Cubo Matemática Educacional
Vol. 3., N°1, ENERO 2001

Using Statistical indicators to monitor the quality of undergraduate mathematics education programs

Kenneth J. Travers
University of Illinois at Urbana-Champaign

Curtis C. McKnight
University of Oklahoma

John A. Dossey
Illinois State University

Providing undergraduate mathematics instruction is a primary mission for many mathematics departments. The lower division undergraduate program (the first two years), that typically includes service courses as well as foundational courses for mathematics majors (for example, the calculus sequence), often accounts for a large proportion of that mission. Depending on the kind of post-secondary institution, the lower division program may also encompass liberal arts courses that are intended to satisfy general education requirements-precalculus courses, developmental or remedial courses, and perhaps, technical/apprenticeship courses

Most mathematics department chairs, curriculum committees, and concerned faculty recognize that accomplishing their undergraduate mathematics education mission entails more than simply offering courses. The students to be served fall into a variety of categories that can be expected to include mathematics majors, majors in mathematics-intensive fields such as engineering, statistics, prospective K-12 teachers of mathematics and those taking mathematics to satisfy general education requirements. Courses and programs must be planned to serve each type of student. Course content and sequence need to be carefully crafted to be suitable, effective, and efficient. For each course, qualified instructors (regular faculty, adjuncts, teaching assistants, etc.) must be assigned and appropriate, effective instructional approaches determined. Instructional factors to be taken into account include class size, use of discussion sections, selection of textbooks, and the role of technology such as graphing calculators and computer algebra software.

Most department chairs and concerned faculty realize that there is a great variety of ways to carry out tasks such as those outlined above. Some courses work better than others-they have content that is appropriate to their target student audiences and are organized so the instruction can be effectively delivered. Some ways for identifying, developing, and assigning qualified instructors work better than others. Some instructional strategies may be more effective than others - or at least they are as effective as possible given limitations of available resources. In short, carrying out the responsibilities of a lower division mathematics education mission is complex and is typically accomplished with varying success. It is a demonstrable fact that departments with the same basic missions, and even offering similar course sequences, may vary significantly in the extent to which they succeed in accomplishing their respective undergraduate instructional missions.

At some point, those persons responsible for their department's instructional programs are likely to be led to ask the natural question, "How is our department doing?" Mathematics departments often proceed by inertia and tradition. At times they may be called on to evaluate and justify themselves formally to groups outside the department (for example, university required self-studies, program reviews, audits, etc.). At other times, certain department members may express the need to carry out a less formal assessment of 'how the department is doing', what needs are we not meeting, what courses are no longer effective, and so on.

In reality, evaluating the quality of a department's undergraduate teaching mission, and addressing those questions that must be answered in making such an evaluation, are often done on an *ad hoc* basis. But this article is directed at those individuals who seek a more systematic approach to monitoring of the quality of the undergraduate program and are looking for ideas as to how this responsibility can be carried out successfully. We therefore consider the question: "How can the quality of a department's undergraduate program be consistently, systematically, and effectively monitored so as to identify trends, problems, successes, and needed changes?"

The effective monitoring of program quality at the institutional level has a counterpart at the national level. Entities that help shape national priorities and policies in collegiate mathematics education (for example, funding agencies such as the National Science Foundation, professional organizations such as the

Conference Board of the Mathematical Sciences, or the American Mathematics Association for Two Year Colleges, Mathematical Association of America, etc.) need answers to many of the same questions as do those individuals at the institutional level [2,3]. The major difference is that, in this case, the questions are asked about the aggregate of U.S. colleges and departments. It is the effectiveness of undergraduate mathematics instruction at the national level that is to be assessed. Methods are needed to identify and understand weaknesses, strengths, and needed changes for the aggregate, and thus in turn for the colleges that make up that aggregate. The question of how to consistently and systematically monitor programs at the national level, we argue, closely parallels the corresponding question at the individual institutional and departmental levels.

How are we doing? How can we improve ?

The questions faced by mathematics department chairs and faculty seeking information to describe and evaluate the current status of their programs require factual, data-based answers. Arguably, many aspects of carrying out a department's teaching mission, just as much of teaching a course effectively, remain a matter of art and experience. For departments and for individual courses, experienced instructors' opinions do indeed matter. However, even this accumulated wisdom can lead to more insightful conclusions when informed by appropriate data describing what actually occurs in the life of the department.

The kinds of questions that are faced likely have to do with at least some of the following: the department itself-its goals and priorities; the curriculum (programs and courses); the instructional staff; the classroom practices commonly found in the department; and the students served by the department. These aspects of undergraduate instruction become focal points for data that are needed to inform experienced opinion and to help a department to accurately identify its teaching strengths, weaknesses, and needs.

Much of these data may consist of detailed, narrative (qualitative) information about specific courses, instructional practices, student attitudes, and so on. We contend, however, that these data should also include numerical (quantitative) information about various aspects of a department's programs. Statistical measures that are used to inform evaluations in this way are called *indicators*. In this article we describe an ongoing project that is exploring the role that statistical indicators can usefully play in monitoring and evaluating the quality of undergraduate mathematics programs. [See Note 1]

The results of this project should be useful to persons who through careful, systematic evaluation desire to improve the quality of undergraduate instruction in their departments. The outcomes of this project, generalized and abstracted from the concrete details of departmental life, may also assist those at the national level who are concerned with priorities and policies to be followed in improving the quality of undergraduate mathematics education. In summary, therefore, descriptive statistical indicators that help inform departments as they evaluate themselves, may also serve to help profile and monitor the national status and needs of undergraduate mathematics education.

