

**UNITED STATES DEPARTMENT OF EDUCATION
NATIONAL MATHEMATICS ADVISORY PANEL MEETING**

September 7, 2007

SUMMARY

The National Math Panel met in open session at the Eric P. Newman Education Center, 320 South Euclid Avenue, St. Louis, Missouri 63110, on Friday, September 7, 2007, at 8:30 a.m.

PANEL AND EX OFFICIO MEMBERS PRESENT:

DR. LARRY FAULKNER	Chair
DR. CAMILLA PERSSON BENBOW	Vice Chair
DR. DEBORAH LOEWENBERG BALL	Member
DR. SUSAN EMBRETSON	Member
DR. FRANCIS (SKIP) FENNELL	Member
DR. BERT FRISTEDT	Member
DR. DAVID GEARY	Member
DR. RUSSELL GERSTEN	Member
DR. TOM LOVELESS	Member
DR. VALERIE REYNA	Member
DR. ROBERT SIEGLER	Member
DR. SANDRA STOTSKY	Member
MR. VERN WILLIAMS	Member
DR. HUNG-HSI WU	Member
DR. IRMA ARISPE	Member
DR. DANIEL B. BERCH	Ex Officio
DR. JOAN FERRINI-MUNDY	Ex Officio
MR. RAYMOND SIMON	Ex Officio
DR. GROVER (RUSS) WHITEHURST	Ex Officio

PANEL AND EX OFFICIO MEMBERS NOT PRESENT:

DR. A. WADE BOYKIN	Member
DR. DOUGLAS CLEMENTS	Member
DR. LIPING MA	Member
DR. WILFRIED SCHMID	Member
DR. JAMES SIMONS	Member

STAFF PRESENT:

MS. TYRRELL FLAWN, EXECUTIVE DIRECTOR
MS. MARIAN BANFIELD
MS. IDA EBLINGER KELLEY
MS. JENNIFER GRABAN
MR. JIM YUN
MR. KYLE ALBERT

CALL TO ORDER:

Chair Faulkner convened the public session of the National Mathematics Advisory Panel and stated that the focus of the session would be to receive reports from the subcommittees and task groups of the Panel. These groups have been carrying out a great volume of work outside of the public sessions. He explained that the Panel is in the process of wrapping up the work of the task groups and will go from this stage of receiving task group reports into a stage in its next meeting in Phoenix, Arizona, largely focused on synthesizing a Panel report. This will be an overarching report with the whole Panel's message to the constituencies interested in this report. He highlighted for the audience that they are at a major turning point in their activity and are about to move out of subcommittee-based activity, into whole-panel activity.

Chair Faulkner then asked if signing services were needed, and they were not.

OPEN SESSION:

TASK GROUP AND SUBCOMMITTEE REPORTS: INSTRUCTIONAL MATERIALS SUBCOMMITTEE

Bob Siegler, Chair; Dan Berch; Bert Fristedt; Vern Williams; Irma Arispe.

Dr. Siegler stated that, as the Subcommittee has just started their work, they are not as far along as the other groups presenting. They are going to be looking at a variety of sources of evidence, much of it from the reports of the other Panel groups. They also have other available materials, such as the National Opinion Research Center (NORC) National Survey of Algebra Teachers, an NRC Report on instructional materials, some mathematicians who have written about evaluating textbooks for accuracy, and a variety of other sources.

Dr. Siegler reported that they have decided to divide their task into three main parts: 1) evaluating textbooks, one important kind of instructional material; 2) evaluating ancillary materials; and 3) evaluating knowledge creation mechanisms.

For textbooks, they are going to be looking at two main things. One of them is the mathematical accuracy of textbooks, and the other is ~~a cluster of concepts and dimensions~~ number of reasons that have led to textbook manufacturers' reports that the average third-grade math textbook is 750 pages and the average eighth-grade math textbook is more than 1100 pages. They will go into these reasons ~~for that~~, and will compare these to textbooks in other countries to see if the country really needs to have textbooks that are this long. These issues get into questions of coherence and sequencing. They also will address whether the sheer length and diversity of topics interfere with the coherence and logical sequencing of textbooks.

Their second focus is ancillary materials, i.e., materials other than textbooks that are used in instruction. They include calculators, computer software, teacher manuals, and support materials for diverse students of varying ability levels.

Finally, the Subcommittee will review knowledge creation mechanisms, the What Works Clearinghouse, and areas that are particularly in need of greater research.

They are charged with writing five to eight paragraphs. The group has identified the topics at this meeting, and then different group members will be drafting various paragraphs prepared in a week from today when they will talk on the phone. They will complete the report a week after that.

Dr. Fristedt added that they would look at existing resources to make recommendations.

Chair Faulkner stated that this group has been charged with developing language that might be effectively incorporated into the Panel report rather than trying to develop a detailed study in this area. The Panel Report, as a whole, is being targeted for something in the range of

thirty published pages, which would have to cover the activities of all the task groups and subcommittees, and deal with introductory material and so forth.

INSTRUCTIONAL MATERIALS SUBCOMMITTEE: QUESTION AND ANSWER PERIOD

Dr. Whitehurst asked if other group members will allude to the National Academy of Science's report on curriculum and textbooks that came out several years ago. Dr. Siegler responded that several group members read that, but they found it only a little bit helpful.

Dr. Ball, in reference to length of textbooks, asked if they were going to try to be analytical about what the sources of the length were, for example, to support the range of capacities that teachers need to teach well or what students are expected to complete. Dr. Siegler responded that textbook manufacturers brought up in Cambridge the issues of extensive use of large color photographs that have little to do with the content that is being captured, or materials that are required in some states but not in the state in which a book is being used. Dr. Ball also asked whether the group would be looking at teacher support materials. Dr. Fristedt replied that teacher materials should address areas of weakness for teachers and focus on helping the teacher recall and get back to that particular aspect of mathematics. What should not be there are things that fill up the margins with extraneous comments. Dr. Siegler added that if they know of any articles on this topic that the Panel should look at, please recommend them to the group.

Dr. Fennell stated that the intersection of the mathematics, the pedagogy, and the marketability of a program lead to a fattening of the teacher materials.

