 out of 3 points, 3 out of 4 points, etc. i.e. a score of at least .66 of the maximum score. In #3.4 we discuss two measures to express whether the group of students has attained its goals or not. For reasons explained in #3.4 we will choose p^* , defined in that section, as measure for the attainment of the goals. This measure expresses how well the group of students, in this paper the group of passed students, has attained the course objectives.

When have the passed students attained the objectives? From Table 2 we have concluded that in the mathematics courses primarily objectives on the levels F1, F2 and P1 were pursued, and, to a less extent, level P2. The other levels were so scarcely represented in the examinations that earlier on is concluded that those levels were not pursued. Restricting ourselves, therefore, to the levels F1, F2, P1 and P2, we chose as criterion for attainment of an objective $\beta = .75$. Thus, if in a cell $p^* \geq .75$ was recorded we said that the objective was attained by the group of passed students.

Now from Table 3 some important conclusions were drawn pertaining to the attainment of the objectives by the group of passed students.

- a. All the objectives but one (viz. (3,F2)), which were attained by the group of passed students ($p^* \geq .75$) appeared to be of the type F1: elementary standard problems to find. Looking at the columns, only the group of objectives F1 was attained by the students; although one cell was clearly not attained.
- b. None of the objectives on the level of P1 and P2 was attained, although some cells were close to the criterion.
- c. A rather loose inspection of the p-value of the items within the cells showed us that a class of items is not homogeneous with respect to difficulty. For example, a F1-cell may contain relatively easy as well as rather difficult items: the classification is essentially only based on *content*, not on *difficulty*. For instance the *number* of steps that must be made to solve the problem is relevant, not the *difficulty* of each step (see the second assumption in #3.3). For clarity, in the presentation of the results to teachers responsible for this and following courses, this point must be stressed, perhaps by including a sample of items from each cell.
- d. Because the two totals in Table 3 were computed over different behavioral levels, it was not useful to apply the criterion of attainment of an objective on the row totals. From the last column of Table 3 it can be seen that there were poor results on 1. mathematical induction, 2. limits and 5. functions of two variables.

TABLE 3: Results of the retrospective analysis

	F1		F2		P1		P2		P3		P4		Total ³	
	m ¹	p* ²	m	p*	m	p*	m	p*	m	p*	m	p*	m	p*
1. Mathematical induction	-	-	-	-	4	.44	-	-	-	-	-	-	4	.44
2. Limits	6	.50	6	.55	4	.41	2	.23	1	.00	1	.24	18	.46
3. Differential calculus	9	.88	6	.80	1	.43	-	-	-	.00	2	.39	16	.82
4. Integral calculus	11	.82	8	.56	3	.57	-	-	-	-	-	-	22	.69
5. Functions of two variables	1	.83	-	-	1	.28	2	.24	-	-	-	-	4	.40
6. Differential equations	2	.72	4	.73	1	.59	-	-	-	-	-	-	7	.69
7. Series	14	.77	2	.22	10	.73	3	.67	-	-	-	-	29	.71
8. Improper integrals	1	.97	-	-	6	.68	2	.10	-	-	-	-	9	.67
Total	45	.78	26	.60	30	.65	9	.48						

¹ m is the number of problems in which the objective is operationalized.

² p* the measure for attainment of the objectives by the group of passed students (see #3.4).

³ Only based upon the F1, F2, P1 and P2 columns.

- e. Because the F1-type problems form usually less than half the number of items of an examination an interesting conclusion can be drawn with respect to the considered group of students. Although the passed students performed well on the attained problems of type F1, their score on these problems was usually not sufficient for getting a pass mark on the examination. To get the pass mark on the examination of the course every student has to master several of the objectives of which we concluded that they were not attained by the whole group of passed students. This means that teachers of following courses must take into account the fact that the group of students has attained only a relatively small part of the course objectives of courses which are prerequisites for his/her courses. This means that apparently it is the practice of mathematics education within the Department of Mathematics to accept passed students who may not have attained most of the course objectives but only taken cognizance of them.

5. Discussion

The procedure for a retrospective analysis of course objectives is not a test in a statistical sense. It is meant as an attempt to bring some order in an unsurveyable amount of examination results, so that discussions on these results will be possible. Then, discussions on the objectives of courses can be conducted using arguments which are based more upon objective data than personal feelings and impressions of teachers.

In the preceding, we have discussed the questions: (a) which objectives were pursued in the mathematics courses and (b) which of them were attained by the students with a pass mark. The answers to the first question are important for improving the considered mathematics course. The answers to the second question are important in connection with the construction of following courses of which the concerned course is one of the prerequisites. With this procedure we can indicate a general knowledge base upon which teachers of following courses may build upon.

It must be emphasized that this indication of general knowledge is only a *minimum*.

Firstly, because of the compensatorial way in which grades are established, it can be stated that all passed students know more than this minimum.

But also the attainment of a lot of objectives can principally not directly be measured by written, time-limited examination sessions. We can mention rather vague objectives like overall familiarity with the main structure of the content, ability to reconstruct forgotten knowledge, inventivity in the tackling of novel problems, etc.: skills that surely will be affected by the given instruction.

We have considered the passed students of every examination. No distinction was made between examinations immediately after the teaching of the course and examinations later on in the year (a "second opportunity"), because the students who failed the examinations immediately after the course were provided an extra opportunity. Therefore, in our choice of the subgroup of students, the "passed students", no one was counted twice.

Some concluding remarks and points for further research have to be mentioned.

1. Under the assumptions presented in #3.3 per cell, the measure of attainment p^* of an objective is computed via the transformation $z_i^* \rightarrow p_i^*$. The transformations are carried out via the cumulative standard normal distribution. From a table of this distribution can be seen that p varies almost proportionally with z outside the tails of the distribution. Thus, to get a quick impression of the results, the weighted means of p_i -values can be computed, as long as the p_i 's are not too small or too large.

2. The number of items per cell will influence the accuracy of the estimation by p^* of the true value of the measure of attainment of an objective. How to bring this element into the method discussed in #3.4 has to be studied.
3. The method presented in this paper can also be applied to examine the quality of instruction. Then, the measures of attainment of an objective have to be interpreted as measures for the quality of instruction. The examinations immediately after teaching of the course have to be analyzed for *all* the students who took them.

Acknowledgement

We owe many thanks to Dr. Wim J. van der Linden and Dr. Robert C. Harris for their willingness to contribute to this paper by intensive discussion of the theoretical part and critical comment on the manuscript.

REFERENCES

- BLOOM, B.S. (Ed.) *Taxonomy of educational objectives: The classification of educational goals; Handbook I: Cognitive domain*. New York: McKay, 1956.
- PLOMP, Tj., & VAN DER MEER, A. *A method for a retrospective analysis of course objectives*. (Twente Educational Report No. 7), Enschede (Netherlands): Twente University of Technology, Department of Applied Education, 1979.
- POLYA, G. *How to solve it*. Garden City (N.Y.): Doubleday, 1957.
- WILSON, J.W. Evaluation of learning in secondary school mathematics. In Bloom et al. *Handbook on formative and summative evaluation of student learning*. New York: McGraw-Hill, 1971.

THE AUTHORS

TJEERD PLOMP graduated as mathematician and wrote a dissertation, in education, on individualized instruction in higher education. After teaching mathematics to engineering students during 12 years, he is now engaged in the preparation of the new Department of Educational Technology at Twente University of Technology.

ADRI VAN DER MEER is a graduate in mathematics from Twente University. At present he is teaching mathematics to engineering students. His research interest is mathematics education, especially applied mathematics.