

CHAPTER 7

RELATIVE MEASUREMENT AND THE SELECTIVE PHILOSOPHY IN EDUCATION

Egbert Warries

Technische Hogeschool Twente, 7500 A E Enschede, The Netherlands

ABSTRACT

Teachers in different countries, when judging the achievement of their students, do not use absolute standards but seem to work by comparison. The paper describes the reason for and the philosophy behind relative judgment in education. The explanation for the frequent use of relative testing in education is that it has a selective function in the classroom. The selective approach does fit in an educational philosophy regarding the place of schooling in our society. The philosophy does not possess a theoretical framework nor a clear description of its content. Six statements typical of the selective philosophy are discussed. They relate to the paucity of jobs on the market, to competition, the distribution of talent, and the supposed attributes of good school tests. The statements are related to each other in order to demonstrate how they can form a coherent system of opinions about how schooling ought to be organized. It is concluded that no philosopher can solve the problem of the educator who realizes that he is measuring instead of educating. Finally, it is argued that the problem of setting up a non-selective organization is a technological matter - a matter of choice for other learning strategies and other measurement techniques.

INTRODUCTION

What kind of philosophy of education ultimately is responsible for relative measurement of learning outcomes? Educational researchers in different countries have noted that teachers, when judging the achievement of their students, do not use absolute standards but seem to work by comparison. The grade assigned to a student's piece of work seldom seems to be connected with an operational definition of the relevant learning goals. In most cases, it appears to be the empirical distribution of the raw student-scores which determines what is considered excellent, moderate, and insufficient. It is

found that, on the average, teachers give failing marks to a fixed proportion of the group taught by them. For Holland, Posthumus (1940, 1960) reported that the mean percentage of failing marks in the yearly student reports in secondary schools is always approximately equal to 25%. For the Dutch universities this observation was borne out by Drion (1960), who speaks of "Posthumus Law" to denote the phenomenon. De Groot (1966) studied the working of this law and found that this 25% remains the same in all six years of pre-university education but that the percentage usually differs for the different courses. In general the difficult courses or the courses with more status in school had more insufficient marks than the rest. Thus, algebra with a percentage of 31% had about five times the number of insufficient marks as history. Bloom (1968), writing about the expectations of teachers at the beginning of a new course or term, states that the instructor expects a third of his pupils to fail or just "get by". Warries (1978), in a treatise on the yield of university education, points to the grading habits of staff members and their influence on the output of the university. Many authors have called for attention to the marking or grading habits in our educational institutions. We seldom find absolute or criterion-referenced testing and we often find relative testing methods in schools. In this paper, we will try to describe the kind of philosophy which is hidden behind the persistent phenomenon of relative judgment in education.

AN EXPLICABLE PHENOMENON

It is undeniable that the current practice of relative measurement has withstood many challenges and criticisms such as that by Glaser (1963). In everyday school and university life, teachers do not frequently use absolute or criterion-referenced methods. Relative judgment is the rule in education, and probably it is wise to consider first the most plausible explanation for this phenomenon, viz. the psychological necessity of such behavior, before going into the drawbacks of relative measurement and the philosophy behind it. In our opinion, people, by nature, tend to compare. Or, as taught by most introductory texts in our psychology courses, human beings tend to impose organization upon what they perceive. Hilgard (1953) says: "We see faces in the clouds". By selecting aspects out of total situations in our environment, we tend to create figures. This process also goes for measuring learning outcomes in the classroom. The total situation can be the essay-test of one student as well as the sheet of raw scores of a group of students. In the former case, we compare the visible outcomes of our teachings with the goals or expectations we had as teachers. In the latter case, our perception takes a different turn because there are more things to compare with. We tend to impose structure on the set of scores, and select certain elements from the total situation when describing the outcomes of our teaching. Time and again we perceive a middle group of moderate achievers, a small group of excellent achievers, and we clearly discern another third or quarter of insufficient scores. Unlike when looking at the clouds, we are not able to see faces in the given situation, but we certainly do impose structure upon our data along the one dimension that we have at our disposal: the

achievement scale. Something in favour of others and ourselves as teachers and as comparative judges of achievement should be added: it is only normal for human spectators to compare elements in a set and to differentiate among the characteristics of different entities, even when these entities are students or their products in a class. What follows aims at making the reader aware of his comparing judging behavior and asks for more attention to the philosophy which is suspected to be behind that behavior.