We begin by briefly examining the idea of educational quality indicators. We then turn to the kinds of specific questions about mathematics instruction that indicators can help us answer more effectively as we monitor and evaluate the teaching and learning of undergraduate mathematics.

It is important to note that indicators are used to help present a picture - in this case, of what takes place in the life of a department as mathematics is taught. Hence, we seek not single, isolated statistical measures, but carefully organized *sets* of indicators.

To help us identify what should make up these sets of data, we propose a framework or model within which such indicators can be developed and organized. Finally, we provide selected examples of illustrative indicators that might be used in evaluating the educational quality of the first two years of undergraduate mathematics.

The exemplars that we provide are obviously are far from a complete set for even simple purposes of monitoring program quality. They are intended to serve only as catalysts for developing more complete networks of data.

What are education indicators?

We are all familiar with the use of *economic indicators* to describe the health and direction of the nation's economy. These indicators - for example, the rate of inflation, the Dow-Jones Industrial Average, the gross domestic product, and others - reflect "performance characteristics" of the economy. Even when these data have complex relations with other aspects of the economy, they provide "benchmarks" - comparisons of the state of the economy with itself at different times. The meaning of those comparisons is often the subject of public discussion and of econometric models that are devised to explain how the benchmarks relate to the state of the economy. The significance of some of the data (for example, housing starts) is relatively uncomplicated. Sets of these indicators and associated benchmarks help inform judgments of the economy's strength and of the direction of its movement (that is, prediction of more likely economic trends by comparison with past performance of the economy).

Education indicators can serve similar purposes. Mathematics departments face questions of evaluating the status of their instructional programs (strengths and weaknesses) and identifying needed or suggested possible areas of improvement. Indicators can provide empirical data to assist mathematicians make informed judgments about the quality and direction of their instructional programs.

The education research literature has considerable information about statistical indicators and their use in providing empirical benchmarks that help in judging the quality of educational programs. The details of that literature are not appropriate for this article. However, some readers may find some of the ideas and techniques useful and wish to pursue it on their own [1]. Part of our project (on which we report here) is to sift this educational literature, translate it into reasonable English with minimal jargon, and make a start at applying it to the monitoring and improvement of lower division mathematics instruction.

That literature contains varied discussions of what qualifies as an educational indicator. Shavelson, *et al*, have stated, in a report prepared for RAND [4]:

Education indicators are single or composite statistics that reflect important aspects of the education system (as economic indicators reflect aspects of the economy). They are expected to tell a great deal about the entire system by reporting the condition of particularly significant features of it. [An education indicator] should provide insight into the "health", quality or effectiveness of the system; and it should be useful in the educational policy context.

Most of the above statement is straightforward. However, some mathematicians may be surprised to hear that they work in an "educational policy context." But, shedding the jargon, anyone who has responsibility for seeing that undergraduate mathematics is taught, and taught well, makes decisions about how to get the job done with the resources available. As "everyday" as that seems, such decisions reflect stated, or implied, educational policies and those making the decisions would qualify as working in an educational policy context as defined in the above quotation.

Education indicators in the sense discussed so far serve at least three purposes:

- Indicators can help us **interpret or understand** what is done at the different levels that affect mathematics instruction. These levels include college-wide activity, department activities, specific educational programs (e.g., the calculus sequence) and individual mathematics classrooms (e.g., instructors' classroom uses of technology or students' grasp of the basic ideas of the calculus). Indicators can be used in a way that facilitates comparing, reliably and objectively, what happens in different classrooms and courses so we can more easily determine what works well and what works less well.

- Indicators can help us **monitor trends**, that is, changes over time in what happens in mathematics classes, courses, and departments (for example, enrollment patterns of females in mathematics-based career tracks).

- Indicators can help us **determine the effects of deliberate changes** we make in instruction, courses, and departmental or college-wide policies (for example, moving to larger class sizes in calculus so that all sections can be taught by faculty rather than teaching assistants or adjuncts, adopting a particular reformed calculus approach, instituting a college-wide mathematics general education requirement, and so on).

Major goals of mathematics departments "self-evaluations" often include: (1) helping us understand what is going on in carrying out service and other instruction for which the department is responsible and (2) evaluating those gradual, unplanned changes as well as deliberate, planned changes (changes in *policy* as well as changes in practice). Indicators can be of help in all these tasks.

Focal factors and questions guiding indicator development.

As we evaluate our mathematics programs using a combination of professional experience and judgment as well as empirical (indicator) data, we focus repeatedly on those factors believed to affect undergraduate instruction. Within each focus, we ask the same types of questions. If we can identify these focuses and factors, as well as essential questions about each, we are well on our way to developing a useful set of educational indicators.

Indicators are of necessity selective. They picture only certain features of what we do in providing mathematics instruction. If they weren't selective, we would drown in a sea of empirical information. Because indicators are selective, we must be sure to select the important features of what we are doing so that the data we obtain will be relevant and helpful in our decision-making. That's why identifying focuses and questions before planning what data we will collect is such an important first step.

In our project, we identified five factors on which to focus when thinking about how successfully undergraduate instruction was being provided. These factors are: (1) the *institution* (two year college, comprehensive university, research university, etc.) and the *department*; (2) the mathematics *curriculum*; (3) the *instructors* (including faculty, teaching assistants, adjuncts, etc.); (4) the *classroom* (that is, the instruction and the assessment of what students are learning that goes on in individual classes); and (5) the *students* in the classes and activities through which undergraduate teaching is provided.

