DOCUMENT RESUME

ED 294 759 SE 049 163

AUTHOR Connell, Michael L.
TITLE What Counts in Elementary Mathematical Evaluation?
PUB DATE 88
NOTE 19p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April, 1988).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Educational Research; Elementary Education; *Elementary School Mathematics; *Evaluation Methods; Informal Assessment; Intermediate Grades; *Mathematics Achievement; Mathematics Education; *Mathematics Skills; *Mathematics Tests; Measurement Techniques; *Student Placement
IDENTIFIERS Mathematics Education Research

ABSTRACT

When standard paper-and-pencil tests are used to measure progress in mathematics, specific assumptions quite apart from those dealing with the statistical validity of the test are often made. These can lead to improper student treatment and often suggest successful teaching which is not warranted. The results of examining 32 fourth through sixth grade students attending a predominately middle class, urban school indicate four distinct groups of students that are not identifiable by these measures. The differences between these groups, although insignificant statistically, are considered by some to be of great importance in the creation of effective learning environments. (Author)

What Counts in Elementary Mathematical Evaluation?

Michael L. Connell

University of Illinois

Presented to the American Educational Research Association. New Orleans, Louisiana, April, 1988.

Running Head: MATHEMATICAL EVALUATION

Author Identification Notes

I gratefully acknowledge the assistance of Donald M. Peck, Peggy Sorenson and the staff of East Sandy Elementary for their assistance in data collection; Becky Davison, Rita Gaskill and Raleen Connell for their help in manuscript preparation and Jeff Instefjord for comments on an earlier version of this paper.

Requests for reprints should be sent to Michael L. Connell, who is now at the University of Illinois, 51 Gerty Drive Room 199, Center for the Study of Reading, Champaign, Illinois 61820.

Abstract

When standard paper and pencil tests are used to measure progress in mathematics, specific assumptions quite apart from those dealing with the statistical validity of the test are often made. These can lead to improper student treatment and often suggest successful teaching which is not warranted. The results of examining 32 fourth through sixth grade students attending a predominately middle class, urban school indicate four distinct groups of students that are not identifiable by these measures. The differences between these groups, although insignificant statistically, are of great importance in the creation of effective learning environments.

What Counts in Mathematical Evaluation?

When standard paper and pencil tests are used to measure progress in mathematics, specific assumptions quite apart from those dealing with the statistical validity of the test are often made. These assumptions can play a large role in the creation of teacher misconceptions in regards to their effectiveness in instruction and lead to mislabeling and inappropriate treatment of students.

A primary underlying assumption made in the use of such instruments is that those children able to respond in a form commensurate with the answer key are in control of the tested operation and those unable to do so are not. As Campione, Brown and Connell (In press) point out, this simple assumption is currently the subject of some controversy. Unfortunately, why this should be a controversial belief is not at all obvious to most users of the instrument. It is only when the implications of this statement are examined that the paired hidden assumptions contained within it are identified.

Central to this examination is what it means for a student to be in control of a mathematical operation. Although a full discussion of what is meant by control is beyond the scope of this paper, it seems that some further description would be warranted. This paper takes the position that being in control requires meeting at least two criterion. The first, that the student be consistently and efficiently able to produce the correct answers. The second, that the student possess understandings in regards to how these answers are produced. (For a more detailed analysis of conceptual control see Schoenfeld, 1983, 1985.)

Looking at the original assumption from this perspective, it becomes clear that of these two criterion, only the first is addressed by normative forms of evaluation. A student's scores on such instruments most accurately reflects their skill in the application of an appropriate process. What is often assumed by the test user is that concept is being measured concurrently. This is rarely the case. Indeed, Burns (1986) has argued that measured success in computation often serves to mask the students lack of higher level understandings.

This paper will attempt to show that this hidden assumption creates an interpretation of testing results which is inaccurate. Further, when applied in the classroom these results have a profound impact upon the educational opportunities and placement of the student.

In this paper the term *process* will be used to refer to the schema, algorithm or method utilized in the computation of an answer. In keeping with our earlier definition, control at a process level would involve the efficient and accurate use of this process in the computation of an answer. *Concept* will refer to the underlying assumptions or mental modeling upon which the process is built. Control at a concept level would be present if these underlying assumptions or mental models accurately reflect the mathematical situation.

Within this convention, let us create four hypothetical groups of students who either meet or fail to meet the dual criterion of process and concept.

Insert Table 1 about here.

Group one students possess a process adequate in dealing with the problem under examination and understand the underlying concepts for this process. For this group, at least, the assumptions behind the interpretation of test scores are accurate. Those in group two, although able to deal with the mechanics of the problem, lack understanding of conceptual underpinnings. Whitney (1985) gives a typical protocol of such students, "In a *school problem*, you just guess what operation to use with given numbers." Group three students would lack an adequate process as well underlying concepts for the problem type. Those in group four lack an adequate process, but do possess a concept appropriate for the problem type.

