NCRMSE Research Review
from the National Center for Research in Mathematical Sciences Education

Volume 3, Number 2
Spring 1994

In This Issue:
- Integrating Statistics into the School Curriculum
- Review of NCRMSE Research

Statistics and School Mathema tics

Students encounter numerous statistical claims in their daily lives. Data are collected, summarized, analyzed, and transformed in most of this country's media, work places, and homes. The collecting, representing, and processing of data are assuming major importance in most nations. While statistics was once taught primarily to college students pursuing professional or academic careers, it is now becoming a part of the school mathematics curriculum.

Mathematics reform documents developed during the last decade stress teaching appropriate statistical content to students in elementary and secondary mathematics classes. This content should, according to the documents, develop students' ability to use statistics to understand their worlds, to create and interpret summaries of data and displays of information, and to e critical of claims and arguments that are based on data. The *Curriculum and Evaluation Standards for School Mathematics* developed by the National Council of Teachers of Mathematics (1980) and projects, such as the Quantitative Literacy (American Statistical Association, 1987) and Used Numbers and Reasoning Under Uncertainty (Technical Education Research Center, Lesley College, and the Consortium for Mathematics and Its Applications, 1989-1990), suggest ways data analysis could be implemented in the nation's schools.

Few activities that involve statistics are now carried out in K-12 mathematics classrooms. While there is little information either from research or from practical experience to aid teachers who would introduce statistics at elementary- and secondary-grade levels, there is information on teaching and learning statistics at the college or university level. This research-based information suggests that a majority of higher education students do not understand elementary statistical concepts (Garfield & Ahlgren, 1988), even after completing several courses (Poston, 1981). In its current form, according to a consensus among educators and researchers (American Statistical Association, 1991; Mosteller, 1988; NCTM, 1989; and Shaughnessy, 1992), statistics education is inadequate.

Learning/Teaching of Statistics

The Learning/Teaching of Statistics Working Group of the National Center for Research in Mathematics Sciences Education (NCRMSE) is studying the ways in which statistical content can best be integrated into the school mathematics curriculum. While NCRMSE

Director Thomas Romberg initiated the Working Group, Susan Lajoie of McGill University now chairs the group. Sharon Derry and Richard Lehrer, University of Wisconsin-Madison, are the group's principal investigators. The group also includes 5 staff members and 17 affiliated members.

Five Working Groups were formed by NCRMSE early in 1991. They focused on the topics: whole numbers, quantities, algebra, implementation of reform, and models of authentic assessment. A sixth Working Group that focused on geometry was formed during late 1991. The Learning/Teaching of Statistics is the seventh and final NCRMSE Working Group. It began its activities early in 1993. The operation and research of each of the other Working Groups have been described in previous issues of NCRMSE Research Review.

While each of the members had completed independent research related to the learning/teaching of statistics before the creation of the Working Group, they began a collaborative program of research in June 1993. At their first meeting, these members summarized their current research and knowledge on learning/teaching statistics. They then identified the kinds of research they thought would be necessary to extend knowledge on the appropriate content, pedagogy, learning, and assessment of statistics for K-12 students.

Several studies by members of the Working Group that are designed to extend knowledge about statistics for K-12 students are now underway. The studies address a) the development of an international database of research literature that relates to the teaching and learning of statistics; b) the identification of the cognitive components of probability and statistics that are related to the K-12 curriculum; C) the development of a curriculum for statistics with Grade 7-8 teachers and implementing of the curriculum in Grad 7-8 classes; d) the development of a technology-based authentic statistics project for Grade 8 mathematics students that includes the testing of new assessment models that use computer and videotape technologies.

Research Literature Database

Chair Lajoie and Affiliated Member Joan Garfield collaborated on the development of a computerized database of research, STAT-FILE, related to the teaching/learning of statistics. They combined their previous resources and eliminated entries that were duplicates. They then used the database to develop and annotated bibliography of the research literature on teaching/learning statistics. At the end of 1993, their bibliography included 110 entries that were relevant to K-12 research. The database can be searched by title, author, key words, and publication or presentation information. Copies of the database were sent to the members of the Working Group.

Cognitive Models Projects

Richard Lehrer and Jeffrey K. Horvath are examining cognitive models of less- and more-skilled Grade 4-5 students' problem-solving performances in statistics. They take the view that curriculum-based performances should be examined when information about student knowledge and understanding is sought, and students should be assessed in ways that provide opportunities for them to learn more about the curriculum or provide opportunities for them to receive assistance with their performance. Their work began with the spinner sum task, a probabilistic assessment task developed by the California Department of Education (Lehrer & Horvath, 1993). They are now extending their work on cognition with a study of students' developing understanding of the nature of chance.

Kevin Collis has begun a study of cognitive functioning in probability and statistics as it relates to the school curriculum. It addresses specifically the understanding of chance and data concepts in the wider social context. As a starting point, he is developing a questionnaire that uses the media—excerpts from newspapers, for example—to learn about students' understanding of the content. The questionnaire will include paper-and-pencil items, supplemented with graphical information when relevant, that are constructed to elicit statistical problem-solving skills. The four-part format (Collis & Romberg, 1992) and think-aloud techniques will be used in the construction of the items.