INSTRUCTIONAL DRAWBACKS OF RELATIVE TESTING

Aside from nice psychological explanations, it is still possible to grumble about that bad habit of teachers on all levels of schooling, namely their persisting inclination to compare learning outcomes before assigning marks or grades. We believe that the essential argument against relative testing is that it belongs to an outdated and sometimes hidden philosophy of education. In fact it could be defended that it forms the visible part of an obsolete outlook on schooling in the modern western world. But let us first look at some practical drawbacks of relative measurement. Doing so, we restrict ourselves to measurement of learning outcomes by educators in normal educational settings, and will not speak about testing as it is used for institutional selection or admission purposes. Neither shall we speak about programmed instruction where it stands to reason that criterion-referenced measurement for the people concerned will have preference over relative measurement. We shall not deal with these two measurement cases because in programmed instruction the choice for some absolute testing method is obvious, and, on the other hand, in the case of admission, a rational decision has to be made about the ranking of applicants on attributes considered relevant for the program in question. So what follows below has mainly to do with grades or report-marks students get for their results on a test paper, a final test, or three months of classwork grades which are assigned to them by their subject-teacher, class-teacher, or professor, and which must inform the student and his parents about the student's progress in school. Now, in this ordinary situation, what are the drawbacks of relative measurement? Three only will be mentioned here.

The first drawback of report-marks based on comparison of the student's test results with those of the others is the lack of information for parents interested in their child's progress in relation to the content of the course. To be sure, there indeed is valuable information in a grade representing the relative achievement of a student, if parents are at least in the possession of additional data needed to interpret relative scores. When parents do know about the average class performance, the quality of instruction, and how these two relate to local or national standards, they sure can make a realistic estimate of the educational prospects in the near future for their daughter or son. But in most cases parents do not possess the extra information, or they have wrong information. In many cases, parents ask questions about their child's result of half a year of school work or they ask what the prospects in the long run are, but relative testing does not have answers to those

kinds of questions. This really is a serious problem. Relative tests do not provide for relevant information for parents. In a small country such as Holland, with cohorts of less than 250,000 individuals, it was not tried to solve this problem by the introduction of nationwide standardized tests. This can be considered fortunate, because in that case we should not only still have had the wrong answers to the parents' question, but we should also have lost the freedom to criticize the tests as they were made then by highly qualified experts.

The second drawback is similar to the first, but here the concern is with the information given to the student himself. What kind of message does a student receive through a single comparative score? Does the score give him any advice on how to continue next week, next month, or the following trimester? Does the score specify for the student how to reach a higher score next time? Must the student who failed his test try preparing himself in another way for the next test? The answer must be negative since his rank order in the group at best signals him to work harder for the next test, do more of the same, and do more than his colleagues. In most cases a score or mark does not tell the student how far he is in his work and what he has to do to better his position. Nor does the student know if he would have gotten the same mark for his work, if he had been in a different group or had had a different teacher. In short, a relative scoring method does not inform the student how to improve his work in school and his test results.