In this paper, concept judgments were based on student products such as physical manipulation of objects, sketches, diagrams, flow charts, verbal descriptions and analogies. In making these decisions, efforts were made to balance a sensitivity to visual imagery with language as suggested by Dawe (1984).

In using this scheme it was often the case that many students used an alternative conceptual framework to that formally presented in the classroom. Many of these were surprisingly effective for the students in meeting the needs of the problem. As Vakili (1985) observed, such alternatives became more common in the more complex problem types. As before, judgments of such alternative frameworks were based upon their accuracy in meeting the demands of the problem setting.

Methodology

A study was conducted utilizing 341 fourth through sixth grade students of an urban school of 765 serving a predominately middle class population.

The first step was the administration of a standard paper and pencil test to the subjects. This test was selected to meet two criterion. First, the test must follow an easily recognizable format so that the results of the study would be an effective demonstration. Second, the test must play an important role in the placement of the subjects taking it.

The test selected for use had been developed by the local school district as a placement tool for a goal based instructional sequence. In content, the test consisted of multiple choice problems utilizing the basic operations as applied to whole numbers, fractions and decimals. The district in their efforts to follow typical standardized testing format had developed an instrument strikingly similar to tests such as the Iowa Test of Basic Skills and the California Achievement Tests.[1]

This test was then given to the entire fourth through sixth grade population and scored. The students were placed into groups on the basis of their sub-scores in each of the tested areas following the guidelines of the test. In the directed use of this test, the earlier outlined assumption was clearly in force. Test scores were the sole authority to identify student mastery of the problem type and eventual placement.

Obviously, mastery of the problem type is a very slippery term. As alluded to earlier, it might mean has full mastery of the concept and the process. Possibly, the correct process was chosen based on superficial factors that although successful in achieving the correct answer do not reflect upon the underlying concept. Other possibilities include the attainment of the correct answer by blindly adhering to a misunderstood rule or by compensating errors.

The next step was aimed at assessing the accuracy of the tests placement. Teachers were trained in the use of a clinical interview as outlined by Peck, Jencks and Connell (in press) in making decisions regarding possession of concept.[2]

Following this, 32 students were randomly selected for interviews and additional placement. The initial placement of these students was created by using the test's results. This created the student placements as shown in Table 2.

Insert Table 2 about here.

A significantly different picture emerged when the additional data from the interviews was taken into account in the student placement.

Using the comparison groups defined earlier the students were placed according to Table 3.

Insert Table 3 about here.

It is important in looking at this data to bear in mind what is required for the test to be accurate. Using the test as sole measure of progress, the inference is made that those in groups one and two are successful and those in three and four are not. Subsequent placements would then be made on this basis. It is arguable that this assumption is accurate only for groups one and three, thus misplacing those students in groups two and four.

Using the case of addition and subtraction of whole numbers as an example, the assumption would have been that 28 students (the combination of Group 1 and Group 2) had mastery and 4 (Group 3 and Group 4) had not. Looking at the interview data shows that 13 students in Group 2 and the student in Group 4 would have been misplaced. For this particular example, 14 students would have been misplaced on the basis of the test alone. Of the 192 placement decisions which were made on the basis of the test alone, over 41% were likewise in error.

ANOVA results were then determined for each test sub-score to ascertain if there was a significant statistical difference between groups. In each of the areas measured by the test, a significant difference between groups ($.01 < p < .05$) was found. To further identify the source of this variance, contrast comparisons were computed for each sub-score showing a significant difference. In each case a statistically significant difference ($.001 < p < .01$) was observed between groups one and four, groups one and three, groups two and four and groups two and three. The differences between groups three and four, as well as the differences between groups one and two, were insignificant statistically.

Although the differences between members of group one and two and those between groups three and four were not observable statistically, they were patently obvious to those conducting the interviews.

Group one students were consistently able to quickly identify the goals of a problem solving situation. They were capable of activating existing knowledge in novel arrangements to meet the needs of the problem situation. They evaluated the results of their approaches in terms of the problem and identified effective sequences to be used again. For these students, the testing assumptions were clearly valid.

Group two students, despite their success in solving the problems, often spontaneously commented that they did not understand what the problem was about. In working problems, they appeared to be compensating for lack of understanding by placing greater reliance upon short term memory concerning specific problem instances and surface features of the problem. Students in this group were unable to identify simple variants of problems and instead treated them as unique.

In examining the performance of the test as a whole, the greatest accuracy demonstrated was the identification of failure. Because of this, group three students were quite accurately placed. These students showed definite needs for work in the concepts and processes evaluated.

In observing the students in group four, however, they were found to have many of the same characteristics of group one students. They were able to identify the goals of a problem solving situation, although often taking longer to do so. They used existing knowledge to attempt to meet the needs of the problem, but were inefficient in their processing of this information. Unlike the students in group one, they often made careless errors in computation or failed to check their answers.