For each of these factors there are key question to be addressed as we monitor the quality of mathematics teaching and learning. We consider each in turn.

1. The Institution and Department. The success of a mathematics program is determined in part by what it is trying to accomplish. To monitor and understand this, we need answers to such questions as:

- *What are the major goals of our department (not just in teaching and not just with undergraduates)?*
- *What role does our undergraduate program play in these goals and what priority does it have among the many things we hope to accomplish (research, representing the mathematics community, graduate instruction, etc.)?*

These are certainly not the only questions we would want answered in order to understand those aspects of our college and of our mathematics department that affect undergraduate instruction. But they are two important questions that can serve as examples here. We would also want to consider other factors such as available resources, what regular methods (if any) are available to help us decide when courses or instructional approaches are needed or are no longer effective, and the commitment of our instructional staff to the continued improvement of the undergraduate program.

The Curriculum. The traditional meaning of 'curriculum' is the course of study provided. College mathematics instruction usually is packaged into courses and sequences of courses that deal with particular mathematical content and have specific goals and, especially for lower division courses, ófficial' instructional approaches to accomplish each course's goals. Mathematics departments (and, in some cases, other divisions of a college or university) are responsible for dealing with those courses, their mathematical content, goals, and approaches. Let's call that combination of courses, sequences, course goals, and expected instructional approaches the "curriculum". Given this definition, questions such as the following arise:

- *How does our curriculum relate to the goals of our department, to the requirements of our "partner disciplines" (such as engineering or science) and to the needs of our student's?*
- *What methods (if any) do we use to monitor how well our curriculum is accomplishing its goals?* Who is responsible for this monitoring (department chairs, curriculum committees; course coordinators, individual instructors, etc.)?

Certainly there are other things about an institution's mathematics curriculum that we would want to know. But these two questions are illustrative of the kinds of information one would seek as quality concerns about the curriculum are explored.

The Faculty (instructional staff). Departments and institutions have goals or missions. The instructional staff refines and makes explicit those goals through actual classroom instruction. Obviously, a department's instructional staff (its qualifications, experience in teaching, beliefs, and many other factors) greatly affects the quality of the department's programs. Two sample questions are:

- *How do our faculty's interests relate to major components of our undergraduate programs?* Do the fulltime, tenure-track faculty share an interest in and responsibility for lower division instruction? Do they willingly and routinely help provide this instruction?
- *What professional development activities for the faculty have taken place in the past three years?*

"Professional development", while a bit of jargon, captures the idea that faculty are professionals and, more specifically, professional teachers as well as mathematicians. The demands, possibilities, and approaches for effective undergraduate mathematics instruction gradually evolve and grow. The best of current teaching practices can continue to be enhanced by taking into account, for example, recent developments in knowledge about how mathematics is learned or by becoming familiar with research on instructional strategies (such as the efficacy of using graphing calculators as a problem-solving tool).

Teaching professionals do their best work when they are informed and aware of changes and new possibilities. "Professional development" is shorthand for activities that are made available to the instructional staff to help them stay informed and equipped to consider the best approaches, old and new, that might be useful in carrying out their teaching responsibilities.

The Classroom. "Classroom" here is shorthand for what actually happens in mathematics class meetings, what happens between instructors and students, as teaching is carried out and (hopefully) learning takes place. Obviously this category includes many kinds of questions for which actual data would be helpful in providing answers.

Classrooms are the "center stage" of undergraduate mathematics instruction. Many activities take place there in fulfilling teaching missions. In particular, instructional activities are carried out and student learning is assessed through tests, homework, the instructor's judgment, and so on.

Two sample subclasses of questions are considered here.

Classroom instruction. Obviously, actual mathematics instruction is the central activity of carrying out a department's teaching mission. A department's plans, uses of resources, policies, and self-evaluation are all aimed at getting qualified instructors into classes appropriate to the students served and in which the instructors use effective methods to provide activities that help students willing to take advantage of these opportunities to master the goals of that class. In considering the kinds of activities that create opportunities as a part of instruction, several questions are relevant. Here are two:

- *What do we expect students in a particular course to be able to do with the mathematical content they encounter and hopefully master?* More formally, with a bit of jargon, "What do we expect our students to know and be able to do?"
- *What kinds of technology are available to support and enhance classroom instruction?* To what extent is the available technology used? Does the department have policies about its use? Does the department help the instructional staff make more effective use of the available technology?

To be sure, these two questions are quite different from each other. This is intentional and done to show the diversity of the kinds of questions we may need to ask about classroom instruction.

Assessing student learning. In addition to teaching, instructors regularly assess, both formally and informally, how well their students are learning. Instructors do so for formal purposes of marking and assigning grades to students. They do so less formally in many cases to see if their teaching is effective, if the message is getting through, if changes are needed, if the instructional pace needs to be picked up or slowed down, and so forth. Even the most traditional, lecture-oriented instructors gauge their "audience" and often make at least small changes in response to what they see. This idea of assessing in the classroom what students are learning leads to an entirely different range of classroom-related questions. Two examples are:

- *What kinds, if any, of formal (collected and marked) assessment activities are in use that go beyond the usual kinds of tests?* "The usual" include quizzes, chapter tests, and department mid-term and final examinations. What other kinds of assessment activities are used in our department? Are projects assigned? What written work is required, apart from numerical or symbolic answers to exercises?