Statistics Curriculum Project

Sharon Derry and Helen Osana have developed instructional simulations that will stimulate students to acquire and use statistical concepts in the context of reasoning about realistic of non-trivial problem-solving situations. They began with a plan to design situations that would improve middle-school students' ability and propensity to reason statistically about the problems they encounter in their daily lives. During the summer of 1993, they worked with a small group of classroom teachers to develop and refine the design for instruction simulation activity that would be used by 7th- and 8th-grade students. The 3-4 week long instructional game is called Vitamin Wars.

During the academic year 1993-1994, Derry and Osana and four classroom teachers began to implement the curriculum in 7th- and 8th-grade classrooms. As an introduction to the curriculum, a film on concepts about statistics and medical research was shown to the students. IT provided a basis for discussion and analysis. Students then listened to presentations by local experts on statistics, argumentation, and scientific research and discussed them in small groups. During the last days of the activity, students were asked to assume the roles of researchers, lawyers, legislators, consumers, or business owners. Acting in these roles, students worked in collaborative groups conducting research and preparing and presenting testimony at a mock legislative hearing on government regulation of the vitamin industry.

The researchers collected field notes, video recordings, and students and teacher ratings of class activities. They also administered a reasoning task. The data will be analyzed to identify the specific events, features of materials, procedures, aspects of the classrooms, and aspects of mentorship that were useful in shaping the research activity or affecting the outcomes. Their classroom observations suggest that role-playing motivates most students, even some who are identified by teachers as having behavioral or motivational difficulties. Despite the use of the same instructional game, the use of identical procedures, and common inservice programs for teachers, each classroom evolved as a unique social and learning environment during the activity. Some of the roles students played also appeared to be more effective than others in promoting the development of reasoning, regardless of the classroom context.

Investigators Derry and Osana are now investigating whether the simulations were more effective than the traditionally used instructional approaches in terms of the students' development of concepts, use of statistical concepts, quality of individual and social reasoning, and quality of individual and group products such as presentations, written interpretations, or others that were based on reasoning.

Authentic Statistics Project

Working Group Chair Lajoie is directing the Authentic Statistics Project. Designed to test new forms of statistics instruction and assessment, the project uses computer and videotape technologies. Its assessment and instructional models are based on situated learning theory and the NCRM Standards. Of special note is its attention to the problem solving, communication, reasoning, and connections emphasized by the NCTM Standards in both their instructional and assessment aspects. The project also designed instructional activities in statistics that would provide students with opportunities to reflect, organize, model, represent, and argue within and across other mathematical domains.

The Authentic Statistics Project developed an innovative approach to assessment, a computerized library of exemplars—text and video examples of student work—that convey to students both models of and criteria for average and above-average performance and can be used by students as benchmarks for their own statistics performances. When assessment criteria are made explicit to students, the researchers theorized, it can precipitate additional learning. A library-of-exemplars study was then undertaken to examine the effectiveness of the library as a tool for clarifying assessment criteria and for promoting statistical understanding. The Project also developed a second study to investigate whether tasks that require students to collect their own data are more effective for promoting understanding than tasks that require students to use preexisting data.

One portion of this project's study on the effects of text and video examples is reported in the next article in this newsletter. Other portions of the project's research were reported in papers presented at the April 1993 American Educational Research Association Conference in Atlanta: *New Ways to Measure Skills of Problem Solving, Reasoning, Communication, and Connectedness,* by Susanne Lajoie, John Lawless, Nancy Lavigne, and Steve Munsie (1993); and *The Use of Hypercard for Facilitating Assessment: A Library of Exemplars for Reifying Statistical Concepts*, by Susanne Lajoie, Nancy Lavigne, and John Lawless (1993).

In addition to the principal investigators Sharon Derry and Richard Lehrer, the Working Group includes Affiliated Members George Bright, Gail Burrill, Kevin Collis, Susan

Friel, Ido Gal, Joan Garfield, Christopher Hancock, Victoria Jacobs, Clifford Konold, James Landwehr, Joel Levin, Kathleen Metz, Thomas Romberg, Andee Rubin, Richard Scheaffer, Ron Serlin, Leona Shauble, and Michael Shaughnessy. It also includes Staff Members Jeff Horvath, Nancy Lavigne, Steve Munsie, Helen Osana, and Tara Wilkie. Two staff members are located at the University of Wisconsin-Madison and three are located at McGill University in Quebec.

Conclusion

Future research on how statistical content can best be integrated into the school mathematics curriculum should include an integrated framework for the instructional and assessment process. Such a framework would involve looking at the statistical content, and how instruction builds on the assessment of the learner in the context of an instructional situation. The Working Group members collaboratively will seek answers to the research questions they have identified.

References

American Statistical Association (1989). *Quantitative Literacy Series*. Palo Alto, CA: Dale Seymour Publications.

American Statistical Association (1991). *Guidelines for the teaching of statistics in K-12 mathematics curriculum*. Landover, MD: Corporate Press.