The third drawback is more pedagogical. We refer to the negative effects on the individual who continuously belongs to that part of the group to which low grades are given. Educational psychologists in the past have given due consideration to individual differences in relevant learning abilities, but gradually we personally have come to the conviction that in most instructional settings those differences between individuals are relatively small in comparison with the psychological similarities of the group members. But, notwithstanding the many abilities and aptitudes that members of most instructional groups have in common, we also have to face reality and to admit that in our present schooling system there do exist consistent differences in achievement between the members of the group. Now what kind of message does a rank-order grade bring to the slowest student? The message leaves no doubt. It reads: "You, low achieving student, shall never do well". Such tidings can hardly do any good for the learning process. Roughly, there are two possible translations or conclusions for the unhappy receiver of this information, and these are well known in educational circles. The one is: "I am a worthless person", and the other is: "Learning is no good at all". So in the first case the student is liable to see himself in general as somebody less valuable than other people. In the second case, the student is developing a negative attitude towards all forms of transfer of knowledge in both informal and formal settings. There is hardly any doubt that in both cases a student will be severely hindered in his future school career (if, in the circumstances, there is any left for him).

ACHIEVEMENT TESTS AND SELECTION

Why is it that, despite the drawbacks mentioned, relative measurement still is the most widely spread method of measuring student's achievement in the educational settings discussed? In this paper, the explanation of the phenomenon is that, in general, the choice of a testing method reflects the educator's or the school's view of the place of school in society, and that in this particular case the educator believes in a selective function of the school. But others give a different explanation. The alternative explanation holds that relative measurement takes place so often simply because teachers have at their easy disposal a number of relatively cheap objective tests which all give the teacher the opportunity to rank students. Three objections to this explanation can be made. The first objection is that tests for sale in the shop also have unattractive properties for classroom application. For example, they seldom match the content of the course as taught by the teacher or represented in the learning material. Second, the observation that these tests are cheap for the teachers is only part of the story. When viewed locally or nationally, professionally made tests are not cheap at all when all the academic and developmental work done to produce these tests is considered. Thus, the cheapness of the tests cannot be an explanation for the relative measurement habits of educators. The third objection is that we can observe that relative judgment habits also do exist in countries where no objective tests are available for schools. In the Netherlands we do not have a testing movement like that in the United States but still we have our Posthumus' Law which states that educators judging educational achievement do compare individual learning outcomes. Apparently, test papers and essay-type tests also are fit for comparing achievement in a group. Summarizing, it can be concluded that the mere availability of objective tests is not a necessary or sufficient reason for the existence of relative measurement. There has to be some other reason.

It can be stated confidently that the explanation for the frequent use of relative testing in education is that it has a function in selection which in turn is considered by many of us to be one of the main functions of educational institutions. What is meant by selection here? Educational scientists, when categorizing tests according to the way they are constructed and used, tend to speak about the selective use of tests in a very restrictive way. The selective use of tests in education, according to the experts, seems to be synonymous with testing for admission to institutions or programs. But next to this definition of selection, conceptualized as a neatly arranged and systematic enterprise, there is the classroom use of tests. Around us we can see many instances of informal, teacher-made, non-standardized tests with an obviously selective character. Perhaps we should first purge from our thoughts the systematic use of professional testing if we want to reflect on the functions of testing inside school. In fact, this is what a small group of educators and researchers - among them, the author of the paper - did when in the mid-sixties, the first large-scale multiple choice testing program for the transition from elementary to secondary education was introduced in Holland. We asked ourselves why, in general, within the walls of the school, testing takes place. We did this because, after we all had been

engaged in the introduction of these new objective tests shortly before, we felt worried about the consequences in the long run for the schools. Our purpose was to make educationists aware of the real reasons for testing in the classroom. We came up with three reasons (Warries, 1971), and selection was one of them (and not the least important). With selection we did not mean admission to or entrance into a program. We had in mind a more gradual process. Something that takes place little by little and in which students belonging to the part of the group that is learning more slowly are separated from the class or, in Holland, have to double class. Tests used for selection were described as tests that students look upon as difficult tests, and also as tests for which preparation cannot be made. In our description, students feel that even by working harder or longer, they can never be successful with a selective test. Once a student is among the low achievers, it is a losing battle. Teachers, on the other hand, consider a selective test as an excellent means to differentiate between what they see as poor students and good students. This description of the selective use of testing in the classroom can be seen as a typical description of the classroom use of relative measurement. In relative measurement, raw scores are ordered and subsequently transformed to some type of standard scores forming a distribution in which scores are spread in such manner that a restricted part of the scores qualifies for the predicate "excellent". Students are right when they consider such a test as difficult, and they are right when they do not believe it is possible to prepare for such a test. Teachers also are right when they see these tests as well-suited to differentiate between students' scores. What we are trying to say is that, if we have the impression that relative testing procedures have too many drawbacks to maintain them as a means for judging the achievement of our students, we have to look sharply at the selective function of relative testing. It is only when we look at the selective processes taking place in our educational institutions that we shall be able to understand why an alternative like criterion-referenced testing is so seldom used in the schools. When selection indeed is the everyday practice of our schools, then this practice is only maintained because it fits in an educational philosophy regarding the place of schooling in our society. Let us look at this alleged educational philosophy.