- *Is assessment used to promote learning rather than to simply assign grades to students and, if so, how?* Various assessment methods are available for deciding what, how much, and how well students know the mathematics that was taught. But is assessment used only for formal purposes to assign grades? Are our assessment methods also used to provide the instructor with a picture of what students do and don't understand so that the instructor can tailor her or his future activities to where the students are mathematically? Are the results of tests and homework used directly in discussion with students to help correct misunderstandings and weak spots or are they simply handed back to inform students of their current status?

Clearly these latter two sets of questions make it clear that gathering information and judging what and how well students are learning can be quite complex. Most of these complexities can be ignored or they can be addressed in an attempt to improve instruction. The specifics of whether and how these complexities are addressed may tell a department important things about its effectiveness, especially when their department's descriptive data are compared and contrasted with the benchmarks of corresponding data from other institutions known to be effective. It is this kind of comparative benchmarking that suggests that national and aggregate collecting of indicator data may help local departments as well as those who help to shape national policy for college mathematics.

The Students. Students are the target audience that must be reached in order for a department to successfully carry out its educational mission. Certainly students who do not do well in mathematics bear some, if not most, of the responsibility for their lack of success. But instructors, departments, and institutions must each take some share in that responsibility. Given this assumption, questions about how students engage in and react to mathematics instruction are another important focus in evaluating the quality of a department's instructional program. Two sample questions that lead to empirical data serving as indicators include:

- *What opportunities are provided for our undergraduate students to take part in the scholarly and social life of the department?* For many departments, teaching is service and students are a part of "classroom life" only. In other cases, departments seek to involve their majors, beginning students, and others in non-classroom aspects of the departments life, either by encouraging participation in mathematical, scholarly activities, by providing social opportunities in which students are encouraged to participate, or both. To help evaluate our program's effectiveness and how it is accomplished, we need to ask, "Are opportunities other than those in classrooms available to our students? Are the students encouraged to share in those activities? To what extent are all students rather than only selected subsets, such as mathematics majors provided such opportunities?"

- *How do our students, present and past, feel about our department and its programs?* Do they feel well served by our courses? Are they getting or did they get what they needed to master the mathematics necessary for their chosen career? Do they feel challenged? Do they feel supported? Do they consider the department and its faculty easy to approach? Do they consider the program a necessary evil or an important opportunity? How do they feel about specific courses and programs? Do our graduated majors have any suggestions for how we might have provided them with better preparation?

As before, these questions and the areas they represent are meant only as samples of the kinds of issues a department might wish to consider in evaluating the quality of what it is doing for undergraduates.No one department is likely to seek answers to all these questions but each department likely has a somewhat similar set of questions to be answered.

Questions such as those in the five focus areas above are often, if not typically, answered in the absence of reliable empirical data. This is not intended to belittle the professional judgment of mathematicians. However, even the most expert judges arrive at better conclusions when informed by accurate data. Decisions about program quality, student outcomes, and staffing too often are based on anecdotal information, tradition, or conjecture.

We as mathematicians are not exempt from at times approaching these kinds of questions at times without seriously engaging our professional experience and expertise. There are many reasons for this, not the least of which is that we have so many other pressing demands on our time. However, accurate and informed answers to these and similar questions are essential as a basis for building and maintaining high quality undergraduate instruction programs. A carefully designed and well developed indicator set and its consistent use can go far toward providing needed direction in developing, monitoring and improving our instructional programs.

The importance of indicator sets rather than single indicators

The categories and questions above suggest how wide is the range of potential indicators to aid in the careful collection of data to enable mathematics departments to make more informed decisions. The range of factors and their associated statistical indicators to be considered are bewilderingly broad. If indicators are to provide useful guidance, selections must be made and organized to provide a set that portrays an accurate, purposeful and integrated picture.

As already noted, it is important from the very beginning of their development, that indicators be viewed as occurring in sets rather than as isolated bits of data. Indicator-based portrayals of mathematics instruction are inherently composites.

Developing a set of indicators for undergraduate mathematics requires careful thought and planning. The operative word here is "set". A single indicator for a complex enterprise such as carrying out a teaching mission would likely be misleading and subject to erroneous interpretations. What is much more useful to departments is, instead, web of related indicators enabling the targeting of areas of importance.

Let us consider one example of the danger of focusing on a single indicator rather than a planned, organized set of related indicators. Figure 1 shows how many students are retained (or, dually, how many drop out) across educational levels of study from grade school to graduate school. The simple story appears to be the massive drop off in mathematics study (notice that the vertical scale is logarithmic). The jargon for how many students are retained over time is "retentivity".

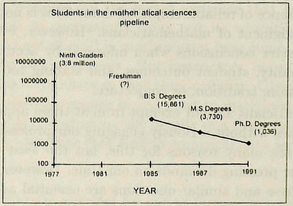

Figure 1. Sample indicator: Retentivity in the mathematical sciences "pipeline"

Figure 1 appears at first to tell a simple story. Further consideration suggests that the obvious interpretation may be misleadingly simple. What appears as a lack of retentivity may be, in fact, a result of increasing selectivity. The criteria for being allowed to continue mathematics study may become increasingly demanding as we move from pre-college to college and from undergraduate to the two levels of post-graduate study. This single indicator may portray increasing selectivity but it cannot alone reveal other aspects of such selectivity. Does increasing selectivity also involve other changes? For example, does it change the composition of the students studying mathematics at later points? Does it reflect different proportions by gender or by ethnic background? That is, are some groups more "at risk" than others as the criteria for continued mathematics study become more demanding (both in the sense of official policies and in the form of demands for more resources to allow one to continue mathematics study)? Are certain courses (for example, the calculus sequence) major "gatekeepers" for the continued study of mathematics and thus more in need of informed change than other courses, or should changes in selectivity be across the board for all courses, perhaps by admission standards at a college-wide level?