Dr. Loveless asked if the Subcommittee will be reviewing the assessment materials that come with textbooks. He knows they vary a great deal in terms of the numbers of quizzes and unit tests. Dr. Siegler stated that they are open to the idea. Dr. Wu remarked on the issue of length, and concurred with the Subcommittee's concern. The most coherent textbooks he has seen are extremely thin and are available from the American Mathematical Society--the Japanese textbooks of Grades 10-11. There are various accounts from publishers that what is in these sources is present because they found that those elements were those things demanded by teachers. Dr. Siegler agreed about the market factors that go into all this, but feels that market factors do not make up the total basis. Other points of concern include the fact that different states require different topics to be taught in different grades, and the clustering of issues.

Dr. Fristedt added that at this point in the Panel's report, they are thinking about different issues that can be brought together in a unified way, such as the work of the Assessment Task Group and the Instructional Practices Task Group section on formative assessment.

Dr. Stotsky asked if there was any mention in research about cost of textbooks. Dr. Siegler stated that they would look at that if time permits.

Dr. Ball asked at what meeting they all would get to discuss exactly what they are going to be concluding about textbooks. Chair Faulkner stated that it would be in the working material for the synthesis groups to start putting into a Panel report.

Chair Faulkner stated that there are conflicts of interest in some cases that limit what people can say. The Subcommittee will not be commenting on individual products, but rather on the state of knowledge.

TASK GROUP AND SUBCOMMITTEE REPORTS: STANDARDS OF EVIDENCE SUBCOMMITTEE

Valerie Reyna, Chair; Camilla Benbow; Wade Boykin; Russ Whitehurst; Mark Lipsey, consultant.

Dr. Reyna stated that the Subcommittee's task was to marshal the best scientific evidence in the service of producing evidence-based mathematics instruction. They define highest quality evidence as evidence that is high in both internal and external validity. That means excellence of design in terms of internal validity, methodology and rigor and scientific soundness. External validity naturally refers to the ability to generalize beyond the sample that is studied, to different and diverse populations in various circumstances. They also distinguished therefore, highest quality evidence, which is high in both internal and external validity, from promising or suggested evidence. In addition, they will identify areas that would benefit from further research. They also recognize that the Panel must consider expert opinion in some situations where scientific trials are inapplicable or unavailable.

The Subcommittee distinguished different kinds of questions that involve survey methodology. Their strongest confidence was reserved for studies that actually test a hypothesis. These are the kinds of studies where it is possible to disprove the hypothesis that one started out with. The Subcommittee is also interested in studies that have been replicated with diverse samples. The What Works Clearinghouse has been dealing with many of these issues, and the Panel cites them as a reference.

Moderately strong evidence might come, for example, from the effects shown in one or two high quality studies, not necessarily independently replicated. Suggestive evidence might be represented by the situation where high quality studies exist in support of a conclusion, but other studies have a null result. Inconsistent evidence would be trumped by high quality study designs. Weak evidence is where there are only low quality studies available.

Effects of interventions are things that involve random assignment to condition. Low attrition is a mark of high quality, valid and reliable measures. A descriptive survey has to have a representative sample, a low non-response rate and evidence that attrition was not biased.

Tests and assessments are subject to a variety of psychometric standards, including some of the measurement issues mentioned, such as validity, reliability and sensitivity.

The methodology statement has a set of references, despite the fact that it is not about empirical evidence.

Dr. Reyna concluded that all of the task groups were charged to have some recommendations, and their recommendation inevitably revolve around standards of methodology. Many other members of the Panel noted that they had to whittle through a number of studies that really did not pass methodological muster. Many of these failed to meet standards because they do not permit strong inferences about causation or about causal mechanisms; therefore, the Subcommittee recommended that the rigor and amount of course work in statistics and experimental design be increased in graduate training and education.

Chair Faulkner stated that the Panel is at a stage where the members need to carry out final discussion and make a formal adoption of this report as the basis for the Panel's activity.

STANDARDS OF EVIDENCE SUBCOMMITTEE: QUESTION AND ANSWER PERIOD

Dr. Fristedt noticed that in some places the term scientific evidence is used and at other places the term evidence is used. He thinks it is important that both terms be used.

Chair Faulkner sought a motion to adopt the Subcommittee report guidelines for standards of evidence.

Dr. Fennell so moved. Dr. Gersten seconded that motion. Chair Faulkner then asked that all in favor of adoption please signify by saying aye. All Panel members stated aye, and there were none opposed.

Chair Faulkner thanked the Subcommittee for their work and stated that the reports that are about to come forward from the task groups have been developed using the standards of evidence that were represented here.

Chair Faulkner turned over the presiding duties to Vice Chair Camilla Benbow as he assisted with the presentation of the Conceptual Knowledge and Skills Group.

TASK GROUP AND SUBCOMMITTEE REPORTS: TASK GROUP ON CONCEPTUAL KNOWLEDGE AND SKILLS

Francis “Skip” Fennell, Chair; Larry R. Faulkner; Liping Ma; Wilfried Schmid; Sandra Stotsky; Hung-Hsi Wu; Tyrrell Flawn, staff

Dr. Fennell began by giving an overview of the section titled, “What are the Major Topics of School Algebra.” These are the essential mathematic concepts and skills that lead to success in algebra and should be learned as a prerequisite to algebra. This leads to the question of how the sequence of topics prior to formal algebra course work affects achievement in Algebra.

Dr. Stotsky gave an overview of the methodology used by the group, which is a combination of peer reviewed and published studies, as well as expert judgment. The task group also provides a context for student achievement in mathematics. The Major Topics in School Algebra reflect the judgment of the mathematicians on the Panel and other mathematicians with whom they have consulted. This is the centerpiece that launches the rest of the report.

The secondary themes included a look at different sources that set forth algebra topics: 1) state standards for Algebra I and II; 2) Algebra I and II textbooks; 3) Singapore’s mathematics curriculum for Grades 7 to 10 [the leading country on the Third International Mathematics and Science Survey (TIMSS)]; 4) Grade 12 National Assessment of Educational Progress (NAEP) test objectives; and 5) proposed American Diploma Project Benchmark and Test Objectives for its Algebra II end of course test, which is now being piloted by a group of states on a voluntary basis.

