Rethinking Mathematics Assessment: Some Reflections on Solution Dynamics as a Way to Enhance Quality Indicators

Elliott Ostler
Neal Grandgenett
Carol Mitchell

This paper is intended to offer some reflections on the difficulties associated with the appropriate use of rubric assessment in mathematics at the secondary level, and to provide an overview of an assessment technique, hereafter referred to as solution dynamics, as a way to enhance popular rubric assessment techniques. Two primary aspects of solution dynamics are presented in this manuscript. The first aspect considers how the tasks assigned in mathematics classrooms might be better organized and developed to demonstrate an evolving student understanding of the subject. The second aspect illustrates how revised scoring parameters reduce the potential for scoring inconsistencies stemming from the non-descript language commonly used in rubrics.

Introduction

Professional teacher organizations have established the importance of assessment as the vanguard of instructional decision making. Specifically, in mathematics, the Principles and Standards for School Mathematics (National Council of Teachers of Mathematics [NCTM], 2000), emphasize assessment as a cornerstone to effective instruction and illustrate the need for teachers to have a solid grasp of what it means to effectively assess their students' abilities. Of course, how specific assessments are carried out in different environments will always vary according to individual needs; nevertheless, the authors still see a great need for innovation in assessment, both in interpretation and in technique.

The U.S. educational industry makes a staggering number of decisions, fiscal and otherwise, based on the "snapshot" results of standardized tests. These tests cause part of the assessment dilemma, forcing a teacher to decide whether to use the results of standardized measures or focus on assessment methods that are more contemporary and meaningful. The standardized assessments most appropriate for large-scale policy decisions are not necessarily those most suitable for instructional decision making. Ostensibly, the most appropriate small-scale assessments would be those allowing teachers to make decisions about their instruction (NCTM, 2000). Yet standardized test results continue to capture the lion's share of attention even for teachers gauging their own success. In fact, despite the research-supported utility of rubric-based assessments that allow teachers to examine quality indicators (Arter & McTighe, 2001; Goodrich, 2000; Stiggins, 2001; Wiggins, 1998), there still appears to be great resistance to transferring the scope of pedagogical decisions made from standardized tests to those more appropriate for evaluating the quality of students' mathematics work.

The purpose of this manuscript is to offer some reflections on item selection and scoring difficulties associated with appropriate use of rubric assessment in secondary mathematics and to introduce an interpretive assessment strategy, hereafter referred to as solution dynamics, as a way to enhance popular rubric assessment techniques. Two primary aspects of solution dynamics are presented in this article: first, how mathematical tasks might be better organized and developed to allow students to demonstrate evolving understanding as they progress through the subjects, and, second, how revised scoring parameters reduce the potential for scoring inconsistencies stemming from the non-descript language commonly used in rubrics.

Dr. Elliott Ostler is a professor of mathematics education at the University of Nebraska at Omaha. He has interests in assessment and technology as they relate to mathematics education and has consulted with institutions such as the College Board to develop teams for implementing curriculum for vertical articulation.

Dr. Neal Grandgenett is a professor of mathematics education at the University of Nebraska at Omaha. He has extensive expertise in technology based learning in mathematics and has authored more than 100 articles in mathematics education. He is currently assisting in building a curriculum for robotics in mathematics.

Dr. Carol Mitchell is a professor of science education at the University of Nebraska at Omaha. She has been involved public education for the past 37 years and is an alumna of the 2002 Oxford Roundtable, Oxford, England. She has numerous publications and provided leadership for several National Science Foundation Grants.
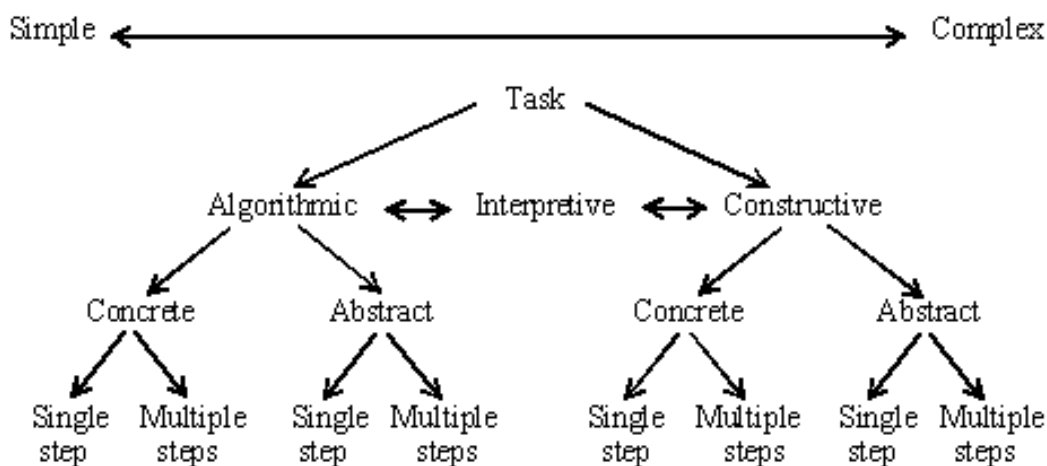
Simple ←————————————————————————→ Complex