A SELECTIVE PHILOSOPHY

In this paper, education is also examined from the teacher's point of view. For those who look at education from the "inside", a philosophy of education is a way of thinking about how schooling in a given environment ought to be organized. In such a way of thinking there is hardly a place for clear sociological notions on the functions of schooling in society or for normative ideas about the ends of educational processes. The selective educational philosophy under consideration does not seem to possess a strong theoretical framework; neither is there a clear description of the philosophy in textbooks used by prospective teachers. In fact, every new teacher seems to understand this philosophy without being trained to do it. Actually, the philosophy consists of statements only, conceptually more or less independent

but strongly interrelated when they are situated in what we consider to be a selective philosophy. Some five years after introduction of objective testing in Holland, the author wrote about relative measurement in the classroom and recorded six utterances which he considered typical for a philosophy of permanent selection in the educational process (Warries, 1970):

- many are called but few are chosen;
- competing for better scores is wholesome;
- in each group some must fall behind;
- school tests separate the dull from the clever;
- school tests must be difficult;
- students should not be informed in advance about school tests.

We will comment on these six statements and will try to relate them to each other in order to demonstrate how they could form a coherent system of opinions about how schooling ought to be organized.

Many are Called but Few are Chosen

This patriarchal utterance from a spokesman of an educational institution toward the candidates to be taught or who seek admission to courses or the examination, seems to carry the message that candidates have to be modest. They ought to know their place, should not expect too much, have to study hard and should understand that it is not a matter of course that all will complete the full program. The admonition is sometimes accompanied by information on the paucity of places in the market for which the institution prepares. Often attention is called to the fact that standards are high and that the school has to be rigorous in its demands.

Competing for Better Scores is Wholesome

Usually, two arguments are advanced to support this statement. One is that competition to get the best or better results in class is a very useful extrinsic motive for learning. Students, according to this line of thought, need an external challenge to work hard since the intrinsic rewards of school learning are not enough for them. Students who know that their score will be calculated after comparison with the outcomes of the whole group will strive after a higher place in the rank order. The other argument is that competition in itself is a good exercise and a learning experience in preparation of the great competition which is said to wait for the students after school.

In Each Group Some Must Fall Behind

Why is this so? Because this repeatedly turns out to be so. In each instructional group there appear to be some students who are the duller ones. They cannot keep pace with others, and the teacher does not have the means to help them out. Though these slower students often ask more of the teacher's time than the others do, seldom are they successful in catching up. Every teacher sooner or later discovers this truth and learns to live with it. Most educators indeed believe that some must always fall behind since these are the less talented and there seems to exist no policy preventing them from gaining admission. In fact, selection specialists admit that they cannot provide a procedure to safely predict academic failure.

School Tests Separate the Dull from the Clever

That means that a good test must make such a separation, since, according to this statement, an achievement test must give information about the fitness of the students to proceed with the study. So success on future work is predicted by success on past work, as appearing in the test results, and, therefore, every school test must give as much information as possible. That means that the test has to produce large distances between students' scores. The more differentiation, the more information for the teacher. A large range of scores is better than a small one, and a flat distribution of scores is better than a bunched one.