Dr. Stotsky explained the the Task Group has a comparisons section that shows how all of these different sources of topics reflect on the major topics the task group has identified. The next section is called “Observations Regarding Rigor,” which includes some material on mathematical errors found in contemporary Algebra I and II textbooks.

The task group report then moves to how to help all students grasp those major concepts and skills, and what they should learn as preparation to arrive at formal algebra, as taught at the end of middle school or early high school. They reviewed international approaches to pre-algebra education, particularly the work of William Schmidt and his colleagues, on the curricula of the A+ countries—the six leading countries on the TIMSS test. They also reviewed comparisons of curriculum approaches in these top-performing countries to U.S. mathematics education.

The task group then reviewed some national approaches to pre-algebra U.S. education, including 1) National Council of Teachers of Mathematics’ (NCTM) *Curriculum Focal Points*, 2) curriculum profiles of the six highest rated U.S. curriculum frameworks, 3) the ACT Curriculum Survey, and 4) the Panel’s National Algebra Teacher Survey. What comes out of this is the “Critical Foundations of Algebra.”

The task group then examined whether the sequence of mathematics topics during formal algebra course work affects algebra achievement. They found that there was no research they could draw on for answering that question.

The next question they reviewed was if there was research on the benefits of an integrated or single subject approach for the study of algebra. While there might be a large

number of studies, no conclusions could be drawn from what was a very deficient body of research.

The next question they looked at was when formal algebra course work would be best addressed. There was a small body of research to draw on that met the criteria for the Panel's standards of evidence.

Chair Faulkner then presented the list of conclusions and recommendations, which included the overarching point that algebra is a gateway to more advanced mathematics and most post-secondary education. All schools and teachers must concentrate on providing a sound and strong mathematics education to elementary and middle school students so that all can enroll and succeed in algebra. It is much more important for students to be soundly prepared for algebra and then well taught in algebra, than to study algebra at any particular grade level. This task group is supportive of beginning students who are ready earlier than at a traditional grade level, perhaps Grade 8, or in some cases even earlier. What is important is that students get legitimate algebra courses, that they are well taught and that the students be prepared for them, rather than that the students get an early start.

Dr. Fennell then presented the list of recommendations.

1) The Task Group recommends that school algebra be consistently understood in terms of the list of Major Topics of School Algebra (MTSA).

2) The list of Major Topics of School Algebra accompanied by a thorough elucidation of the mathematical connections among these topics should be the main focus of Algebra I and Algebra II standards in state curriculum frameworks, in Algebra I and Algebra II courses, in textbooks for these two levels of algebra, whether integrated or otherwise, and of course, assessments of these two levels of algebra. Supporting that statement, the task group also recommends use of the Major Topics of School Algebra in revisions of math standards at the high school level, in state curriculum frameworks, in high school textbooks organized by an integrated approach and in grade level state assessments using an integrated approach at the high school level by Grade 11, at the latest.

3) Proficiency with whole numbers, fractions and particular aspects of geometry, are the Critical Foundations of Algebra. Emphasis on these essential components and skills must be provided at the elementary and middle grade levels. The coherence and hierarchical nature of mathematics dictate the foundational skills that are necessary for the learning of algebra. By the nature of algebra, the most important among them is proficiency with fractions, which the Task Group defines to include decimals, percents and negative fractions. The teaching of fractions must be acknowledged as critically important and improved before an increase in student achievement in algebra can be expected.

4) International studies show that high achieving nations teach for mastery in a few topics, in comparison with the mile-wide, inch-deep curriculum. A coherent progression with an emphasis on mastery of key topics should become the norm in elementary and middle school curricula. There should be a de-emphasis on the spiral approach that continually revisits topics year after year without closure.

5) Federal and state policy should give incentives to schools to offer an authentic Algebra I course in Grade 8 and to prepare a higher percentage of students to enter the study of algebra by Grade 8. Care must be taken to ensure that such a course addresses algebra as described in Recommendation 2.

6) Publishers must ensure the mathematical correctness of their materials. Those involved with developing mathematics textbooks and related instructional materials need to engage mathematicians as well as mathematic educators in writing, editing, and reviewing these materials.

7) Adequate preparation of students for algebra requires their teachers to have a strong mathematics background. To this end, the Major Topics of School Algebra and the Critical

Foundations of Algebra must be fundamental in the mathematics preparation of elementary and middle school teachers.

TASK GROUP ON CONCEPTUAL KNOWLEDGE AND SKILLS: QUESTION AND ANSWER PERIOD

Mr. Simon stated that he has always felt from the beginning that the Conceptual Knowledge and Skills section was going to be the heart of this whole report because that is where a majority of the people who implement U.S. mathematics education will go. Dr. Fennell asked if there was anything missing. Mr. Simon stated that the Critical Foundations section does not go far enough to give guidance to the teachers in the state departments that are setting standards and looking at this research. While not telling the states what they need to do—respecting their curricula and individual differences—the Conceptual Knowledge and Skills Task Group and Panel need to assist the growing number of schools and parents that do not know what to do. He asked the task group to consider putting in some benchmarks at certain grade levels that would also address the spiraling issue they talked about.

Dr. Fristedt agreed with much that they have written, but he has some problems with the organization and the messages that typical readers might get out of it. For example, he saw the following words—algebra, formal algebra, authentic algebra, legitimate algebra, algebra, Algebra I, and Algebra II. He stated that this is a communication issue that should be dealt with in a very systematic manner. He also asked about the word leading up to algebra, which was arithmetic. He shared concern about being too specific about what is Algebra I and Algebra II, especially when the point is that the list of topics under algebra be fundamental.

Dr. Clements asked if the “Critical Foundations” represent the entire elementary curriculum or just those foundations leading to algebra. Dr. Fennell responded that they are recommending that children from preschool through algebra, regardless of what grade that happens, engage in a full curriculum in mathematics. By a full curriculum, he means measurement; some opportunity to analyze data; adding, subtracting, multiplying, and dividing whole numbers; fractions, including decimals, and percents; and particular aspects of geometry.