Task
→ Algorithmic ←→ Interpretive ←→ Constructive

Algorithmic → Concrete, Abstract
Constructive → Concrete, Abstract

Concrete → Single step, Multiple steps
Abstract → Single step, Multiple steps
Concrete → Single step, Multiple steps
Abstract → Single step, Multiple steps

Figure 1.

Solution Dynamics Defined

Solution dynamics can be thought of as a way to analyze, organize, and rank student solutions based on the inherent level of sophistication represented in the tasks. This is akin to how a performance rubric might be used, but instead of measuring student performance with vague descriptors, we will make statements concerning the complexity of the tasks. Specifically, solution dynamics considers what that complexity implies for student understandings needed for completing the tasks. In some sense, the analysis of mathematical tasks for solution dynamics assessment will also determine which tasks are most effective for instructional purposes.

The solution dynamics process uses the same general techniques for ranking the complexity of problems that are used to rank the difficulty of problems in standardized tests, but the nature of the tasks require that student solutions be more open-ended. For example, if a student correctly completes a math problem of moderate difficulty on a standardized test, we may come to the conclusion that the student understands the nature of mathematics related to solving such problems. However, given the opportunity to investigate further, we may find that the student took a long time to solve the problem by using a low level trial-and-error technique, or that he or she may even have simply guessed. A rubric assessment of the same type of problem could possibly determine that the correct solution illustrates some understanding of how to complete the task, but this type of scoring would not necessarily be able to provide specific references to quality because of the nature of the way the task was presented. For example, the student's solution may receive a score of "progressing," or a 1 on a 0-3 scale, which is actually no more effective for instructional decision making than a multiple choice answer to such a question.

On the other hand, in the solution dynamics model a group of teachers would first look specifically at the task and provide an organizational structure of possible solution techniques, each of which would be ranked by the complexity of the mathematics needed. Student solutions would then be mapped to the ranked structure template (See Figure 1) for a score. At first glance, this may appear to simply be a subtle new twist on an existing rubric technique, and to some extent, it is; however, by creating a ranked structure of possible solutions for a given task, teachers have not only been forced to analyze the importance and validity of the task, but also to review a template which provides the vertically articulated concepts immediately above and below what the student's solution illustrates.

Example Solution Dynamics Task: Optimizing the Volume of a Box

Problem: Suppose a rectangular (three-dimensional) box is to be created by using a 20-inch by 20-inch square sheet of plastic (See Figure 2). Square corners will be cut from the original sheet of plastic and the rectangular tabs on each side will be folded up to create the sides of the box as illustrated below. What size corner pieces need to be removed so that the box will have the greatest possible volume? This is considered a good solution dynamics task because there is great potential for a number of possible unique solutions, starting at an arithmetic level and ending at a calculus level. The same problem can be used in each of a number of successive courses but the solutions will change (become dynamic) as the material in the courses becomes more sophisticated.
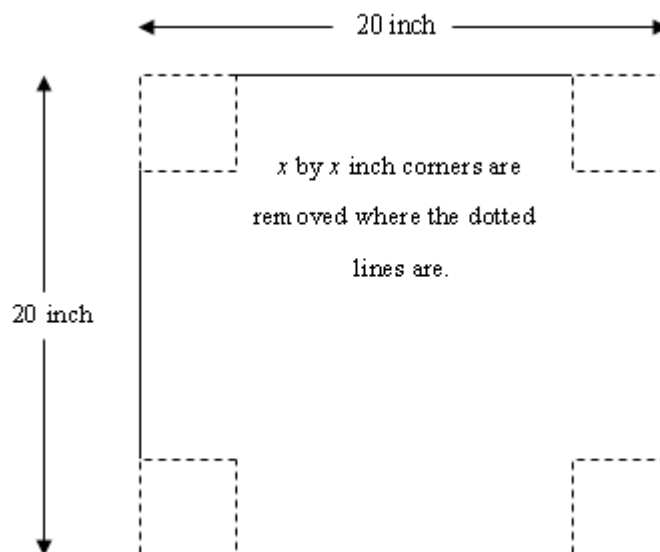
Figure 2.

