

Assessment in one Learning Theory Based Approach to Teaching

Ed Dubinsky, Georgia State University

I am engaged in a number of curriculum development projects¹ based on theoretical and empirical research in how mathematics can be learned. The research is done in connection with a loosely organized group of mathematicians and mathematics educators known as the Research in Undergraduate Mathematics Education Community, or RUMEC.² The educational strategy which arises out of this research involves a number of innovations including: cooperative learning, students constructing mathematical concepts on the computer, de-emphasizing lectures in favor of problem solving and discussions designed to stimulate student constructions of mathematical concepts.

Implementing these innovations raises a number of assessment questions. How do we estimate what individual students have learned if most of their work is in a group? If students construct mathematical concepts on the computer, how can we tell if they have made similar constructions in their minds? If our theoretical perspective implies that a student may know something quite well but not necessarily display that knowledge in every instance, what is the meaning of answers to specific questions on a timed test?

In this article, I will describe how the curriculum development projects relate to these issues. I begin with a very brief sketch of the theoretical framework in which the research takes place and the overall pedagogical strategies it leads to. Then I will describe some ways in which research has influenced the assessment component of the curriculum development. Finally I will outline our approach to assessment.

A theoretical framework

Our theory begins with an hypothesis on the nature of mathematical knowledge and how it develops.

An individual's mathematical knowledge is her or his tendency to respond to perceived mathematical problem situations by reflecting on them in a social context and constructing or reconstructing mathematical actions, processes and objects and organizing these in schemas to use in dealing with the situations. [1].

There are a number of important issues raised by this statement and many of them relate to assessment. For example, the fact that one only has a "tendency" rather than a certainty to respond in various ways brings into question the meaning of written answers in a timed exam. Another issue is that often the student perceives a very different problem than what the test-maker intended and it is unclear how we should evaluate a thoughtful solution to a different problem. The position that learning occurs in response to situations leaves very much open the

¹See [2], [6], [4].

²For more about RUMEC, visit our web site at <http://rumeccs.gsu.edu/>.

sequence of topics which a student will learn. In fact, different students learn different pieces of the material at different times, so the timing of specific assessments becomes important. Finally, the position that learning takes place in a social context raises questions about how to assess individual knowledge.

The last part of our hypothesis relates directly to how the learning might actually take place. It is the role of our research to try to develop theoretical and operational understandings of the complex constructions we call actions, processes, objects and schemas (we use these as technical terms which are fully described in our publications) and then to relate those understandings to specific mathematical topics. For some details on how we go about this, see [1] and some of our research reports which are beginning to appear in the literature (for current information, visit the RUMEC web site mentioned above.)

Given our understandings of the mental constructions involved in learning mathematics, it is the role of pedagogy to develop strategies for getting students to make them and apply them to the problem situations. Following is a list of the major strategies used in courses that we develop. For more information see [1], [7], [3].

- Students construct mathematical concepts on the computer to foster direct mental constructions and provide an experiential base for reflection.
- Students work in cooperative groups that are not changed for the entire course.
- Lectures are de-emphasized in favor of small-group problem solving to help students reflect on their computer constructions and convert them to mental constructions of mathematical concepts.
- Students are repeatedly confronted with the entire panorama of the material of the course and have various experiences that help different students learn different portions of this material at different times. We refer to this arrangement as an *holistic spray*.

Some inputs to assessment from research

The position on assessment which follows from our theoretical framework is that assessment should ask two kinds of questions: Has the student made the mental constructions (specific actions, processes, objects and schemas) which the research calls for? and: Has the student learned the mathematics in the course? Positive answers to the first kind of question allow the assertion, supported by research, that the mathematics based on these mental constructions has been learned. This permits us to test, albeit indirectly, for knowledge that the second kind of question may not get to.

Unfortunately, it is not practical, as part of standard course procedures, to test students for mental constructions they may have made. In our research, we use interviews, teaching experiments and other methods, all of which require enormous amounts of time and energy, to get at such questions. So we must introduce another indirect component to our assessment. This

involves two stages: design and implementation of a very specific pedagogical approach, referred to as the ACE teaching cycle, designed to get students to make certain mental constructions and use them to construct mathematical knowledge; and application, to a particular group of students, of certain assertions, based on research, about the effect of this pedagogical strategy on students' making mental constructions.

The ACE teaching cycle is a course structure in which there is a weekly repetition of a cycle of (A) activities in a computer lab, (C) classroom discussion based on those activities, and (E) exercises. The computer activities are intended to directly foster the specific mental constructions which, according to our research, can lead to understanding the mathematics we are concerned with; the classroom discussions are intended to get students to reflect on these constructions and use them to develop understandings of mathematical concepts; and the exercises, which are fairly traditional, are expected to help the students reinforce and extend their developing mathematical knowledge. (For more details, see [1].)