If we modify our policies to retain more students and decrease this selectivity, what other consequences will derive from this policy change ? Will simply "opening up the pipeline" be enough to increase participation by traditionally underrepresented groups? Will such an "opening up" result in weaker standards and therefore a less adequate mathematical preparation for all? To what extent can we retain larger numbers of students, especially from underrepresented groups, and still maintain our standards for expected mathematics mastery? Could greater retention be accomplished by focusing on key, "gatekeeper" courses and sequences rather than by across the board changes in selectivity?

If an institution looks only at its drop-out rate (a single indicator), without regard for the type of students or the differing impact of various courses, the picture that emerges probably hides as much or more than it reveals and changes based on these data would be misguided. If that institution assumes a simple relationship between the student dropout rate and a corresponding, undifferentiated index for selectivity of admissions, then neither the data nor what they seem to reveal about a key relationship will likely inform an effective policy.

Global indicators such as retention rates and selectivity of admissions can provide important information about the 'health' of a program, department, or institution. However informed decision-making and planning requires knowing more about other factors that are associated with enrollments and successful completion of particular courses by various types of students. A set of more specific, related indicators on drop-out rates and on the effects of selective admissions allows us to more precisely identify and understand problem areas that call for special attention. Indicators need to be organized into related, integrated sets in order to inform not mis-inform our plans for effective mathematics programs.

It is also important that indicators be interpreted in context. The setting from which indicator data are obtained may be important to understanding the significance of the indicator and for selecting appropriate benchmarks. For example, comparing budget allocations for undergraduate mathematics in a small four-year liberal arts college to related allocations in a large research university would necessarily be misleading, regardless of whether it seemed to favor higher or lower budgets. The two types of institutions have different overall missions and, in particular, different responsibilities for providing mathematics instruction.

Various mathematics teaching missions have differing budgetary implications. Were we to change the proportion of an institution's budget allocated to mathematics instruction or even the proportion of the institution's mathematics department's budget devoted to undergraduate instruction, the results would remain fundamentally incomparable because of the differing priorities for instructing undergraduates in the two types of colleges. However, for comparing institutions with similar missions, the same budgetary allocation indicator could provide useful benchmark data and have considerable interpretive power.

Single indicators are best used to raise questions or to identify potential problems or issues. They are less useful for rendering overall judgments about the how adequate a program or institution is or how well it performs. That latter purpose is best served by sets of related indicators that provide more specific, detailed information and that serve as context for each other in informing the common sense of experienced professionals. Even sets of indicators are best used as sources of information for reflection rather than decision criteria. Used in this thoughtful way, they can lead to more insightful planning and decisions.

Given these considerations we must conclude that indicators need to be carefully constructed in structured sets of related indicators that provide full, rich, and contextually sufficient pictures of programs. This is true whether the indicators are used for external or internal analysis and evaluation. Externally, demands for indicator data traditionally have focused on key instructional outcomes and insight into how instruction is organized and delivered. This concern has been accompanied by calls for evaluating the "value added" by mathematics instructional programs. Unfortunately, requests for data on how much "value" is added by a department or institution's efforts have often used a simplistic "input-output" model with little attention paid to the qualitative differences by which programs add value to their "inputs" - that is, to student demographics, availability of instructional resources, whether the instructional work force keeps current in terms of classroom uses of technology, and so on. Accurately picturing what our departments accomplish (their "outputs") through their teaching efforts, given what they have to work with (their "inputs"), must take into account not only "how much" (numbers of students entering or completing calculus, continuing in mathematics-intensive majors, etc.) but "how well" (what students can actually do mathematically after instruction that they would not do before).

Growing fiscal pressures and changes in student enrollments and what those students seek from mathematics programs are leading to increased internal evaluation of undergraduate teaching as many colleges and universities take new looks at the mathematics instruction that they deliver. Evaluation in that demanding context obviously will not have much effect if based solely on anecdote and opinion. Empirical data are essential for making convincing cases for the worth of what we do. These evaluations also seem to have moved from focusing only on "inspecting the end results" of mathematics instruction to considering more broadly how this is accomplished and how efficiently and effectively the means are used. This broadened focus makes data from well-conceived, structured indicator sets even more important.

A Model for Undergraduate Mathematics Indicators

How can we develop the kinds of organized, integrated indicator sets we have been arguing for? One effective method is to use a generic planning model that systematically identifies the factors to be considered and their relations to each other. Figure 2 presents a schematic overview of a model that provides a framework for thinking about related indicators that help to describe the goals, status, and quality of undergraduate mathematics programs. Using this framework not only suggests areas needing indicator data but also helps to identify areas missing from our consideration.

I	II	III	IV	
	Intentions	Transactions	Outcomes	
System/ Institution				
Department				
Classroom				
Student				

Figure 2: An organizational framework for a system

of indicators for undergraduate mathematics instruction

In the model, five clusters of issues related to undergraduate mathematics are identified. These major components are indicated in the left-most column of Figure 2: department, curriculum, faculty (instructional staff), classroom practices, and students. These are, in fact, the focuses we identified in the earlier discussion.