Level 1 (arithmetic-based) solution. The student creates a chart (See Table 1) that records the volumes of all possible boxes with whole number increments being removed from the corners. Such a chart might look something like Table 1.

Table 1

Example of arithmetic-based solution

Corner Removed	Resulting Bases	Height	Volume
1 x 1 inch	18 x 18 inch	1 inch	342 inch^3
2 x 2 inch	16 x 16 inch	2 inch	512 inch^3
3 x 3 inch	14 x 14 inch	3 inch	588 inch^3
4 x 4 inch	12 x 12 inch	4 inch	576 inch^3
5 x 5 inch	10 x 10 inch	5 inch	500 inch^3

By the time the student has reached the fifth entry in the chart, they will probably be able to recognize that the volume is decreasing and that the optimal corner piece to remove is a 3-inch by 3-inch section. This kind of solution indicates the student recognizes that the corner piece removed has the same dimension as the height of the box and that the base of the box decreases steadily as larger and larger corner pieces are removed. They are likely to make a number of other observations as well; however, at this level they may not yet have the ability to efficiently test fractional increments, making their solution incomplete.

Level 2 (algebraic) solution. Students will use the same basic diagram to provide context, but this representation of the solution indicates that they recognize the volume of the box is a function of the corner piece removed. When examining the pattern that emerges from the chart in the first level, students may derive the following formula: $V = (x)(20 - 2x)^2$. Using this formula, students can test both whole number and fractional increments of corner piece dimensions much more efficiently than was possible with a chart. Yet this solution is still limited in that it does not allow for an efficient determination of an exact solution.

Level 3 (advanced algebra/calculus-based) solutions. Once again, students will use the same diagram to provide context for the problem. An advanced understanding of this problem will illustrate that students not only recognize the functional relationship between the volume of the box and the dimension of the corners removed, but that they understand that the volume can be graphed as a function of the dimension of that corner. They may also recognize that a maximum volume can be determined by closely examining the resulting graph or that by calculating the derivative of the function, they can determine an exact maximum point, which would represent a maximum volume of the box.

By analyzing the solution to a problem in terms of levels of sophistication, not only can we place a student on a scale, we can surmise with some accuracy what they know, and what they need to know in order to achieve the next level of complexity. The general tree diagram in Figure 1 adapted from Craig (2002) can help determine the complexity of mathematical tasks based on a continuum, which progresses from simple to complex. Galbraith and Haines (2000) conducted research that clearly indicated that mechanical

processes, here referred to as algorithmic processes, were easier than interpretive problems, which, in turn, were easier than constructive problems. Algorithmic processes consisted of mechanical solutions where students needed only to follow a sequenced set of steps to solve a problem. Interpretive problems were those problems presented in more abstract forms (i.e. word problems) from which the correct processes had to be interpreted. Constructive problems were those that required a combination of the two lower categories. Certainly the use of the model in Figure 1 does not allow for the ranking of mathematical tasks to be an exact science, but it does guide teachers to focus on the hierarchy of difficulty innate to a task.

The following example illustrates how solutions on another simple mathematical task might be ranked on a solution dynamics rubric as the mathematics used to solve the problem becomes more sophisticated. Note that the same problem is used year after year so that growth in the understanding of the processes related to this specific problem can be tracked. The differences in the complexity of the mathematics at each scoring level have been greatly exaggerated in this example in order to help differentiate between the elements in Figure 1. With an actual solution dynamics task, the differences would be more subtle and require the attention of a team of mathematics teachers to study the nuances of expected students' solutions.

The levels of the task shown in Figure 3 are certainly subject to interpretation, but illustrate how solutions become *dynamic* by focusing on the sophistication of the mathematics and the process of derivation rather than on the actual formula for the area of the circle as an answer. This particular task is one of the most basic examples of solution dynamics and one that has been used successfully by the authors in calculus courses. Allowing students to observe the evolution of complexity in a mathematical task provides context to the processes of integration.