The second stage of this component is an application of our ongoing research. Our investigations use laborious methods combining both quantitative and qualitative data to determine what mental constructions students appear to be making, and which mental constructions appear to lead to development of mathematical understanding. One outcome of these studies is to permit us to assert, not with certainty, but with some support, that if the pedagogy operated as we intended, that is, the student participated in all of the course activities, cooperated in her or his group, completed the assignments, did reasonably well in exams, etc., then the mental constructions were made.

Because this last point is somewhat different from the kinds of assessments most of us have been used to, perhaps an example will help communicate what we have in mind. Consider the chain rule. We would like students to be able to use this to compute the derivative of a "function of a function" in standard examples, but we would also like the student to understand the rule well enough so that later it can be used to understand (and perhaps even derive, from the Fundamental Theorem of Calculus) Leibniz' formula for the derivative of a function defined by an integral whose endpoints are functions.

Our research suggests that a key to understanding the chain rule might be an understanding that certain definitions of functions amount to describing them as the composition of two functions, which itself is understood as the sequential coordination of two processes. Our research also suggests that if students successfully perform certain computer tasks and participate in certain discussions, then they are likely to construct such an understanding of the chain rule and also will be reasonably competent in applying this rule in traditional examples.

In principle we could simply perform the same research on the students in our classes and get the assessment directly. But this would be vastly impractical since the research involves interviews and transcribing and analyses that could take years. Instead we ask if the students did perform the computer tasks, did participate in the discussions, and did cooperate in their groups (we determine this by keeping records of their written work, classroom participation,

and meetings with groups). We also ask (by testing) if they can use the chain rule to compute various derivatives. If the answer to these questions is yes, then, *given the research* we have reason to hope that the students not only learned to use the chain rule, but also developed an understanding that could help them understand Leibniz' formula in a subsequent course.

There is a second consequence of our theoretical position which moves us away from thinking of the course as a set of material which the students must learn so that assessment must measure how much of it they did learn. Rather we think of the students as beginning with a certain knowledge and the goal of the course is to increase that knowledge as much as possible. Thus, in making up an examination, for example, we don't think so much of questions that cover as large a portion of the material as possible, but we try to ask the hardest possible questions about the material we believe the students have learned. The expectation is that the students will do well on such tests and a part of our assessment of how well the course went in terms of what was intended (in the sense of the previous paragraphs) consists of assessing how hard the tests were and how much material they covered.

It could be argued that in this second consequence we are throwing out the requirement that, for example, everyone must learn a certain amount of material in order to get an A. We would respond that, in fact, such a requirement cannot be, and is not, implemented. It is simply impossible, given the realities in which we work, to take a course such as Calculus I, list a set of material and then determine with any degree of accuracy that a given student has learned this or that portion (i.e., numerical percentage) of it. We accept this reality, for example, when we give an exam limited to one, or even two hours and, of necessity, select only *a portion* of the material to test. We are making an assumption that students who score x on such a test understand x amount of the selected material and *also x amount of the material not tested!* We don't see this as a more compelling conclusion about how much of the material was learned than the conclusions we draw using our research.

We also accept the reality when we curve our results, basing our grades not on a given amount of material which we judge to warrant an A, but based on how well the brightest students in the class perform on the exam. Again, assumptions are being made that are not more certain than ones being made in our approach to assessment. As an aside, I would like to forestall an argument that curving grades is a practice not used very often today. I think it may be used more than we think, perhaps implicitly. For example, consider a large engineering oriented school with thousands of students each year taking calculus to satisfy engineering requirement. The grades in such a course generally fall along a certain bell-shaped distribution. Imagine, for example, what would be the reaction if the student performance were significantly lower (three-quarters of the class failed) or higher (more than half the class got an A). Are we prepared to deny that there is (perhaps implicit) curving here? Do we think that this situation represents a reasonable standard of a given amount of material for an A? If so, what would a list of that material — as determined by what is on the tests — look like?

Finally, let me mention one other input, this time from general research in cooperative learn-

ing. The results regarding this pedagogical strategy are mixed. There are reports showing large gains as well as others that do not show much advantage from it, and there do not appear to be many results in which cooperative learning was harmful. Studies that have taken a closer look report that there are conditions under which cooperative learning is more likely to be beneficial. One of the most important conditions, according to Slavin ([8]) is that students are rewarded individually for the performance of their group³. As will be seen in the next section, we make heavy use of this principle.

An approach to assessment

In our courses, students are assigned to permanent groups (of 3 or 4) very early in the course and they do most of their work in these groups, including some of the tests. We use the following assessment items. Because the first of these, computer assignments, are designed to stimulate mental constructions and often ask students to do things that are new and different for them, the grading is relatively lenient and tries to measure mental effort as much as correctness. All of the other instruments are graded in standard ways. Each of the first two exams listed below is held throughout an entire day so that students do not have time limits. They are allowed to leave and return during the day, on the honor system that nothing related to the course will be done during that day except when they are in the exam room.

Weekly computer assignments. Students have lab time to work on these in their groups, but not enough for the whole assignment and they must spend large amounts of time on their own, either individually or in collaboration with their group. The assignment is submitted as a group.