Dr. Loveless asked about the curriculum of the six countries that perform at the top on TIMSS, and whether the task group reviewed the bottom six or the middle six. Dr. Fennell stated that they did not. Dr. Clements added that it is important to at least have a caveat about that. Dr. Loveless added that when they run regressions and use curricular variables to explain variance in TIMSS scores they do not get a huge effect.

Dr. Wu agreed with Dr. Fristedt about the multi-faceted use of the word algebra. That should be clarified to make sure that it is clear. He also thought the task group had agreed to use “Combinations and permutations as applications of the binomial theorem and Pascal’s Triangle” as the descriptor for Combinatorics and Finite Probability in the list of algebra topics. Dr. Fennell noted that the report does list it out separately.

Dr. Benbow seconded Mr. Simon’s suggestion earlier about benchmarks and thinks it will be very helpful to state education departments in terms of developing curricula to be proud of. She also asked the Task Group to talk a little bit more about students taking algebra before the eighth grade. Dr. Fennell stated that they have to make sure that all children have access to algebra when they are ready. For students who are ready, they should have access to it. The student needs some prerequisite knowledge before that mathematics is begun.

Dr. Stotsky added that the report talks about offering algebra in Grades 7 and 8, as the research does mention that possibility.

Dr. Siegler stated, in reference to age and grade norms for teaching particular topics that there is nothing in either the empirical evidence from psychology nor in the logic of mathematics that says that topic X should be taught in grade N . The NCTM *Focal Points* does the best job possible currently to provide reasonable recommendations, and the Panel should steer clear on

linking specific grade levels to particular topics. Dr. Fennell noted the need to take all of this under advisement, but that is a very good point.

Chair Faulkner reminded the audience that the reports they are hearing are very substantial—far along drafts—but they are not complete. They are complete enough for the task groups to give a strong indication of what those reports will say at the end; even so, significant changes can be expected.

TASK GROUP AND SUBCOMMITTEE REPORTS: TASK GROUP ON LEARNING PROCESSES

Dave Geary, Chair; Wade Boykin; Susan Embretson; Valerie Reyna; Bob Siegler; Dan Berch; Jennifer Graban, staff

Dr. Geary stated that the methodologies used in the conclusions of this task group were based primarily on studies that test explicit hypotheses about the mechanism promoting the learning of mathematics. The evidence regarded as strongest for this purpose is that which shows convergent results across procedures and study types. When the evidence is not strong, conclusions are qualified, and suggestions for research that will strengthen the ability to draw conclusions are provided. There were multiple approaches, procedures, and study types reviewed and assessed with regard to convergent results using a variety of methodologies.

With respect to the literature search, the Task Group looked at key mathematical content terms linked with learning and cognitive processes. Their first search studied core peer-reviewed learning, cognition, and developmental journals. They then conducted a second search that included other empirical journals, indexed and Psych Info and Web of Science.

Criteria for inclusion were that the studies were published in English; participants were age three and older; publication took place in a peer-reviewed empirical journal, review of empirical research, book, or annual review; and the methodology was experimental, quasi-experimental or correlational.

Dr. Geary then gave a brief overview of some of the types of things that the Task Group covered under general principles, from cognition to learning. There is a great deal of scientific knowledge on learning and cognition that can be applied to improve student achievement, but it is not currently being used in the nation's classrooms.

Basic research and factors that promote learning provide an essential grounding for the development and evaluation of effective educational practices. As an example, inherent limits on working memory capacity can impede proficient performance in mathematics. Practice can offset this limitation by achieving automaticity, which frees up working memory resources.

The learning of facts, algorithms and concepts are interrelated. Conceptual knowledge aids in the choice of algorithms. Practice of algorithms can provide a context for making inferences about concepts. Committing facts to long-term memory allows attention to be focused on more complex problem features. Conceptual understanding promotes transfer of learning to solve new problems and results in better long-term retention, higher-order thinking, and problem solving.

Mathematical knowledge that children from both low- and middle-income families bring to school influences their learning for many years thereafter, probably throughout their education. Dr. Geary stated that several effective programs have been developed to improve the mathematical knowledge of preschoolers and kindergartners, especially those from at-risk backgrounds. Nonetheless, many children and adults in the United States do not solve simple arithmetic problems as fast and efficiently as their peers in other nations, because they have not practiced these problems frequently enough. The learning of algorithms to solve complex arithmetic problems is influenced by working memory, conceptual knowledge, degree of mastery

of basic facts, and practice. Learning is most effective when practice using algorithms is combined with instruction of related concepts.

In the area of social, motivational, and affective influences, Dr. Geary noted that Vygotsky's socio-cultural perspective has been influential in education. The theory treats learning as a social induction process by which learners become increasingly able to function independently through the tutelage of more knowledgeable peers and adults. However, due to a shortage of controlled experiments, the usefulness of this approach for improving math learning is difficult to evaluate at this time.

There is empirical research on other factors that influence and can improve mathematical competence. Self-regulation--the ability to set goals, plan, monitor, and evaluate progress--is correlated with mathematics achievement. Anxiety about mathematics performance lowers test scores, but there are interventions that significantly reduce anxiety and improve test scores. Young children's intrinsic motivation to learn is positively correlated with academic outcomes in mathematics, but it declines across grades as material becomes increasingly complex. There are educational interventions that can influence intrinsic motivation in later grades.

Relative to children in nations with high mathematics achievement, children in the United States tend to attribute mathematical achievement more to ability than to effort. Experimental studies have demonstrated that children's beliefs about the relative importance of effort and ability can be changed, and that increased emphasis on the importance of effort is related to improved mathematics grades.

Dr. Geary then turned to the section on what children bring to school. Most children begin school with a fair amount of numerical knowledge. The mathematical knowledge that children from low- and middle-income families bring to school influences their learning mathematics and achievement for many years thereafter. The numerical knowledge of children from low-income backgrounds lags even before they start school. Promising instructional programs exist for increasing low-income preschooler's numerical knowledge. Studies that evaluate the effectiveness of the scaled up application of these programs are recommended.

The Task Group reviewed the research on mathematical content areas and they organize their recommendations around classroom practices or research needed to facilitate these practices. This includes training of teachers and future researchers; curriculum, including content and textbooks; and basic and applied research in these areas. For all of the areas, a pipeline of research must be funded that extends from the basic science of learning to field studies in classrooms.