Collis, K., & Romberg, T.A. (1992). *Mathematics Problem Solving Profiles*. P.O. Box 210, Hawthorn, Victoria 3122, Australia: Australian Council for Education Research.

Garfield, J., & Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: Implications for research. *Journal for Research in Mathematics Education*, 19(1), 44-63.

Lajoie, S., Lavigne, N., & Lawless, J. (1993). *The use of Hypercard for facilitating assessment: A Library of exemplars for reifying statistical concepts*. Paper presented at the annual conference of the American Educational Research Association in Atlanta, GA.

Lajoie, S., Lawless, J., Lavigne, N., & Munsie, S. (1993). *New ways to measure skills of problem solving, reasoning, communication, and connectedness*. Paper presented at the annual conference of the American Educational Research Association in Atlanta, GA.

Lehrer, R., & Horvath, J.K. (1993). *Cognitive models of authentic assessment*. Paper presented at the annual conference of the American Educational Research Association, Atlanta, GA.

Mosteller, F. (1988). Broadening the scope of statistics and statistical education. *The American Statistician*, 42(2), 93-99.

National Council of Teachers of Mathematics. (1989). *Curriculum and Evaluation standards for school mathematics*. Reston, VA: Author.

Poston, H.O. (1981). Review of statistical teaching materials for 11-16 year-olds. *The American Statistician*, 35(4), 258-259.

Shaughnessy, J.M. (1992). Research in probability and statistics: Reflections and directions. In D. Grouws (Ed.), *Handbook for research in mathematics teaching and learning* (pp. 465-494). New York: Macmillan Publishing.

Integrating Statistics into the School Curriculum

Statistics is assuming a new role in the mathematics curriculum for K-12 students in the 1990s. Because statistics previously has not been taught to students in elementary or high school grades, little is known about students' experiences with statistics acquired outside of their classrooms and about instruction that can effectively enhance students' understanding of statistics in their classrooms. This section presents excerpts from papers on two studies that were carried out to develop a knowledge base that would enhance statistics instruction in middle school classrooms. The first paper, *Statistics in Middle School: An Exploration of Students' Informal Knowledge*, was completed by Victoria R. Jacobs and Susanne Lajoie. The second paper, *How Do Group Composition and Gender Influence the Learning of Statistics?*, was completed by Susanne Lajoie and Nancy C. Lavigne.

An Exploration of Students' Informal Knowledge

Much of the recent research in mathematics education underscores the importance of understanding the extent of children's informal knowledge before preparing formal instructional programs. To some, students' prior knowledge is an essential starting point from which to build additional instruction. It can also provide teachers with a framework for thinking about their students' development in this content domain. An enrichment program was developed for middle school students who were identified by teachers as interested in challenges beyond their regular mathematics curriculum. The instructional activities that comprised the program provided the context from which students' statistical understanding could be studied. Both the content and the pedagogy of the enrichment program were designed to provide students with opportunities for extended thought and discussion about statistical problems. They were also designed to provide teachers with activities that elicit or encourage discussion about statistical content.

Content

The content of the enrichment program was designed to guide students in explorations of the concepts that make up inferential statistics. It was based on the assumption that chance and sampling—the use of results obtained from a sample in reaching conclusions about a larger population—are the core concepts of inferential statistics. Specific lessons focused on chance and its role in everyday decisions and on sampling and the logic underlying the factors that affect a researcher's ability to draw accurate conclusions about the population from the sample results. The inferential statistical content was selected because it is consistent with that recommended for the middle school age group in the curriculum *Standards* published by the National Council of Teachers of Mathematics and the guidelines prepared by The American Statistical Association. It was also selected because, while there is little research on students' understanding of any statistical content, many current projects emphasize descriptive rather than inferential statistics.

Pedagogy

The NCTM *Curriculum and Evaluation Standards* for School Mathematics (1989) recommend a constructivist pedagogy that is based on a new vision of mathematics learning. Learning mathematics is defined as doing mathematics. This view promotes an activity-based curriculum where students solve problems and emphasis is placed on their informal knowledge and their ability to communicate about the processes they use to reach solutions. Classes emphasize problem solving rather than computational skills and explanations of how solutions are reached rather than single answers. Assessment is incorporated into instruction and modified to reflect new beliefs about knowledge and learning.

Method

A researcher/teacher met with ten high-ability middle school students in a Wisconsin school for 13 weeks during the students' regular mathematics class. The enrichment lessons were considered part of the students' regular curriculum and the teachers incorporated the students' performance in the program into their semester grades in mathematics.

The activities in the lessons were designed to promote statistical discussion. Measures of performance during the 13-week period included indices of class participation, weekly homework assignments, and a comprehensive examination administered as a pretest and posttest. All class dialogues were audiotapes, and protocol analysis provided the core of the analyses that were completed.