School Tests Must be Difficult

Every country seems to have its own folklore about the difficulty of final examinations, and there hardly exists an institution where the finals are not dreaded by the students. School subjects easy to learn are preferably not entered into the final examination, and predetermined norms that accidentally result in a high percentage of excellent scores are revised in order to produce the usual proportion of fail-grades. The same goes for testing inside the classroom. There seems to exist a general feeling among policy makers and teachers that school tests ought to be difficult. A test that is done well by all the students in a class cannot be taken as a good test. In many cases, the difficulty of the test and the marks to be given are established only after the results are known. Sometimes, the latter is done systematically by splitting the distribution of raw scores, using, e.g. percentages of 10, 20, 40, 20 and 10%. In this manner tests being too easy are avoided most effectively. When such "grading on the curve" is not applied, many teachers nevertheless do not set the pass-fail cutting score until they have been able to look at the results of their students.

Students Should Not be Informed in Advance About School Tests

Many teachers are convinced that the content and wording of a good test must remain a secret so that the students cannot prepare themselves for it. In any case, they must not be able to prepare in such a way that they are sure to know all the right answers. In this opinion, a good test should not bear reference to the detailed contents of the course taught. It should consist of items that can be answered right only by those who not only know the content of the book, but also know how to use their heads. In this notion there is something of a battle between school and students. Both parties try to win by being smarter than the other. A good test item is one invented by the teacher in a brilliant moment, and a bad item is one of the type students can answer easily. A teacher who beforehand reveals what kind of questions he is going to ask, would break the rules of the "game". His colleagues, and probably even his students, would consider him a bad sport.

A COHERENT BELIEF SYSTEM

In what way do these six statements relate to each other? How do they form a coherent belief system for the people working inside a school or for those identifying with them? In the center of the belief system is the practice of continuous selection. Teachers and policy makers in a selective institution frequently compare and sort their students. This is done by decisive judgments on the students' achievements which the students interpret as a negative decision to go on with the program they are in. They are assigned to a different program which is of lower status, or they are rejected without any further recommendation. When, in an instructional group, such a procedure has been practiced for a period of three to six years, the final group is only a fraction of the original group and few are chosen indeed. Those who are in the midst of such a procedure need a rational basis for their actions and they will find it in the first adage which states that few are chosen. This adage forms a sound basis for selection since it stands for the belief that there simply are too few places for all students, certainly with the standards we want to use. It functions as a central rule in a selective policy. The institution (the people involved, say) is obliged to select because society, the university, or another institution cannot take them all. Of course, the next question to be answered is why society or the next institute cannot take all who have been taught. Here we know of two answers. The first one actually lays outside the belief system under consideration and is mentioned only briefly. It is the explanation of sociologists who hold that in a capitalistic society places deliberately are kept scarce so that the benefits of the good positions do not have to be shared with so many. The other answer is that only part of the group is placeable since it is a simple empirical fact that some will fall behind at every measuring point in the course, since they eventually will not meet the final standards. Apparently, some law is governing human behavior stating that some must fall behind in instruction so that, indeed, after some repetition of

the phenomenon few are left. This seems to be a plausible explanation of the statement that few are chosen. On the other hand, the supposed truth that some have to fall behind in everyday school life has to be translated into a measuring instrument. There will be a need for a good instrument to measure who is behind. The instrument may be simple, for instance, some questions posed by the teacher, or it maybe a well constructed, sophisticated achievement test. In any case the desired instrument must indicate who will not survive in the educational process by separating the dull from the bright students.

But when a good test to separate the dull from the bright is desired, two technical conditions should be fulfilled. One condition is that such a test must differentiate very well. The other condition is that testing knowledge of the course content is not sufficient, but also some general characteristic or behavior responsible for the learning outcomes must be tested. So, tests must be difficult enough, and they must be such that students cannot prepare for it by learning what is in the book. Why must tests be difficult? It seems as if everyone inside a selective institution is of the opinion that tests have to prove that some fall behind. In contradiction to what many believe, the standards are seldom absolute, criteria for success are seldom fixed, and final examinations seldom have norms that are set with reference to the desired behavior.