The column's topic order suggests a 'top down' view of undergraduate mathematics education, starting with a consideration of the department in its institutional setting. Mathematics education takes place in a variety of post-secondary institutions. Within these institutions, with different missions, departments have differing goals and priorities. It is the *department* that organizes programs, develops mathematics curricula, selects faculty, staffs courses, provides instructional opportunities for students, assesses student progress, conducts research, and otherwise provides much of the context within which the teaching and learning of mathematics take place. That is, mathematics departments make plans and create environments for the other levels of mathematics instruction, noted in the remaining four rows in Figure 2.

Now let us refer to the columns of Figure 2, labeled 'Initial conditions', 'Actions' and 'Outcomes'. For each of the five levels (rows of Figure 2), there are three aspects of that level that may be considered. These are specified by the three columns of Figure 2. Certain conditions occur before the actions taken. These we label Ínitial conditions'. Actions have results which we label "Outcomes". Between the stage set by the initial conditions and the outcomes in that setting are the actions themselves, a sort of transaction among participants whether those participants are the persons in seeking to change a curriculum or course, the instructors and students interacting in a classroom, or whatever players are appropriate for the actions taken in that setting. These factors we label "actions".

Figure 2's diagram has three columns that reflect these aspects that shape a mathematics program's activities - from institutional and department intentions to evaluating student outcomes. The first column focuses on the initial conditions and contexts for each of the five levels of undergraduate mathematics programs. For example, at the department level (Level I), a department's goals are conditions flowing from the teaching mission and priorities assigned it by the institution of which it is a part. Column Two focuses on the actions by which initial conditions moves toward program outcomes. An example, again at Level I, might be having a departmental committee that regularly reviews that department's goals and priorities. Column Three represents the outcomes of actions that take place in the setting defined by the initial conditions. Continuing our example, at Level I this might be statements of a department's priorities that emerge from reviewing the department's missions and goals.

The fifteen cells of Figure 2 are created by combining the five levels and the three evaluative viewpoints that we have just described. Cell (1,1), for example, specifies indicators that reflect a program's current goals or intentions, and the charge given to it by the broader institution. This cell might also include indicators on the recency, breadth, and consistency of those goals and intentions. Cell (1,2) includes departmental actions to plan what to do to carry out their assigned mission and meet those goals. This could include the departmental structure for monitoring whether goals are attained and for initiating new plans to better attain or revise goals based on past attempts to accomplish them. Cell (1,3) would encompass indicators whose focus is on whether departmental goals are attained. For example, it might include data on to what degree the program's various goals are being met and by whom.

Similar descriptions hold for each of the other four levels (the rows, Levels II through V) of an undergraduate program - that is, for curriculum, faculty (instructional staff), classroom practices, and students. We conclude this article with exemplars of indicators categorized by a few selected cells of the model in Figure 2. These indicators might well be included in a set effective initial indicators for departments to use. More importantly, they should help make the idea of an indicator more concrete and be of assistance as mathematics departments undertake the design of an indicator system to study and evaluate their own structure and functioning.

Exemplars of indicators

As discussed earlier, there is a large variety of possible statistical measures to use as indicator data. A department must select among those possibilities and develop an organized indicator set that meets its needs. The framework in Figure 2 can help identify and organize a set of indicators and is also useful in checking any resulting set completeness. Following are illustrative indicators, based on selected cells of Figure 2's framework.

Level 1: Department. Let's first consider antecedents, transactions and outcomes at the departmental level. Figure 3 presents the kind of question that might solicit information on what proportion of a department's teaching resources are devoted to lower division mathematics teaching. These data might be considered indicators of what a department does to meet its goal for undergraduate instruction. It may well be more appropriate, however, to regard these data as telling us something about the relative priority the department puts on lower division instruction.

Department level data: Initial conditions

Sample Indicators: Lower Division Teaching vs. Other Departmental Missions (Four year colleges)
-Percentage of tenure track faculty loads devoted to lower division instruction versus service, graduate, or research efforts of the department.

(Two year colleges)
-Percentage of full time faculty teaching load devoted to developmental vs. transfer courses

Figure 3. Indicator of the priority a department gives lower division mathematics instruction

This example captures one interesting property of indicators. Data used as indicators are often the basis for inferring indirectly something about a program that would be hard to get at by direct questions. For example, we could imagine directly seeking data on the relative priority a mathematics department gives to lower division instruction by presenting department chairs with a list of possible goals and missions and asking them to rank order the possibilities to reflect their department's actual priorities. The result would be one informed person's professional judgment. It would be informed but relatively subjective. It might well confound desired goals with those the department actually devotes resources to. The data here are a more objective basis for assessing the relative priority of lower division teaching. Obviously this is not an either/or situation. Both data sources could be used, as well as similar rankings by departmental faculty or by college administrators. The results could then be used for "triangulation" in helping to identify the true priorities seen in the department's actions and whether they differ from the expectations and perceptions of the persons involved.

Level II: Curriculum. Figure 4 presents a question that might be used to survey faculty with at least occasional responsibilities for lower division instruction. The question is designed to get at the relative priority that instructional personnel give to three important things they might expect from students. This information might be interesting descriptive information in itself. More importantly for indicators, these data could be used to construct a picture of the relative priorities faculty hold for students learning mathematics - that is, whether mathematics is perceived mainly to involve mastering procedures, applying mathematics, or reasoning logically.