Cognitive studies indicate that many children do not master whole number arithmetic. In comparison to children of many other nations, it takes U.S. children many more years to become fast and efficient at solving basic arithmetic problems. They frequently make errors when using standard algorithms. Error patterns suggest poor conceptual knowledge.

By the end of elementary school, a majority of children do not appear to understand many basic concepts, including the distributive property and the inverse relation between division and multiplication. The research base for core arithmetical procedures and concepts that are crucial for learning algebra, such as division algorithms and the distributive property, is inadequate. Few curricula in the United States provide sufficient practice and strong conceptual context for this practice. Studies of how to best organize this practice and with well defined outcomes are needed to guide curriculum development.

Priorities include expanding the research base on children's learning of core concepts, promoting better understanding of the reciprocal relation between procedural and conceptual learning, and development of mechanisms that facilitate the translation of basic research into knowledge useable in the classroom.

The task group then looked at fractions, which are formally introduced in elementary school, yet remain difficult for many adults. Twenty-seven percent of eighth- graders cannot

correctly shade $\frac{1}{3}$ of a rectangle in the 2005 NAEP. Forty-five percent could not solve a word problem involving dividing fractions. For adults, poor understanding of fractions, decimals, and proportions is associated with poor medical outcomes (e.g, in the management of prescription doses).

Preschoolers show an intuitive awareness of fractions based on whole/part relations and sharing. Studies also show improved performance between ages four and seven, but understanding of fractions lags far behind understanding of whole numbers. As with whole numbers, conceptual and procedural knowledge of fractions reinforce and bootstrap one another and influence such varied tasks as estimation, word problems, and computations.

A key mechanism linking conceptual and procedural knowledge is the ability to represent fractions, first physically and ultimately with a mental number line. On-task time, motivation, working memory, well-learned basic arithmetic skills, and reading ability also determine performance on fractions problems. An absence of a coherent and empirically supported theory of fraction tasks is a major stumbling block to developing practical interventions to improve performance. Instruction focusing on conceptual knowledge of fractions is likely to have the broadest and largest impact on problem solving performance, provided that it is aimed at accurate solutions of specific problems.

The Task Group found that estimation is an important part of mathematical cognition because it is used in everyday life and in scientific, mathematical, and technical professions, and because it is closely related to overall mathematics achievement. Poor estimation performance often reveals underlying difficulties with understanding mathematics in general. Children's estimation of the magnitudes of fractions is especially poor, and programs for helping children in this area are urgently needed.

For geometry, Dr. Geary noted that the LP Task Group focused only on those aspects of geometry that were highlighted by the Conceptual Knowledge and Skills Task Group. Of the five mathematical content areas assessed by the 2003 TIMSS, U.S. eighth-graders' performance in geometry items was weakest. U.S. eighth-graders exhibited no significant improvement in geometry between 1999 and 2003 on the TIMSS, despite significant gains in algebra during this period. In comparison to high-achieving nations, the U.S. devotes only about half as much time to the study of geometry.

The component of geometry most directly relevant for the early learning of algebra is that of similar triangles. However it is difficult to draw firm scientifically based conclusions from the empirical research on children's acquisition of similarity and related concepts.

Piaget theorized that the representation of space develops from topological, to projective, to Euclidean. The task group concludes that the mathematical inaccuracies of this hypothesis along with the mounting negative empirical evidence, suggests that it should no longer inform the design of instructional approaches in geometry.

One of the challenges to effective learning in geometry is the persistence of misconceptions and the resistance to instruction. One example of this is the illusion of linearity, where students incorrectly believe that if the perimeter of a geometric figure is enlarged k times, its area or volume is enlarged k times as well.

Young children possess at least an implicit understanding of basic facets of Euclidean concepts, although formal instruction is needed to ensure that children adequately build upon and make explicit this core knowledge so they can learn formal mathematical geometry.

Despite the widespread use of mathematical manipulatives, evidence regarding their usefulness is tenuous at best. Students must eventually transition from concrete, hands-on, or visual representations to internalized abstract representations. The crucial steps in making such transitions are not clearly understood at present. Studies are needed to demonstrate whether and

to what extent knowledge about similar triangles enhances the understanding that the slope of a straight line is the same regardless of the two points chosen, and thus leading to a mathematical understanding of linearity.

Cognitive studies of algebra focused on linear equations and word problems have revealed that many students in high school algebra courses are unprepared for learning the basics of algebra. The errors students make when solving algebraic equations reveal that many students do not have a firm understanding of the basic principles of arithmetic, and many do not understand the concept of mathematical equality. Students have difficulty grasping the syntax or structure of algebraic expressions and do not understand procedures for transforming equations or why transformations are done the way they are.

There are many gaps in our current understanding of how students learn algebra and the preparation that is needed by the time they enter the algebra classroom. Research efforts to close these gaps are recommended.

The task group then addressed individual and group differences, beginning with learning disabilities. The empirical evidence suggests that between 5 and 10% of students will experience a significant learning disability or learning difficulty in mathematics before completing high school. The corresponding cognitive deficits include a compromised working memory system and difficulties with basic concepts. These contribute to difficulties with whole number arithmetic learning. At the same time, much less is known about how these difficulties are related to learning fractions, estimation, geometry, and algebra.

The task group recommends funding of longitudinal and brain imaging studies that assess cognitive mechanisms underlying learning disabilities and core mathematical domains. Promising intervention studies are in progress and funding for additional studies is recommended.

The few cognitive studies of the sources of the accelerated learning capacity of mathematically gifted students suggest an enhanced ability to remember and process numerical and spatial information. Cognitive and brain imaging studies of the mechanisms that underlie their accelerated learning are needed to better understand how to help these students achieve their full potential.

For nationally representative samples, the average mathematics scores of boys and girls are very similar. When differences are found they are small and typically favor boys. There are consistently more boys than girls at both the low and high ends of mathematical performance on standardized tests, though differences at the high end have decreased significantly. Media attention to the over-representation of boys at the high end of mathematical performance has obscured the fact that relative to high achieving countries the achievement of both boys and girls in the United States is poor.