The author knows of one case where circumstances made it necessary to set norms before a national examination. In 1968 the Dutch national testing institute constructed a fifty-item English reading comprehension test for the first final examination of a new type of secondary school. After careful consideration, the committee responsible agreed on the test, but afterwards the committee was blamed because the mean raw score, which turned out to become 43.6, was considered too high. One year later, the population mean was 34.0 which proved that the committee had learned their lesson.

A good test is a test difficult enough to reliably split the group in dull and clever students. But, apparently, the difficulty dimension alone does not suffice to weed the poor students because hard working dull students can also reach the goals of course. Since we try to separate the dull from the bright and not the non-learners from the learners, we thus need to test the potential for learning and not only the result of learning. What is it that teachers have in mind when they say that a really good test cannot be prepared for? Are they conceiving of an aptitude specific to the school subject? A general learning aptitude? An innate or developed general ability? Whatever it may be they are after, consciously or unknowingly, the second technical condition is fulfilled. Something else other than learning output is measured.

Three further remarks are to be made on the internal relationships of the belief system. First, the custom of comparing students' learning outcomes to evaluate them seems to need vindication. This is found partly in the supposed merits of competition between members of a group. It is hard to accept, but in this technological age where teamwork is so important everywhere, there are philosophers telling us that, apart from being a

stimulus for learning, competition is also a good exercise for the great struggle in life after school. So the use of relative measurement procedures is reinforced by the alleged benefit of providing an extra learning experience. The second remark concerns the role of general intelligence in the measurement of learning outcomes. Behind four of the six statements given above there must be an hypothesis stating that some type of (normally distributed) general ability is responsible for student achievement in the classroom. Some students ought to have insufficient results since some have to be in the least talented part of the group. The detection of the dull students comes up as a task for the test because the educator believes in the existence of a normally distributed variable measured on a dull-clever scale. There cannot be much difference between such a variable and general intelligence. In any case, the notion of general ability can be detected in the belief that students must not be able to prepare for a test. Also, it is only one step from the belief that school tests have to be difficult and have to show large variability of scores, to the belief that performances on school tests are reflecting the distribution of a developed ability in the group. That some learners must always fall behind is easier to understand when it is believed that some general ability is mainly responsible for learning outcomes.

There is an implicit assumption in some of the selective opinions under discussion. The assumption is that performance on school tests is related more to some innate or developed potential of the student than to the deliberate efforts of schools and teachers. The last remark pertains to the reversibility of the relationships within the belief system. Selection is also caused and maintained by the belief that competition is wholesome. A part of the class has to fall behind just because of the way achievement tests are constructed and scored. There are always dull students to be discovered in a group when the definition of achievement leads to that conclusion. The conviction that few can be chosen holds since so few graduates are ultimately produced by the schools. Thus, parts of the belief system not only provide a rationale for other parts, but they also strengthen the other parts. Actually, the whole belief system strengthens itself as long as its different parts hold out.

CONCLUDING REMARKS

Is there a way out? Probably there is no way out from within. When you are in the midst of a system, it does not help to realize that you are working for a capitalistic society, is subject to laws of perception, you are measuring instead of educating, and that human beings are more than competitors. When one suspects that his selective philosophy is a negative philosophy that only serves his selective practices, then a different practice should be tried, that is a different organization of the components in an educational institution. Starting from the concept of a different organization, the next step is to try to develop a rational solution to the problems stated by the blueprint of such an organization. No philosopher can solve your problems, neither can sociology, or cognitive psychology. The problem of setting up a

non-selective organization is a technological matter, a matter of choice for other learning strategies and other measurement techniques. It seems to us that a strategy like mastery learning (Bloom, 1968) and measurement methods such as criterion-referenced measurement give improved solutions to the problem. The above treatise on a selective philosophy may be forgotten as soon as we are working on the solution of practical problems.