It was interesting to note that the number in this group remained relatively small, comprising between 1 and 5 students. One possible reason might be that since they already have a conceptual understanding of the material, they would benefit from rote drill and practice instruction. As this is a common form of remediation, their needs would be filled quite often.

Several observations seem justified on the basis of the descriptions above. First, the test was unable to distinguish between conceptual and inadvertent errors. Thus, it did not distinguish those students responding in a rote mechanical manner from those who ..nely understood what they were doing and able to interpret their ..s. The tests suggested successful teaching which was n..t warranted in terms of the students ability to address conceptual issues as would be essential in making interpretations and solving problems.

Summary

In many American schools mathematical placement consists of paper and pencil tests which provide only one type measure - the ability of the student to accurately process the problems given. For many purposes this is all that is needed, but the scores usually yield no information regarding other questions crucial to the formation of a productive educational environment.

We must look at the reasons behind why, as educators, we bother to test. Do we give a test to separate the students into those who pass and those who fail? Or are we trying to find out in what areas the student needs additional work for success? If we are testing for the latter we must have insight into what will best aid in the student's growth.

With the results of standard forms of testing we are often unable to distinguish students receiving the correct answer with full conceptual understanding from those students who either made a lucky guess or are totally rule dependent.

But perhaps of primary importance, we are unable to ascertain what concepts successful and unsuccessful children are using. This latter knowledge can be invaluable in determining the domain knowledge, used heuristics and control strategies utilized by the students. As argued by Collins, Brown and Newman (in press) such information is critical in the design of effective learning environments.

References

Burns, M. (April, 1985). Teaching "what to do" in arithmetic vs. teaching "what to do and why". *Educational Leadership*, pp. 34-38.

Campione, J. C., Brown, A. L., & Connell, M. L. (In press). Metacognition: On the importance of understanding what you are doing. To appear in R. I. Charles & E. A. Silver (Eds.), *Research agenda for mathematics education: Teaching and assessment of mathematical problem solving*. Hillsdale, NJ: Lawrence Erlbaum.

Connell, M. L. (1980). [Validation and Reliability Studies] Unpublished Raw Data.

Collins, A. L., Brown, J. S., & Newman, S. E. (in press). Cognitive apprenticeship: Teaching the craft of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Cognition and instruction: Issues and agendas*. Hillsdale, NJ: Erlbaum.

Dawe, L. (1984). Language, visual imagery, and intuition in mathematical thought. *Research in Mathematics Education in Australia*, May, 41-46.

Peck D. M., Jencks, S. M., & Connell, M. L. (in press). *Improving instruction via brief interviews*. Arithmetic Teacher .

Schoenfeld, A. H. (1985). *Mathematical problem solving*. Orlando: Academic Press.

Schoenfeld, A. H. (1983). Beyond the purely cognitive: Belief systems, social cognitions, and metacognitions as driving forces in intellectual performance. *Cognitive Science, 7*, 329-363.

Vakali, M. (1985). Children's thinking in arithmetic word problem solving. *The Journal of Experimental Education, 53*(2), 106-113.

Whitney, H. (1985). Taking responsibility in school mathematics. *Journal of Mathematical Behavior, 4*(3), 219-235.

Footnotes

[1]This is not intended as a particular criticism of these named instruments, but rather to provide an example of the format of the pre-test. Experience has shown that performing evaluations similar to those described in this paper provide consistent patterns, regardless of the source used as pre-test.

[2]The problem of differences between teacher placement had been addressed in an earlier study. Connell (1980) examined this problem through two experiments. In the first, interviews were recorded on audio tape and all notes, sketches, and diagrams used by the student were kept. At the end of four weeks the teacher was again presented with the same interview and asked to regroup the student. At this same time, the identical data was presented to a different teacher with no experience with the student involved. This teacher was also requested to group the student. The results of these experiments showed a 91% agreement when the same teacher grouped a student and an 83% agreement when different teachers scored the same interview.

As part of this same study, internal reliability estimates were made on the basis of eleven sets of interviews, each consisting of thirty two students. Depending on the measure used these estimates ranged from a low of .778 to a high of .873. Admittedly, data of this type does not guarantee the validity of these techniques, but it does provide supporting evidence concerning their consistency and reliability.

Table 1

Concept/Process Groups

	Has Concept	Lacks Concept
Has Process	Group One	Group Two
Lacks Process	Group Four	Group Three

Table 2

Test Placements

	Predicted by Test					
Passed	28	26	18	14	15	13
Failed	4	6	14	18	17	19
	+,-	X,/	+,-	X,/	+,-	X,/

Whole Numbers---Fractions---Decimals---

Table 3

Group Placements

	Number of Students					
One	15	10	7	6	8	6
Two	13	16	11	8	7	7
Three	3	5	9	14	14	16
Four	1	1	5	4	3	3
	+,-	X,/	+,-	X,/	+,-	X,/

Whole Numbers---Fractions---Decimals---