Weekly exercises. These are almost entirely apart from the computer and are fairly traditional. They are done entirely on the students' own time, and again the submission is by group.

First exam. This is a group exam. It comes about 40% through the course and the students take it as a group, turning in only one exam for the entire group. Every student in a group receives the same grade.

Second exam. This comes half way between the first exam and the end of the course. It is taken individually, but each student receives two grades: her or his score on the exam, and the average of the scores of all of the members of the student's group.

Final Exam. This exam is given in the standard way during the standard time period. It is taken individually and students receive only their individual score.

Classroom participation. Much of the class time is taken up with small group problem solving and discussion of the problems and their solutions. Both individual and group participation are recorded.

³There are some opposing views in the literature ([5]) but they are more about the general question of using rewards, such as tests, to motivate students.

For the final grade, all but the last item are given equal weight and the last is used to resolve borderline cases. Thus an individual student's grade is determined, essentially, by six scores, four of which are group scores and two are individual. This imbalance between group and individual rewards is moderated by one other consideration. If the student's individual scores differ sharply from her or his group scores, then as much as a single letter upgrade or downgrade in the direction of the individual scores will be given.

Uses of assessments

We can summarize the uses we make of the various assessment activities as follows: assigning grades, reconsidering specific parts of the course, and reconsidering our entire approach.

We described in the previous section how assignment scores, exam scores and classroom participation of individuals and groups are combined in determining the grade of an individual student. As always, such measures leave some students on the borderline between two possible grades. In these cases we apply our feeling, based on our research, that when all of the components of our course work as we intended them to, then learning takes place. Thus, if the course seemed to go well overall, in its own terms, and students appeared to buy into our approach, then we will tend to choose the higher grade. Otherwise, we will not give students much "benefit of the doubt" in determining the final grade.

The combination of course information about each student and research on students in general provides a sort of triangulation that can be used in a formative way in some cases. If the research tells us that students who experience our approach are likely to learn a particular concept, but they do not do well on this point in the actual course, then we look hard at the specific implementation as it regards that concept. If, on the other hand, students perform well on a particular concept but research suggests that their understanding leaves much to be desired, then we worry that the performance might be due to memorization or other superficial strategies. Finally, if both student performance in courses and subsequent research suggests that they "are not getting it", then we think about local revisions to our approach.

This latter occurs from time to time. For example, in the C^4L calculus reform project, we have a set of computer activities designed to help students develop an understanding of the limit concept by looking at the problem of adjusting the horizontal dimension of a graphics window so as to keep a particular curve within a specified vertical dimension. Students don't like these problems very much and don't do exceptionally well on related exam questions. Moreover, there is nothing in the research to suggest anything striking in their understanding of the limit concept. Therefore we have reconsidered and adjusted our approach to limits.

Finally, there is the possibility that, over a period of time, performance of students in courses and research could lead us to a more general feeling of malaise with respect to our overall methods. In this case, more systemic changes, including the possibility of rejecting the entire approach would be considered. So far, this has not happened.

Conclusion

There are two comments to make in evaluating our approach. One is that it is clear that this approach to assessment reflects the principles we espouse and what we think research tells us. In particular, we have addressed the questions raised at the beginning of this article. The second is that we cannot say with certainty how effective is our assessment. We do have research reports that encourage us but, in the end, teaching, like parenting, is an activity in which we can never really know how effective were our efforts. We can only try as hard as we can to determine and implement what seems to us to be the most effective approaches, and then hope for the best.

1 Bibliography

References

- [1] M. Asiala, N. Brown, D. DeVries, E. Dubinsky, D. Mathews and K. Thomas, *A Framework for Research and Development in Undergraduate Mathematics Education*, Research in Collegiate Mathematics Education II, CBMS Issues in Mathematics Education, 6, 1996, 1-32.
- [2] E. Dubinsky, A Learning Theory Approach to Calculus, in (Z. Karian, ed.) *Symbolic Computation in Undergraduate Mathematics Education*, MAA Notes, 24, pp. 48-55, 1992.
- [3] E. Dubinsky, ISETL: A Programming Language for Learning Mathematics, *Comm. in Pure and Applied Mathematics*, 48, 1995, 1-25.
- [4] W. E. Fenton and E. Dubinsky *Introduction to Discrete Mathematics with ISETL*, Springer, 1996
- [5] A. Kohn, Effects of rewards on prosocial behavior, *Cooperative Learning*, 10, 3, 1990, 23-24.
- [6] U. Leron and E. Dubinsky, *An Abstract Algebra Story*, American Mathematical Monthly, 102, 3, 1995, 227-242.
- [7] Reynolds, B., Hagelgans, N., Schwingendorf, K., Vidakovic, D., Dubinsky, E., Shahin, M., & Wimbish, G., *A Practical Guide to Cooperative Learning in Collegiate Mathematics* (MAA Notes Number 37). Washington, DC: The Mathematical Association of America, 1995.
- [8] Slavin, R. E., When does cooperative learning increase student achievement? *Psychological Bulletin*, 94, 1983, 429-445.

November 17, 2000