Results

At the conclusion of the 13-week enrichment program, students had developed a workable definition of chance, could recognize it in their daily lives, and describe how it affects the decisions they make. While most students were able to compute an exact probability, all were able to use the language of probability—e.g., more likely, less likely, impossible, etc.—to describe chance events. They had little trouble distinguishing more likely outcomes from less likely outcomes, but difficulty understanding chance events with outcomes that were equally likely. With regard to sampling, students easily identified its use and importance in their daily lives. They also were able to apply sampling logic to realistic situations. While all of the students participated in statistical discussions, some had difficulty in two areas. Students should be able to use the results obtained from the application of inferential statistics to inform their decisions about everyday situations; some of the students, however, had difficulty incorporating the results into their discussions. Students were also inconsistent in their ability to explain representative sampling procedures. This inconsistency seemed to be linked to the context of research questions. Opinion questions were treated differently from other questions by students who seemed to think that there should be an equal chance for each possible option—e.g., 50-50 or 33-33-33, and so forth.

Students' performances on a 50-50 point tests suggested that their general level of statistical reasoning had increased as a result of their participation in the project. The test rewarded correct explanations as well as correct answers. In addition to their cognitive gains, students indicated both through verbal anecdotes and through individually written course evaluations that they enjoyed the course.

Useful Instructional Materials

Students were very articulate during classroom activities designed to stimulate discussion about statistical content. They responded verbally to open-ended questions, shared their strategies, and became comfortable with questions that could have multiple answers. They struggled, however, with open-ended written assignments. The quality of reasoning in their homework and essays on tests did not match the level they demonstrated during their oral interactions. This finding suggested to the researchers that students' conceptual understanding may be underestimated if evaluations are based only on written work.

The study found that two activities were particularly useful in promoting discussions that enabled students to learn from one another. A student-generated test and a students-directed class can provide educators with unique information about students' understanding. There is little research to guide the implementation of these activities.

The study also found that students had difficulty identifying why the content of statistics was classified as mathematics. When they reviewed the program, they viewed the division, multiplication, and fraction procedures they completed with numbers as mathematics. The students' views reflect society's view of mathematics as a nonconceptual field made up of numbers and their manipulation—a view that is destructive because it encourages the use of memorized algorithms without understanding.

Conclusions

This work provides some initial guidance for mathematics educators who must begin to implement statistics in the Grade K-12 curriculum. Its suggestions for instructional and assessment activities can provide an initial glimpse of students' informal conceptions of statistics. Although generalizing from the results is limited by the size and ability level of the sample, the methodology used in the study provides information that can be used by others interested in examining students' thinking about statistics.

The Influence of Group Composition and Gender

Theories of learning and instruction are increasingly considering learning that occurs in specific situations or contexts. The situations or contexts often involve small groups of individuals working together on a common task. Constructivist and situated learning perspectives emphasize the importance of social interaction that promotes thinking and the development of problem-solving skills in classrooms. These perspectives are based on the assumption that a cooperative or shared learning environment permits students to learn from others and can reduce students' anxiety about learning when they feel willing to share information. Group work encourages students to share their knowledge with peers and exposes them to multiple points of view. Small groups can enable students to develop mathematical power by developing higher-order thinking skills such as problem solving, reasoning, and communication, and behavioral attributes such as persistence.

Despite the success of small group learning situations, little attention has been paid to the nature of collaboration and the dynamics of small-group interactions that affect learning. Research shows that peer collaboration can cause students to shift perspectives but that joint decision-making is necessary for effective learning. Group interaction may result in some students relying on others, accepting little personal responsibility, and doing little independent thinking. Collaboration is a complex phenomenon and this complexity must be considered when collaborative classroom activities are used.

This study was designed to test the ways in which group composition influences learning. In particular, it looked at group composition in terms of gender to determine how gender influenced the learning of statistics by 8^{th}-grade mathematics students. The goal of the research was to examine how best to instruct students and assess their learning in this new area of school mathematics.

According to the NCTM *Standards*, 8^{th}-graders should be provided with opportunities to do statistics—that is, to systematically collect, organize, and describe data; to construct, read, and interpret tables, charts, and graphs; to make inferences and arguments based on data and evaluate arguments based on data analyses; and to develop an appreciation for statistical methods as a powerful means for decision making. The study uses these guidelines by placing statistics instruction in an experimental context carried out in the classroom using computers that graph and analyze student data. Groups of students construct research questions, collect and analyze data, and display their findings and interpretations for the class. Both individual and group data are used to describe the transitions that occur in student learning.

To ensure that students were assessed fairly on their group projects, two conditions were developed: an exemplar condition—*video and text*—where a computer hypercard stack was designed to describe assessment criteria, and a *text-only* condition, also computer-administered, that provided a list of assessment criteria with textual descriptions of what these criteria meant and how they were weighted. The criteria included: quality of the research question—how clear and specific the research question is; data collection—how students go about gathering information that pertains to their question; data presentation—how data are summarized and presented and the types of tables, charts, and/or graphs constructed to represent the data; data analysis and interpretation—what statistics are selected to analyze a data set, and how students demonstrate their understanding and interpretation of the data analysis; presentation style—how the group explains the goals and findings of their project to the class; and creativity—how unique the project is. The video and text condition added a video component that allowed students to select one of the criteria and obtain a textual and digitized video demonstration of average and above-average performance by previous students who had designed statistics projects. The *text only* and *video and text* conditions were prepared to make the assessment criteria apparent to students.