Faculty (instructional staff) level data: Initial conditions

Sample Indicator: Expectations for student performance in mathematics

(For the target class) What is the importance of each of these abilities?

-Ability to perform routine mathematical procedures (those which are primarily algorithmic and with limited contingent behaviors).

Not at all important Somewhat important Extremely important

-Ability to understand or create an appropriate mathematical model (representation) of an everyday situation and to express questions from that situation terms of the model.

Not at all important Somewhat important Extremely important

-Ability to provide plausible and mathematically-based justification for a problem solution strategy, conjecture, etc.

Not at all important Somewhat important Extremely important

Figure 4. A question to build an indicator of faculty expectations for what students can do with the mathematics they learn

Indicators might include not only the mean priorities or emphases among the three but also the strength of agreement on these priorities among the department's faculty (obtained by examining the standard deviations of the responses). We might reasonably expect that consistent overall faculty expectations would affect what they chose to emphasize in lectures, discussions, use of textbooks, exercises assigned, and tests used to assess students. In this sense, these data would be information on the Initial Conditions, that is, on instructor beliefs that affect what those instructors actually did (Actions) in the classroom and the results of those actions (Outcomes).

Level III: Faculty. Figure 5 shows a question that is intended to provide data on the extent of various types of faculty interactions with undergraduate mathematics students other than in formal classes settings- through designing and supervising courses, advising and counseling students, participating in on-going faculty research projects or being directly involved in student activities. The data, most likely provided from departmental records, indicate what faculty members are actually doing because of their interest in undergraduate mathematics students or even from involvement not necessarily motivated by interest (that is, which might reflect a departmental policy or tradition of wide-spread faculty involvement with undergraduates). Since it is intended to describe what faculty are actually doing, these data should be considered as indicators of "Actions".

Faculty (instructional staff) level data: Actions

Sample Indicator: Faculty interest-involvement with undergraduate students

Note: These data should be reported by faculty rank: years in rank; full or part-time status.

-Proportion of the faculty involved in significant extra-class activities associated with program design and monitoring
-Proportion of the faculty involved in significant extra-class activities associated with student activities such as math club, Putnam team, actuarial exam preparation
-Proportion of the faculty involved with undergraduates in significant program or other advisement/counseling activities
-Proportion of the faculty involved in other extra-class student activities

Figure 5. Faculty extra-class involvement with undergraduate mathematics students

Level IV: Classroom. Figure 6 presents a question for students around which an individual or small group project could be built either for in-class or out-of-class use. In itself it has no value as an indicator. Suppose, however, that this item was collected when instructors in a mathematics department were asked for a sample of the typical activities they used in teaching calculus. Consider the process that might lead an item such as this to be in that sample. Instructors likely differ on what they believe is appropriate that students be able to do with mathematics and with the tasks they give them during a course to help attain the ability to do what the instructors think they should be able to do. The tasks collected in the sample would likely range from routine textbook exercises to more demanding, real data problems such as the task suggested by Figure 6.

A department could collect a random sample, or even a complete inventory, of tasks used by instructors who had recently taught calculus in the department. A simple category system could be used to sort the sampled tasks into categories from routine exercises to extended projects. The task in Figure 6 would likely fall toward but not at the latter end (that is, extended projects) of the category set. The relative proportions of tasks of different types would give a relatively objective, empirical profile of the kinds of tasks instructors are actually using to teach calculus.

In one sense, these data might be regarded as "Actions", things actually done in the classroom. If, however, the profile formed is used to indicate the results of faculty beliefs and practices about what students should be expected to do, then these data might better be considered as "Outcomes", specimens from the classroom that reveal the outcomes of instructor planning and decision-making. Clearly, how an indicator is used - what inferences are drawn from its data and the aspects of instruction its data are considered to portray - is not always unique and inherent in the data. An indicator consists not only of data collected but the use to which it is put, that is, what it is interpreted as indicating. Clarifying how collected data will be interpreted is an important part of the process of designing indicators.

Student level data: Outcomes

Sample Indicator: A freshman calculus project
The Kingdome has a hydronic heating system which includes three boilers that produce together 16,500,000 BTUs per hour. It takes 0.0345 BTUs per cubic foot to raise the temperature one degree. On Saturday, February 17th at 11:00 am the Kingdome doors will open for the annual Seattle Home Show. The second part of your task as a consultant is to determine when the maintenance crew should turn this heating system on in order to bring the temperature from a predicted forty-five degrees to seventy-one degrees (the standard temperature for such an event). The Kingdome must be up to temperature in time for the 11 am opening on the seventeenth.

Problem courtesy of Betty Hawkins, Shoreline Community College, Seattle Washington.

Figure 6. Sample student project indicating types of activities used

Level V: Students. Figure 7 presents two kinds of data a department might use to gather information on student needs and difficulties. The emphasis here, however, is on process , that is, on what departments do to find out about student needs and how well the department is meeting them. A simple model of one aspect of a mathematics department's relationship to its undergraduate students would be that the department desires to be responsive to student needs (an antecedent condition), gathers data that tell them what students feel a need for (a transaction) by means of data sources like the two stated in Figure 7. A survey of the department for what sources of data it uses to understand and respond to student needs would include several items like the two shown in Figure 7. A department representative might check all sources actually used by the department. These data on data sources indicate something of the outcome of the departments desire and actions to be responsive to student needs. Of course any outcome data can be used to make inferences about transactions and about initial goals. Placing an indicator into one column of Figure 2's framework ignores these possibilities for inference. This example shows how the framework is meant to be used flexibly and suggestively as a catalyst to stimulate thinking, and not as a set of cells by which to confine one's ideas.