The mathematical tasks assigned would be used to help reinforce concepts being taught at each course level. Certainly a teacher would not expect a student to use a complex mathematical technique to solve a very simple problem, but often a simple problem can provide a very powerful context for illustrating how complex mathematical ideas can be applied to various situations. In the example above we saw that a simple task can be used to demonstrate how both simple and complex mathematics can be applied to a situation. The derivation of the formula for the area of a circle is simply a convenient task that can be repeated through multiple levels of instruction to allow students to demonstrate an understanding of increasingly sophisticated thinking within a familiar context. Because the task remains the same, teachers can get a sense of what students know about evolving levels of mathematics based on how they might approach the solution.

Why Not Rubrics Alone

Rubrics are popular tools for assessment and can no doubt provide insight to student understanding in a variety of subjects and contexts if they are carefully constructed. Rubrics by themselves, however, have some inherent flaws that inhibit consistent scoring and decision making (Popham, 1997). The three most problematic flaws are as follows: rubrics are actually secondary scoring instruments but are often misunderstood to be the primary instrument; the language used in the quality descriptors, although consistent, is too vague to make meaningful decisions; and quantity indicators are often mistaken for quality descriptors. We elaborate on each of these three flaws below.

Rubrics are secondary scoring instruments

Students do not perform on a rubric. Students perform a task that is then scored by a rubric. This simple misunderstanding creates confusion about the nature of rubrics and how they should be used. It is not unusual to hear people talk about how students performed on the rubric, when in fact they mean how students scored on a preset task as interpreted by the rubric. This being the case, it should be at least as important to consider the innate value of the mathematical task as it is to consider the performance level descriptors used to rank the students' understanding of the task. Unfortunately, task considerations tend to be passed over in lieu of more careful consideration of the rubric scale.

The language used in the quality descriptors of rubrics is too vague

Tierney & Simon (2004) argue for the need to state the performance criteria and the attributes clearly. They also argue for the need to describe the qualitative degrees of performance more consistently between the performance levels of the rubric. They indicate that these modifications make the task, criteria, and attributes clearer to students and allow a broader use of the rubric. These are noble concepts, and the claim they make about clarity may be true, but the terminology they suggest is part and parcel of the problem with broad-use assessments: non-descript language. In one example, the terms they suggest using
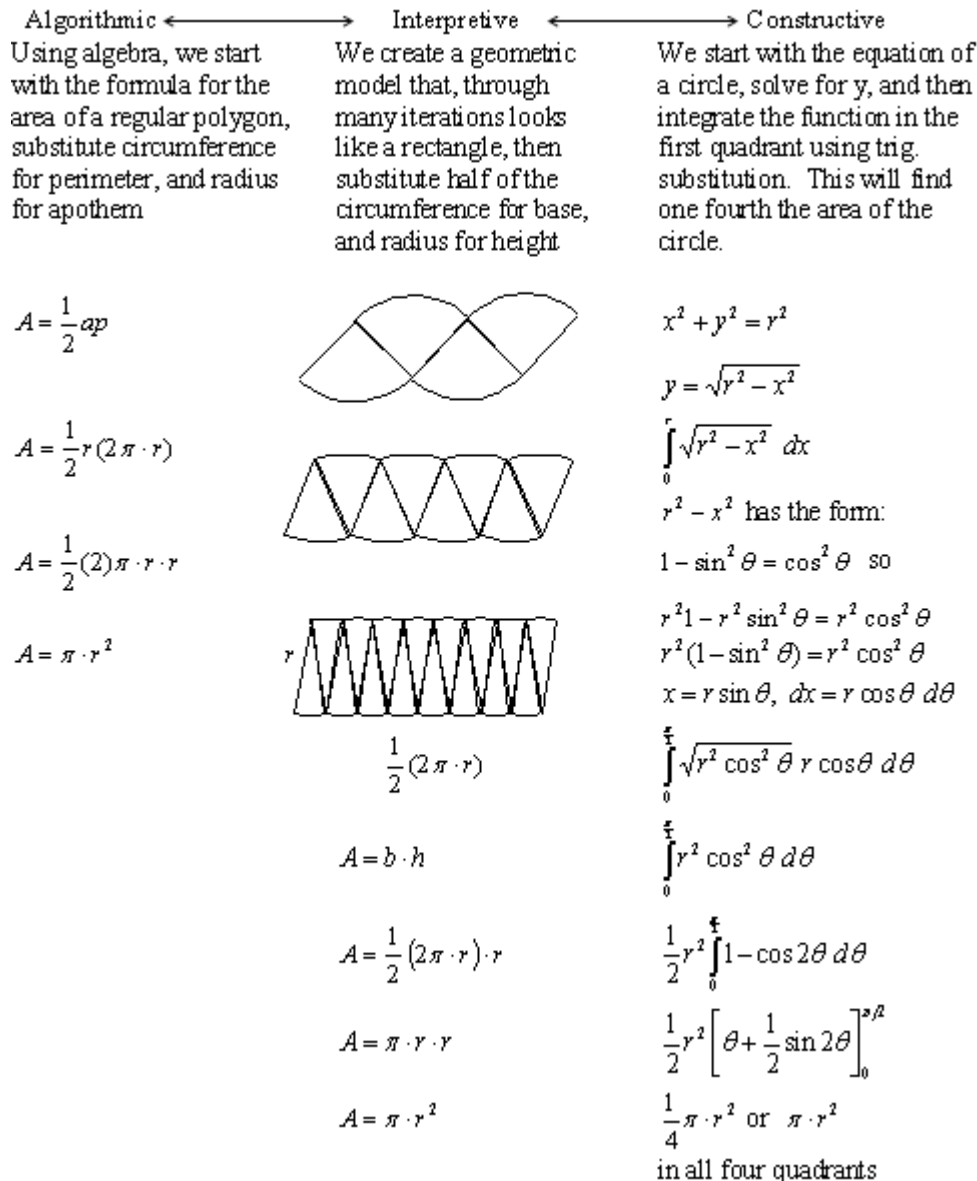
Algorithmic ⟷	Interpretive ⟷	Constructive
Using algebra, we start with the formula for the area of a regular polygon, substitute circumference for perimeter, and radius for apothem	We create a geometric model that, through many iterations looks like a rectangle, then substitute half of the circumference for base, and radius for height	We start with the equation of a circle, solve for y, and then integrate the function in the first quadrant using trig. substitution. This will find one fourth the area of the circle.