Cognitive learning theories support small-group activities that have been found to facilitate learning for each individual by providing multiple perspectives as well as by negotiations during problem-solving activities (Blumenfeld, Soloway, Marx, Krajcik, Guzdial, & Palinscar, 1991; Cognition and Technology Group at Vanderbilt, 1990; Williams, 1992). There are few guidelines for assigning students to groups that are based on promoting learning for individual group members or on how gender affects group activities. An earlier pilot project found that group work did not always result in a positive learning experience for all group members. In the pilot, groups made up of mixed-ability levels completed a statistics project. Informal observations of group activity suggested that when females were in male-dominated groups they did not participate.

Other research that involved computer-based tasks (Underwood & McCaffrey, 1990) found that single-gender pairs improved participation and performance when compared to individual activity, but mixed pairs did not. In this research, single-gender pairs appeared to share task components and discuss possible solutions, while mixed-gender pairs tended to separate task components and then complete them separately. Other research (Webb, 1984) indicated that mixed-gender groups can be detrimental to females' mathematics achievement. Based on these findings, the current study was designed for single-gender groups to see whether gender differences would occur on measures of statistics problem solving, reasoning, or communication. All groups, regardless of gender, mixed students with varying ability levels. The predictions were that females would perform as well as males in single-gender groups on measures of statistical knowledge and that groups exposed to the *text and video* condition would outperform the groups in the *text only* condition.

Subjects for this study were from an 8^{th}-grade mathematics classroom. Twenty-one students, 9 females and 12 males, were divided into 8 groups consisting of either two or three students. Students were grouped with same-gender peers, resulting in four groups of females and four groups of males. Teachers' rankings—e.g., high, medium, low—of each students' performance as measured by classroom assignments and examinations—were used to form groups of students with mixed mathematics achivement. When groups were made up of two students, a high- and low-ability student would be placed together. When groups were made up of three students, a high-, medium-, and low-ability student would be placed together. Each of the groups worked at one Apple Macintosh computer work stations and were supervised by a researcher for a period of two weeks.

The eight groups were assigned randomly to the *text only* condition—in which textual descriptions of the criteria for assessing group projects were presented on the computer—or the *video and text* condition—in which the textual descriptions were supplemented with digitized video clips representing two levels of performance. The randomization process resulted in three groups of females and one of males being assigned to the *video and text* condition and three groups of males and one of female being assigned to the *text only* condition.

Instructional activities and assessment tasks were designed to promote learning and communication in an authentic learning environment. A pretest/posttest design was used to assess changes in student performance. Journals were used as group measures of statistical communication, problem solving, and reasoning. Structured journals contained specific prompts designed to encourage groups to define and explain concepts, to reason about data and graphs used in the tutorial, and to identify areas of difficulty. The prompts were designed as a means to foster learning. Journals served as ongoing measures of group performance and were analyzed in terms of the quality of statistical communication, problem-solving, and reasoning. All group interaction and presentations were audiotaped and videotaped.

Results

The prediction that females would perform as well as males in single-gender groupings on measures of statistical knowledge was not only confirmed but the results exceeded our expectations. The prediction that groups in the video and text condition would outperform those in the text only condition was not confirmed. The statistical knowledge of all of the groups increased from pre- to posttest, but groups of females working together benefited more from the instruction than did groups of males. Analysis of group journals suggests they do provide useful information: Female groups responded to prompts more often than males groups. Based on their responses, it appears that females were more interested in planning and in understanding concepts through definitions, while males seemed somewhat more inclined to deal with questions that required interpreting the information that was presented in the data and graphs.

Conclusions

This study suggests that gender plays an important role in group problem solving using computer-based learning environments. While gender differences did not exist on a pretest of statistical knowledge, Grade 8 females outperformed males after the instructions. It appears that providing females with opportunities to work with other females on group projects that require computers has a more positive impact than single-gender cooperative learning situations for males. This finding is similar to that of Johnson and Johnson (1985), who found that a combination of cooperative learning and computer-assisted instruction had a positive impact on female students' attitudes toward computers.

There were also gender differences in the way students documented their group projects. Structured journals were given to each group and students were asked to document the statistical concepts they were learning, as well as their project ideas. Females tended to document information about statistical concepts and their plans for how to conduct their group projects, whereas the entries for the male groups were sparse in such categories, but robust where they were asked how they would apply statistical concepts in certain situations.

Both of the conditions used in the study, *video and text* and *text only*, produced significant changes in the statistics performances of students. The finding was not expected because it appeared that the *video and text* exemplars would make scoring criteria more apparent to students and thus serve as a more effective instructional tool when compared to *text only* exemplars. Further study of this phenomenon is needed before firm conclusions are drawn about the comparison. Future research on gender differences in learning statistics through same-gender groups could address the ability composition of groups. Students were assigned to mixed-ability groups to ensure that every group had a similar opportunity for success. There may, however, be a confounding between gender and ability composition in groups that must be considered in fostering group problem solving in statistics using computers. That multiple means of assessment provided a better profile of learning has been established in previous research. What needs additional examination is the relationship between individual and group assessment.