The number of sources used and which are used can be used to indicate how seriously and effectively a department has acted in determining student needs. How much is enough? That is the place for benchmarking, for comparing one department's data with those from another comparable department. It is an argument for collecting indicator data at a national level and reporting useful benchmarks by which individual departments can better judge themselves.

Student level data: Outcomes

Sample Indicator: Identifying student needs and difficulties
-Extent of use of results derived from, student satisfaction surveys or reaction questionnaires
Not at all *Somewhat* *A great deal*
-Extent to which student responses on surveys (or in interviews) indicate feeling that their needs and opinions are listened to and taken into account by departmental personnel
Not at all *Somewhat* *A great deal*

Figure 7. Examples of data used by departments to identify student needs

Concluding remarks

This article was written primarily for those department personnel who are responsible for developing and monitoring lower division mathematics programs. At the campus level, this might include department chairs, committees, and individuals who provide curricular leadership in mathematics departments. A more inclusive set of concerned individuals on a campus might include those in deans' offices and others responsible for the status and quality of curricula at the institutional level. Many of the proposed indicators might also be useful for those in state boards of higher education or regional accrediting agencies.

Though secondary to this particular project and to this article, the questions explored here are also significant for those at the federal level. The questions raised provide some ideas for an indicator system at the national aggregate level to monitor the health of the Nation's system of undergraduate mathematics education. The issues raised here, we feel, strongly make a case for moving beyond simple "input-output" models of aggregate data that regard what actually happens in college classrooms and mathematics departments as irrelevant factors better left in the "black box" of an input-output model.

No effective national indicator system can ignore qualitative differences achieved by different departments as they carry out the assigned responsibilities for undergraduate mathematics instruction. Informed, process-sensitive data rooted in the real functioning of mathematics classes and departments can reveal qualitative differences in how outcomes were attained. This kind of indicator data can go far toward explaining the "why's of the outcomes". Why describe what is happening in US college mathematics classes if this is not done in a way that suggests how deficiencies identified by such an aggregate assessment might be met? Well-grounded and informed indicators could provide valuable ideas for those concerned with undergraduate mathematics education initiatives at the National Science Foundation, for private foundations, and for those involved with undergraduate mathematics programs as part of accrediting (pre-college) teacher education programs.

Note 1:

This article draws on the work of a national project appointed by the American Educational Research Association (AERA) to develop a conceptual framework for gathering data to assess the teaching and learning of mathematics during the first two years of undergraduate education. The project consisted of a Steering Committee, and National Advisory Panel and authors of invited papers. The members of these groups are listed, respectively, below. The charge to the project stated, in part:

The focus ... is to be on undergraduate mathematics education indicators. Concern is to be primarily with lower division programs for the entire population of students, not just those majoring in mathematics. Concern is also to be for the broad spectrum of public and private institutions including community colleges, liberal arts colleges, comprehensive universities and research universities.
The key findings of that project are summarized in this article. **A detailed report is available from the lead author.**

Project Personnel

Steering Committee:

- Nabeel Alsalaam, National Center for Education Statistics
- John Dossey, Illinois State University
- James Lightbourne, National Science Foundation
- Carolyn Mahoney, California State University at San Marcos
- Curtis McKnight, University of Oklahoma
- Janet Ray, Seattle Central Community College
- Alan Schoenfeld, University of California at Berkeley
- Kenneth Travers, University of Illinois at Urbana-Champaign (chair)
- Larry Suter, National Science Foundation

National Advisory Panel:

- Paul Holland, University of California at Berkeley
- Raymond Johnson, University of Maryland -College Park
- Rogers Newman, Southern University
- Mary R. Parker, Austin Community College
- Paul Sally, University of Chicago
- David Smith, Duke University
- Alan Tucker, SUNY at Stony Brook

- Ann Watkins (corresponding member), California State University, Northridge
- Paul Williams, University of Wisconsin, Madison

contributing authors:

- Eric Dey, University of Michigan
- John A. Dossey, Illinois State University
- Peter Ewell, National Center for Higher Education Management Systems
- Sylvia Hurtado, University of Michigan
- Curtis McKnight, University of Oklahoma
- Allan Schoenfeld, University of California at Berkeley
- Janet Ray, Seattle Central Community College
- Larry Suter, National Science Foundation

This project was funded by The National Science Foundation and the National Center for Education Statistics. Opinions expressed herein do not necessarily represent the views of the AERA Grants Board or of the funding agencies.

Referencias

[1] **Burstein L., Oakes J. and Guiton G.**, *Education indicators. In M. Alkin (ed.)*, Encyclopedia of educational research (6th Ed.), pp 409-418, 1992.

[2] **George, Melvin D. et al**, *Shaping the Future: New expectations for undergraduate education in science, mathematics, engineering and technology.* Arlington, VA: National Science Foundation. Report by Advisory Committe to Directorate for Education and Human Resources.

[3] Recent CBMS report.

[4] **Shavelson, Richard, Lorraine McDonnell, Jeannie Oakes, Niel Carey with Larry Picus**. *Indicator Systems for Monitoring Mathematics and Science Education.* Santa Monica, CA: RAND, page 8, 1987.