The section on race, ethnicity and socio-economic status is drafted but is still in preparation and not ready for discussion.

The task group's final content topic is brain sciences in mathematics learning. They note that brain sciences research has potential for contributing unique knowledge regarding mathematical learning and cognition, and for eventually informing educational practice. Funding of brain imaging studies that focus on children's learning in core mathematical domain is recommended. At the same time, the application of research in the brain sciences to classroom teaching and student learning in mathematics is premature, and structural programs in mathematics that claim to be based on brain sciences research remain to be validated.

Overall, the task group stated that for all areas, a pipeline of research must be funded that extends from basic science of learning to field studies of classrooms. They recommend incentives to encourage partnerships between basic and applied researchers. The effective interventions should be scaled up and evaluated in classrooms. Educational research must be integrated with basic research in cognition, motivation, neuroscience, and social psychology. Educationally relevant research need not be conducted in classrooms, as research conducted in laboratories under carefully controlled conditions can often be directly applied in classrooms.

For teacher training and curricula, the task group recommends instruction in scientific method in evaluating research evidence and comprehensive courses on contemporary cognitive science research on children's learning. Curricula should provide sufficient time on-task to ensure acquisition and long-term retention of both conceptual and procedural knowledge, and should be based on results from contemporary, rigorous, empirical research on learning.

TASK GROUP ON LEARNING PROCESSES: QUESTION AND ANSWER PERIOD

Dr. Whitehurst asked what the three most important things that policy makers or educators should do to translate that research into changes in current practice. Dr. Geary responded that one is that from a basic learning perspective, one cannot separate conceptual and procedural learning. Much of the Math Wars has been based on this false dichotomy that one teaches children concepts or one teaches them procedures, when in fact there is empirical evidence that these areas intersect. We also need to better understand how those intersections occur, particularly for the core content areas in algebra and leading up to algebra.

Dr. Siegler stated that another recommendation would be that programs for improving low-income preschoolers' mathematical knowledge are at the point where scaling up is appropriate because they met criteria such as those of the What Works Clearinghouse.

And finally, the other policy recommendation has to do with fractions, given the absolutely essential quality of understanding fractions for learning algebra and the currently low level of understanding in that area.

Dr. Berch added that another important point would be providing sufficient time on task. And there is a good deal of evidence about this and it has some major implications across the various domains for accurate performance.

Dr. Reyna added that their recommendation about teacher training is also important. The theory is that if there is an increase not only in teacher training on the essentials of children's learning, but also for all personnel throughout education, her theory is that it would create more demand for research at the level of practical implementation in the classroom.

Dr. Ball asked how they would sort out whether what they see in literature is evidence of how kids develop and learn, or whether they see the effects of instruction. Dr. Geary responded that it is a very difficult one to answer because they do not have random assignment to classrooms versus the park for x number of years. The most important point they can make is that the assumption about readiness in stages is now known not to be correct. We know children are capable of learning much more than they have been learning. But how much more, we do not know.

Dr. Siegler stated that Dr. Ball's question is a version of the heredity and environment question that pervades all of psychology and social science. One way that people are addressing it in the area of mathematical development is through cross-national studies.

Dr. Reyna added that there is also the notion of random assignment to separate these influences. Fred Morrison, for example, is doing a series of studies looking at this issue.

Dr. Wu added on the topic of fractions that the problem with evaluating the non-learning of fractions is that it is not so much because children cannot learn fractions, but rather they have been taught so badly in schools. He asked whether it would be possible to run an experiment where one control group is taught the usual way, a separate class of students is taught more correctly, and then their learning achievement would be evaluated. Dr. Reyna stated that it is not only remotely possible, but also they strongly advise that it be done. And it is not even that expensive.

Dr. Fristedt asked about the decreasing interest in mathematics as students increase in grades and whether they have evidence about the cause. He wondered if it was the length of the

book. Dr. Geary stated that it might be that, as well as the factors that students are maturing and going through social peer issues, among other factors.

Dr. Gersten recommended that the Task Group continue to focus on the most important issues. He also stated that it seems critical for the Panel to have an empirical basis for what it is about fractions that is critical for success in algebra. Dr. Geary responded that they were surprised that there were no appropriate data in the longitudinal studies that would allow them to look at early predictors of outcomes in algebra.

Chair Faulkner stated that these presentations would be posted on the website of the Panel, which is at the U.S. Department of Education website.

TASK GROUP AND SUBCOMMITTEE REPORTS: TASK GROUP ON INSTRUCTIONAL PRACTICES

Russell Gersten, Chair; Camilla Benbow; Doug Clements; Bert Fristedt; Tom Loveless; Vern Williams; Irma Arispe; Joan Ferrini-Mundy; Marian Banfield, staff

Dr. Gersten began by describing the very strict standards of evidence used by the task group. Their state of the art statistical techniques limited the number of studies they found. With these types of studies, if there is a pattern of findings in three or more of these high quality studies, they can draw inferences about effective practices. Their list of topics came from issues about which people often write and make recommendations. They also chose some subjects because they seemed to: 1) be topics of importance, 2) have come up in NCTM surveys, and 3) have existing quality experimental research. Twelve high-quality studies met all rigorous standards.

One finding deals with teachers' use of a formative assessment system. These students learn more math on an array of math achievement measures than those of teachers who do not use formative assessment. This finding is consistently replicated. The effect size is 0.2. The second finding deals with enhancements, which are what can happen after formative assessment. This could include tutoring in specific areas or expert math teachers or math coaches coming up with ideas to improve performance. The effect size doubled in this study. The limitations of these studies include that all studies but one were done at the elementary level, most of the enhancements were done with special education students, and all of this research was done with one type of formative assessment.

The next finding was on teacher-directed versus student-centered instruction. When looking at studies that pit one against the other, the only finding that emerged with significant effect sizes was called Team Assisted Individualization, which Bob Slavin and his colleagues developed and studied in the '80s and '90s. This became infused in the Success for All program. It was the only intervention with consistently positive effects in the area of computation although not in conceptual understanding.