Review of NCRMSE Research

Are NCTM's Curriculum Standards World Class?

In April 1994, NCRMSE Director Thomas A. Romberg delivered a paper, The Mathematical Sciences Education Perspective of World Class Standards, at the Annual Meeting of The National Council on Measurement in Education. Excerpts are abridged and the data provided in tables and figures removed in the version of the paper presented here. The entire paper is available from the Educational Resources Information Center (ERIC) System under the number TM 021 285.

The use of the term *world class standards* in the current political debates grows out of the rhetoric surrounding the National Educational Goals (U.S. Department of Education, 1990). Goals 3 and 4 mention mathematics, and mathematics is implicit in Goal 5. In the debates, the *Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989) are often cited as the exemplar for establishing content standards for other core disciplines. This paper examines the problem of judging what is world class and then explains the mathematical sciences education views about world-class standards.

Since NCTM's *Curriculum Standards* have become our operational definition of a world-class mathematics program, questions about whether they are indeed world class need to be answered. To do this, my staff and I at the National Center for Research in Mathematical Sciences Education (Romberg et al., 1991) examined the mathematics frameworks of eight countries (Austrailia, France, Germany, Japan, The Netherlands, Spain, Norway, and The United Kingdom) and compared those frameworks with NCTM's *Curriculum Standards*. We found considerable variation in what is taught, in when ideas are introduced, and in what is emphasized. Thus, there is no international norm against which one can compare American views of what it is important to teach and learn in school mathematics. Nevertheless, we believe that the following statements about the vision of school mathematics presented in NCTM's *Curriculum and Evaluation Standards for School Mathematics* qualify the *Standards* as a world-class mathematics reform document.

?? The NCTM *Standards*, when compared with the national curricula of other countries, do not represent a "radical" or "romantic" version of school mathematics (p. 40).

?? The manner in which countries build a detailed rationale for these reforms and, specifically, which changes are emphasized, depends on their past practices (p. 41). For example, "number sense" and "estimation" are specifically mentioned in NCTM's document for Grades K-4, but not in those of other countries. This does not mean the topics are unimportant in other countries, only that they have been central in their curricula for decades—but not in ours—thus, no emphasis is needed.

?? The four standards for mathematics teaching and learning, problem solving, communication, reasoning, and connections are reflected in all eight national curricula, not just in NCTM's *Standards*. The terms used for these standards may differ, but the underlying themes are consistent (p. 41).

?? We are convinced that the variation in emphasis with respect to particular mathematical topics also is related to past cultural practices in different countries (p. 41).

?? In Grades K-4, the *Standards*, while including topics new to the American curriculum, still put more emphasis on whole number arithmetic than other countries. This same finding applies to all work with numbers up to Grade 8. However, there is no standard on "number" for Grades 9-12. Other countries appear to be more balanced in their approach to this important aspect of mathematics (p. 41).

?? While geometry is now included at all levels in the *Standards*, which, of course, is not reflected in most American classrooms, it still receives less emphasis on statistics, probability, and discrete mathematics than do other countries (p. 42).

??
Although the beginning principles of calculus are now being suggested for all students in the *Standards*, most other countries have long assumed this to be important and, in fact, expect much more than is advocated by the NCTM (p. 42).

?? Although different programs for students are common in other countries, their students are expected to study mathematics every year they are in school and are often offered several options. In the United States, the radical recommendation in the *Standards* that all American students study real mathematics for at least three years of high school falls short of the expectations of most other countries (p. 42).

The most striking difference between the NCTM *Standards* and the curriculum documents from other countries lies in the emphasis the other countries place on the social and attitudinal aspects of schooling. Schools need to be “a secure environment and place of trust,” “social behaviors” need to be taught, “students should realize that mathematics is relevant,” “students should gain pleasure from mathematics,” and “personal qualities should be nurtured” are statements that appear often in these documents. Such statements put an emphasis on what happens in classrooms that is different from the focus on either cognitive learning or economic imperatives in the *Standards* (p. 42).

In conclusion, one cannot study the curricular documents from these countries without realizing that the current mathematics curriculum in the United States is far from being world class. On the other hand, NCTM’s *Curriculum Standards* (1989) present a vision of content that is significantly in line with what other countries are now doing and with what they are planning to do. The expectations expressed in this vision, if realized in the schools, would bring all American students more in line with the expectations for students in the rest of the world.