$$A = \frac{1}{2}ap$$

$$A = \frac{1}{2}r(2\pi \cdot r)$$

$$A = \frac{1}{2}(2)\pi \cdot r \cdot r$$

$$A = \pi \cdot r^2$$

$$\frac{1}{2}(2\pi \cdot r)$$

$$A = b \cdot h$$

$$A = \frac{1}{2}(2\pi \cdot r) \cdot r$$

$$A = \pi \cdot r \cdot r$$

$$A = \pi \cdot r^2$$

$$x^2 + y^2 = r^2$$

$$y = \sqrt{r^2 - x^2}$$

$$\int_0^r \sqrt{r^2 - x^2}\ dx$$

$r^2 - x^2$ has the form:

$$1 - \sin^2\theta = \cos^2\theta \quad \text{so}$$

$$r^2 1 - r^2\sin^2\theta = r^2\cos^2\theta$$

$$r^2(1 - \sin^2\theta) = r^2\cos^2\theta$$

$$x = r\sin\theta,\ dx = r\cos\theta\ d\theta$$

$$\int_0^{\frac{\pi}{2}} \sqrt{r^2\cos^2\theta}\ r\cos\theta\ d\theta$$

$$\int_0^{\frac{\pi}{2}} r^2\cos^2\theta\ d\theta$$

$$\frac{1}{2}r^2\int_0^{\frac{\pi}{2}} 1 - \cos 2\theta\ d\theta$$

$$\frac{1}{2}r^2\left[\theta + \frac{1}{2}\sin 2\theta\right]_0^{\pi/2}$$

$$\frac{1}{4}\pi \cdot r^2 \quad \text{or} \quad \pi \cdot r^2$$

in all four quadrants

Figure 3. Levels of solutions for the task of deriving the area of a circle.

to provide consistency and clarity are *few*, *some*, *most*, and *all*. These are not bad terms, but they are only indications of clarity or quality when antecedent to some very specific requirements provided in the initial task. For example suppose a timed, 100-item, single-digit multiplication test were being used as an assessment. A student answering 45 items correctly would probably fall into the "few" or "some" category of the rubric. We might surmise from that score that the student has difficulty with multiplication. However, suppose the student only answered 45 questions and was correct on all completed items. It is possible that the student simply writes slowly but knows the information very well. The terms *few*, *some*, *most*, and

all generally do nothing more than a checklist would, particularly when they are applied in the manner indicated above. If however, the assessment instrument included items that gradually became more difficult, the terms *few*, *some*, *most*, and *all* would be more appropriate because they would be antecedent to levels of difficulty within the test rather than just looking at quantity of similar items completed. This idea leads into the next point.