The next finding is of a hybrid program with a teacher-directed part, a very explicit instruction part, a way for students to work with each other, and for teachers to use formative assessment. The most important finding was that there is no data from high-quality experiments to support student-centered instruction or direct or teacher-directed instruction, or any other instructional regimen for the average student or the high-ability student.

A review of the research on real-world problems has focused on whether it really helps kids learn math if part of their instruction involves the real-world type of problems that are in many texts. The task group also looked at whether there are better ways to teach students so that they can solve these more complex real-world word problems. For the second question, not enough is known. There are some promising ideas in the paper that are discussed in a very tentative way in the report.

On question one, five high quality studies showed a pooled effect that is significant. The study involved students who were taught with real-world problems, including specific topics involving geometry, fractions, or multi-step problems. However, when studying only typical math achievement, the word problems, or occasionally computation problems involving fractions, there was no longer a significant effect. Only one of the five studies on real-world problem solving reviewed had any negative outcome.

Dr. Gersten then presented the findings concerning gifted students, which consistent showed no known negative impacts. It was also noted that for enrichment, despite the wide advocacy for it, there is virtually no research, save for one study. This study also showed that a mathematically sound enrichment program is likely to include acceleration.

The research on low-achieving students, the lowest third of the population in math, revealed a small number of studies. The Task Group could not look at school-wide reform, because in addition to math, it also had reading, professional development, and other factors involved. They categorized the studies either as explicit instruction or other strategies. Explicit instruction, while they have a sense of what it is, is a construct that is very hard to unpack. Future research in instructional practice is a major goal.

There were five studies that reviewed explicit instruction for low-achieving students, and they had a pooled effect size of 0.97. Most of the studies focused on word problems, not real-world problems. There was an array in these five studies. Some of them used Connected Math concepts where there are clear models. Other models have parts where the teacher models the concepts, but they are much more interactive and probe for misconceptions. They are mixtures with explicit components with other kinds of practices added into them, but both types seem effective.

One thing that was noteworthy, with lower achieving students and children with learning disabilities, is careful sequencing of examples as they described the instructional approaches. Another promising study was done with first graders who were lacking in the basic kindergarten, preschool foundational skills. Teachers' assistants worked with them in a very structured intense way half an hour per week, every other day.

Some of the newer studies relied on cognitive psychology, including how to build quick retrieval of arithmetic facts and combinations both in multiplication and addition through intense multi-faceted instruction. ~~They are~~ The task group found significant results in these studies. Children with math disabilities do not have this retrieval skill, so the studies are nice models and prototypes of practices that could work for this population.

Another promising theme was in practices that made the strategic move from concrete objects to visual representations, and then going back and forth between the two using carefully orchestrated and sequenced methods. Just one study was significant there, in a case where middle school students with learning disabilities learned about algebra.

The Task Group then looked at technology, including calculators, as there is great interest in their effects. There was no evidence of harmful effects, but the studies are limited because they are old. There seems to be some facilitative effects on word problems. The Task Group also carried out a meta-analysis of the meta-analyses of technology use. There is no clear finding there. It depends on the software and goals.

A couple of cross cutting themes include the Task Group's sense that their work does not represent a judgment, or a personal or professional decision as to which studies and findings are included and excluded. They used external objective criteria that were worked out both with the researchers at Abt Associates and the Panel.

Overall, the Task Group has three robust findings. The first is with formative assessment, and there is enough evidence that they can recommend it be used with the caveats. Second, for part of the day, having some type of explicit instruction for the lower third of students has been effective. Third, there is some promise to the serious use of real-world problem solving

on an array of mathematical tasks, but it has not shown the same positive results in terms of traditional achievement.

Dr. Fristedt commented that there seems to be a definitional problem of what real world means. If it is interpreted in the broadest sense, it is good, but when other words that are used like story problems and word problems, he would like to know if they synonyms or are there distinctions. Also related to real-world problems, he wondered should teachers wait to use them until students understand the concepts being taught. The second comment he made was that he was disappointed that the group did not get into details that everyday teachers confront in the classroom about which things should come first. Dr. Gersten replied that the Task Group will ask mathematicians to take a look at the example sequences and the mathematical correctness.

Dr. Clements commented on the concrete to visual learning issue, and recommended that the Panel be very careful about feeding into what is a long standing idea that learning precedes from concrete, to visual, to abstract. A definition of visual meaning needing visual supports is very different from the creation of visual mental structures or visualization skills. The research does not support Buehner's original sequences of absolute steps.

TASK GROUP ON INSTRUCTIONAL PRACTICES: QUESTION AND ANSWER PERIOD

Mr. Williams commented that he has been hearing as a teacher over the last five years that a child will not understand a topic unless it is introduced through a real-world problem. If there is not research to support this, that finding should be made loud and clear.

Dr. Stotsky stated in regards to real-world problems, it would be helpful if there were a clarification that it may be promising, but there is no evidence to support the focus or the emphasis that has been placed on it. Dr. Gersten responded that they tried to be clear that if the goal is raising achievement vs. using the more typical word problems, real-world problems are not going to work. Dr. Stotsky followed up by stating that they said that was for assessment and she is saying in general this has been emphasized for instruction in textbooks.

Dr. Siegler stated that he was concerned about grouping together all computer software and technology under the same heading, as they are all very different. He does not feel that they should say they do not have enough research to meet this criterion. Maybe it is the wrong criterion. Dr. Clements responded that with the time constraints of the presentation, they were not able to go into the various types of software, and then within those various types, varying substantiations of those types.

Dr. Berch asked about the need to unpack the word explicit, which he feels needs to be done to make clearer distinctions between direct instruction, explicit instruction, guided inquiry, et cetera. It seems to him they can make distinctions between more of the extremes and make a clear recommendation. Dr. Loveless stated that they defined direct instruction as teacher-directed instruction. Direct Instruction, which is the Engelmann/Carnine model, is a scripted model form of instruction. Teacher-directed instruction refers to a larger pool of interventions where the teacher is the center of instruction.