Judging Whether Something is World Class

Given the difficulty of our attempt to judge whether NCTM’s Standards are world class, it is apparent that the political rhetoric in the National Goals (1991) implies we know how to judge whether students’ achievements in any country are indeed world class. In retrospect, to judge something implies either certifying that certain criteria have been met, or rank ordering a set-of-somethings on the basis of specific criteria. For National Education Goal 3, the somethings are students whose work is to be compared against the certification criterion, competency in challenging subject matter, for Grades 4, 8, and 12. Competency is to be defined so that all students learn to use their minds, are prepared for responsible citizenship, further learning, and productive employment—a difficult set of inferences. For Goal 5, similar certification criteria are implied for all adults. One would need to establish the connections between specific knowledge and skills in mathematics to be both employable and a responsible citizen. For National Education Goal 4, the something is a composite profile of U.S. students on achievement to rank order the American profile with those of other countries. Note that the judgments for Goals 3 and 4 are quite different judgments based on different criteria. Then, to judge whether something is world class involves either determining both that the certification criteria are comparable among nations and that the percentage of somethings that meet those criteria are comparable, or that the method of rank ordering on an attribute is reasonable across nations.

Thus, for one to argue that judgments related to Goals 3 and 5 are world class, one would have to build the case that the American certification criteria for all students (and adults) and goals are comparable with those of other nations, and also the percentage of students (adults) who meet those criteria is comparable. Evidence to build such a case for Goal 3 could be drawn by comparing the expectations in different countries for all students at these grade levels, along with comparisons of instructional programs and of school cultures. Evidence for Goal 5 for mathematics would involve a similar argument. For Goal 4, one could build a case for appropriate rank ordering of profiles if one could agree on how and when to assess mathematical achievement. In summary, to build such arguments about what is world class, one would need at least to examine the variations in student outcomes, and expectations for students, programs, and even school cultures across countries. The mathematical sciences education community assumes that any such comparisons should be based on our vision of school mathematics and not on current practice.

Compare Outcomes

A straightforward way of making comparison between the mathematical achievement of students in several countries is to administer a common test to a sample of students at the same level of schooling in each country. This has been done several times in the recent past and another similar test is now in the final stages of planning. The central question that needs to be addressed is: How valid is the test as a measure of student achievement across nations?

To study the world-class validity of such tests, during the past few years my staff and I were asked by the National Center for Educational Statistics to examine the items administered in the two past comparative studies with respect to NCTM's *Curriculum Standards*: The test battery for the Second International Mathematics Study (SIMS) administered by the International Association for the Evaluation of Educational Achievement (1985) and the battery for the International Assessment of Educational Progress (IAEP) administered by Educational Testing Service (1990). The question we addressed was: Does the content of these tests reflect that expressed in the *Standards*? Of course, such a comparison is problematic, since these tests were developed before the *Standards* were written.

The results of our analysis of each battery at each grade level were similar (Romberg, Smith, Smith, & Wilson, 1992). From the categorization of the Grade 8 SIMS battery one can only conclude that the content coverage is out of balance. Performance profiles based on these tests cannot be used to make a valid judgment about world-class mathematics achievement for American students, let alone those of any other country.

Compare Expectations

For expectations one could examine that high-stakes examinations administered in different countries. For example, at Grade 12 one could examine the high school completion exams, such as those given in the different states in Australia, or in Norway, or in The Netherlands. But what would they be compared with in the U.S.? Chantal Shafroth (1993) recently made such a comparison. For the United States, she used the SAT Level I and Level II tests, and the AP Calculus Test. Her analyses are followed with an examination of the different types of questions posed. Even a cursory look at this report indicates that there are vast differences in mathematical expectations among those countries and between them and the United States.

A second type of high stakes examination is college admission tests. For example, the Mathematics Association of America recently published a set of admission examinations from several Japanese universities (Wu, 1993). The reason for publishing the examinations was to demonstrate to American mathematicians and mathematics educators the fact that the Japanese ask entering college students mathematical questions that many American college majors in mathematics would have difficulty answering.

Similar comparison could be made with respect to the examinations administered in some countries at the other levels of schooling. In fact, the New Standards Project (Resnick, Nolan, & Resnick, 1994) is currently studying examinations administered in The Netherlands and France for 8^{th} graders to compare with tests they are developing.

In summary, while very incomplete, the evidence from these reports indicates that there are quite different mathematical expectations for students in other countries. It should be noted that the differences in both terminal and university entrance examinations are not just that some countries expect their students “to cover more” mathematics. This assumes that the discipline is a linear sequence of topics. The fact is that countries differ in what they consider to be the important mathematics all students should learn. Some of these differences may be overt, such as teaching transformation geometry or measurement with metric units; others are covert, such as a focus on number sense rather than on computational proficiency.

Compare Programs

To compare programs, one could examine curricular frameworks. Geoffrey Howson (1991) examined the national curriculum frameworks for mathematics for 14 countries. The United States was not one of these, since we do not have a centralized education system. Howson concluded “There is no ‘easy’ way of comparing what is done and what is achieved in various countries. Major differences in philosophy and structure make simplistic comparison dangerous” (p. 7).

He goes on to state, “National school systems reflect social cultures and traditions, and are much influenced by economic considerations, past and present. Perhaps the simplest measure of the latter is the length of compulsory schooling. This can be as high as 12 years, but Portugal, for example, is only now moving away from a 6-year system” (p. 7). He also found that while politicians argue that there is a need to “forge a closer link between the national curriculum and assessment procedures than would appear to exist in any of the other countries below…there are references to assessment in few national curricula…" (p. 28).