Quantity indicators are often mistaken for quality descriptors

As far as student performance on a given task is concerned, the demonstration of basic knowledge does not necessarily require a rubric. Once again, the nature

of the task needs to be a primary consideration. For instance, if a teacher wants students to know basic facts like multiplication tables, a rubric is probably not necessary. If a teacher were to create a rubric where the scale indicators showed increased student performance by the number of problems they correctly answered (i.e. "Beginning" = 20 problems, "Progressing" = 30 problems, "Advanced" = 40 problems, etc.) the categories would not be indications of conceptual quality nor are the descriptors assigned to the scales necessarily set by any externally valid criteria. It is therefore unnecessary to provide a rubric scale that counts or quantifies the number of correct answers. For a task such as this, a checklist would be more appropriate. Quality indicators are more appropriate to tasks that require some higher-level thinking and rubric levels that clearly indicate the quality of thinking, or the lack thereof.

Conclusion

Ultimately, there are two primary factors that make a solution dynamics approach a potentially effective way to clarify and increase the accuracy of rubric-based assessment. First, a solution dynamics model considers the evolution of a mathematical solution over time. Second, this approach specifically considers the quality of the student performance and the difficulty of the task within the same instrument. Both of these factors, though somewhat obvious, emphasize ideas that are generally absent in the explanation of rubric assessment. Problems such as the derivation of the area formula for a circle, as illustrated earlier, have been used with great success in a solution dynamics format by the authors to show not only the evolution of students' simple mathematical models to complex ones, but also to illustrate natural connections and applications between scientific and mathematical content. This has been particularly true in calculus courses where students tend to lack the conceptual understanding behind processes like integration.

Though it is probably not realistic to expect large gains in mathematical understanding to come in a single academic year for every student, the selection of the right kinds of dynamic mathematical problems can better illustrate the dynamic nature of the mathematics the students are learning and therefore help facilitate the conceptual evolution of mathematical knowledge that represents a transition from algorithmic to abstract thinking. It is important and appropriate to engage in assessment techniques that measure students' progress over a successive period of years. Attention to evolving representations of student solutions allow for this to

happen. A focused effort on vertical articulation, and in particular, efforts to build dynamic solution exercises (specific mathematical tasks that lend themselves well to solution dynamics assessment) will provide a more comprehensive view about students' understanding of mathematics and its various components, concepts, and skills.

Romberg (2000) argues that, with appropriate guidance from teachers, students can build a coherent understanding of mathematics and that their understanding about the symbolic processes of mathematics can evolve into increasingly abstract and scientific reasoning. This, of course, happens through opportunities to participate in appropriate kinds of mathematical tasks. As mentioned previously, a coherent understanding of anything does not happen with most students over the course of a single academic year. The evolution in a student's thinking that allows them to demonstrate a transition from algorithmic to abstract semiotics presumably happens over a period of years. It follows then that developing the kinds of *appropriate mathematical assessments*, the dynamic kinds that allow for this transition to be measured over time, can most appropriately be done by a team of mathematics educators. Each considers the nuances of what the others do, and then documents their part in the process through thoughtful solution dynamics assessments.

References

Arter, J., & McTighe, J. (2001). *Scoring rubrics in the classroom: Using performance criteria for assessing a improving student performance.* Thousand Oaks, CA: Corwin Press/Sage Publications.

Craig, T. (2002). Factors affecting students' perceptions of difficulty in calculus word problems. 2nd International Conference on the Teaching of Mathematics. Crete, Greece, July 1-6, 2002. Retrieved November 28, 2008 from http://www.math.uoc.gr/~ictm2/Proceedings/pap411.pdf.

Galbraith, P., & Haines, C. (mis(understandings) of beginning undergraduates. *International Journal of Mathematical Education in Science and Technology. 31*(5). 651–678.

Goodrich, A. (2000). Using rubrics to promote thinking and learning. *Educational Leadership, 57*, 13–18.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school Mathematics.* Reston, VA: Author.

Popham, W. J. (1997). What's wrong—and what's right—with rubrics. *Educational Leadership*, *55*, 72–75.

Romberg, T. (2000). Changing the teaching and learning of mathematics. *Australian Mathematics Teacher*, *56*(4), 6–9.

Stiggins, R.J. (2001). Student-involved classroom assessment (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.

Tierney, R. & Simon, M. (2004). What's still wrong with rubrics: focusing on the consistency of performance criteria across scale levels. *Practical Assessment Research & Evaluation*, 9(2). Retrieved November 28, 2008 from http://PAREonline.net/getvn.asp?v=9&n=2 .