Ken Travers and Ian Westbury (1989) reported on a more extensive, but perhaps less useful, analysis of mathematics curricula across the 22 countries that participated in SIMS. The focus of this study was on the relationships between the "intended curriculum," the "actual curriculum," and the "attained curriculum." The utility of their analysis is hampered by the focus on variation across countries, on specific features (making it difficult to get a sense of any one country's program), and on the relationships to the SIMS item pool.

Compare Cultures

The differences in school cultures across nations with respect to the teaching of mathematics has been systematically studied by Stevenson and Lee. They studied the context of achievement for a sample of American, Chinese, and Japanese children (1990) and concluded that the performance of American children in their study was due to several factors that were neither "elusive nor subtle."

Some of the most salient reasons for poor performance appear to be the following: Insufficient time and emphasis were devoted to academic activities; children's academic achievement was not a widely shared goal; children and their parents overestimated the children's accomplishments; parental standards for achievement were low; there was little direct involvement of parents in children's schoolwork; and an emphasis on nativism may have undermined the belief that all but seriously disabled children should be able to master the content of the elementary school curriculum. (p. 103)

In a less formal study, Jan de Lange (1992), whose staff at the Freudenthal Institute in The Netherlands has been working with mine to develop instructional materials for middle school students, has reported on the vast differences in the culture of schools in America and The Netherlands. These differences include governance, the role of administrators, the role of parents, the daily rites and rituals of schooling, scheduling, interruptions, athletics, and so on. While schools have been established by all societies to educate their children, there are vast differences in how different cultures have actually created and defined schooling.

Mathematical Sciences Education View

The Mathematical Sciences Education community believes that comparative studies are very important. We can learn by understanding how different countries decide what mathematics their students should learn; how they teach that mathematics; how they expect their students to learn that mathematics; and how they determine student progress, proficiency, or achievement. Such comparisons can make out *commonplace* actions and beliefs problematic.

Second, the community does not see the study of what other countries do as an attempt simply to keep abreast. We believe in the old adage: “Anyone who just wants to keep abreast is bound to be second best.” Comparative studies should be seen as opportunities for us to learn and reflect on our actions, and not simply as an attempt to copy the ideas of others.

Third, for achievement, the community recommends either the development of a more balanced examination system that is aligned with principles articulated in NCTM’s Standards, or the selective adaptation of methods other countries use to judge achievement and compare our students using procedures based on the examination systems of those countries.

Fourth, the community doubts that the overall effort of attempting to develop a common test battery similar to those in SIMS, IAEP, or NAEP upon which one can validly compare student achievement across countries at any grade level is worth the cost and effort. We are particularly concerned that future studies (including TIMSS) will compare student achievements across countries in a horse-race fashion on examinations that fail to reflect world-class aspirations for the students in any country, let alone those in the United States.

Finally, given the differences in schools and cultures, the community doubts that there can be any agreed upon criteria that could be called world class. To strive for such is merely political rhetoric. As such, the rhetoric may detract or undermine our efforts to make needed changes in schooling. Our work on mathematics and the teaching and learning or the subject in schools is only a small part of a much more serious need to restructure schooling in America.

References

De Lange, J. (1992, August). *Mathematics in American schools*. Paper presented at the International Congress on Mathematics Education, Quebec City, Canada.

Educational Testing Service. (1990). *International Assessment of educational progress*: 1991 mathematics assessment. Secured and publicly released test items.

Howson, G. (1991). *National curricula in mathematics*. England: The Mathematical Association.

International Association for the Evaluation of Educational Achievement. (1985). *Second study of mathematics: Technical report IV, instrument book*. Publicly released test item.

National Council of Teachers of Mathematics (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

Resnick, L.B., Nolan, K.J., & Resnick, D.P. (1994). *Benchmarking education standards*. Paper presented at the Conference on Uses of International Education Data. Washington, DC: National Academy of Sciences.

Romberg, T.A., Allison, J., Clarke, B., Clarke, D., Pedro, J., & Spence, M. (1991, November). *School mathematics expectations: A comparison of curricular documents of eight countries with the NCTM Standards of the U.S*. Prepared for the New Standards Project.

Romberg, T.A., Smith, M., Smith, S., & Wilson, L, (1992, June). *The feasibility of using international data to set achievement levels for the National Assessment of Educational Progress (NAEP).* Madison: National Center for Research in Mathematical Sciences Education, University of Wisconsin-Madison.

Shafroth, C. (1993). A comparison of university entrance examinations in the United States and in Europe. *Focus*, 13(3), 1, 11-14.

Stevenson, H.W., & Lee, S. (1990). Context of achievement. *Monographs of the Society for Research in Child Development*, 55(1-2) (Serial No. 221).

Travers, K.J., & Westbury, I. (1989). *The IEA study of mathematics I: Analysis of mathematics curricula*. New York: Pergamon Press.

U.S. Department of Education (1991). *America 2000: An education strategy*. Washington, DC: Author.

Wu, L.E.T. (Ed.). (1993). *Japanese university entrance examination problems in mathematics*. Washington, DC: The Mathematical Association of America.