The task group states that they do have more research that looks at different forms of teacher-directed instruction, and they applied these rigorous qualifications to the research to screen out what they could not rely on. The research typically includes the teacher-directed group as the control group. It's also called the traditional instruction group, and in those cases, it was not explicitly described. They know that teacher-directed instruction is very popular in classrooms, but they do not know very much about what kinds of teacher-directed instruction are effective, what kinds are ineffective, and even how many different kinds there are. The task group's main recommendation is that they have more experimental studies that look at different kinds of teacher-directed instruction, so we can learn more about it.

Dr. Ma commented that the teacher survey about student regimens for algebra mentioned word problems, and she wondered if there is any evidence of a distinction between real-world problems and word problems. Dr. Gersten stated that the analysis does not look at algebra, but there is no significant impact on the more typical word problems from these real world experiences.

Dr. Fennell asked if there was a way to acknowledge that the ability to solve problems is important for anybody learning mathematics at any level, without worrying how real-world is defined. Dr. Gersten stated that most of the studies on real-world problems define word problems such as those that appear in the various state assessments and are multi-stepped problems. That is something they are going to continue to probe.

Mr. Williams stated that real-world problems at times really do not concentrate purely on the mathematics. They might be used for motivational purposes and get into other topics that might not be related to the mathematics involved. When teachers generally discuss word problems, they are thinking more of typical algebraic problems involving distance, rate, time, et cetera.

Dr. Fennell stated that it is also important to think about whose world it is—the world of the child or the teacher.

TASK GROUP AND SUBCOMMITTEE REPORTS: TASK GROUP ON TEACHERS

Deborah Loewenberg Ball, Chair; James H. Simons; Hung-Hsi Wu; Ray Simon; Grover J. (Russ) Whitehurst; Jim Yun.

Dr. Ball presented the synthesis of the work that this Task Group has been doing on teachers. As they work across the questions they have been investigating, the Task Group starts with the assertion, from research and other sources, that teachers are crucial and they make an enormous difference. The group reviewed studies that examined the contribution made by the teacher to achievement gains of students. These showed that a large portion of the variability in student achievement gains is due to who the teacher is. Yet, they have learned very little from these studies about exactly what it is about these teachers who are making a difference for students, whether it is what they know or do specifically. So, what they want to know next is what it is that these teachers do, what it is they know, or something about them that would help explain why some teachers make greater gains for kids than others.

Linking the findings of the Learning Processes and Instructional Practices Task Groups helps them understand what it is that teachers are doing when they make greater differences for students than others.

Working from the strong hypotheses about teachers' mathematical knowledge, the Task Group is led to an investigation of the evidence on the relationship of teachers' content knowledge to students' achievement. Additionally, they are looking at how states' teachers' assessments can rigorously measure this kind of content knowledge.

The Task Group is also looking at what is known about how to train, recruit, retain, and reward teachers in a way that supports more teachers who can produce consistent achievement gains in students.

In addition, the Task Group is addressing the scale problem, and the need to prepare large numbers of teachers who know enough and are skillful enough to produce achievement gains in students. This task includes a review of the specialization of elementary or middle school teachers that might help to address this enormous need.

In regards to methodology, the Task Group tried to identify the available scientific evidence to address their questions. They searched different databases and looked manually based on recommendations from people, testimony, and other areas that met their criteria. They

organized the evidence into categories based on study strength using the criteria developed as a Panel. They noticed that the strengths of the available evidence varied quite a bit across their four questions.

For each of the questions, they organized the findings into 1) what they think they know based on the criteria, 2) what they do not know, and, 3) what is not supported by the research. They also had cases of consistent non-effects, and the task group proposes that the Panel consider those as things that they ought to stop saying. In these cases, there is no substantial evidence to continue to claim that certain things are true.

The task group's findings on teacher content knowledge show that teachers' mathematical knowledge is a positive factor in student achievement. Proximal measures, or tests of the relevant knowledge that teachers actually use to teach mathematics, show a stronger signal than more distal indicators like certification status.

In the category of what the Task Group does not know is what teachers need to know to teach particular topics to particular students. That does not show up in the research they reviewed. They also did not find evidence about how teachers' mathematical knowledge affects instruction in student achievement.

In the "not supported by research" category is the belief that elementary teachers who take more university math courses are more effective. They do see, however, some signs of this for secondary school teachers. They also do not see in the research that students who are taught by teachers who are certified or licensed in math consistently learn more than those taught by teachers who are not.

The next area of review included the area of teacher education; which is defined to include teacher preparation, both conventional and alternative pathways; induction programs; and professional development. The Task Group members do not have anything they can claim to know about teacher education from the research reviewed. Specifically, they do not know what features of teacher preparation or professional development produce changes in teachers' knowledge or in their students' learning. In the "not supported by research" category is the belief that different pathways into teaching at entry produce differential effects in teachers' effectiveness. Very few studies met the criteria that allowed the task group to make claims of what produces what kinds of changes for teachers' capacity.

In the area of teacher incentives or what might attract people into the profession, attract them to certain locations where we especially need teachers, or might reward them for producing achievement, the research says that the salary differential between teaching and other technical fields is quite large. It is not large at entry, but it increases dramatically across the first 10 years.

The Task Group also finds that the exit rate of math and science teachers is greater than other teachers, and that teachers are more likely to cite dissatisfaction with salaries as one of several reasons for leaving the profession.

In addition, location-based pay, which is used to refer to pay to attract teachers to high need areas, can *retain* experienced teachers. But they do not know whether and how location based pay helps to *attract* teachers to high need areas. Performance pay, bonuses for student achievement gains, for teachers also can enhance students' achievement. Little information is available, though, as to how to best design these sorts of pay schemes effectively.

In the area of math specialists, a term used for at least three different models of types of specialization at the elementary level, the Task Group observes that this term is actually obscuring a conversation about what the idea is really about. What they do not know is whether a math specialist in any of these models leads to greater gains in student achievement. What is not supported by research is that most high-performing countries use math specialists at the elementary level.

The task group then presented its recommendations. They recommend that teachers should be required at least to know the mathematics they are teaching, and that certification and licensure examines should at least test well the content that teachers actually teach. They also

think that alternative pathways into teaching need to be developed and studied. In addition, they think it is worth researching and pursuing alternative salary schemes, including differential pay for teachers of mathematics, and pay based on location and performance.

The task group also believes that where there is a shortage of elementary school math teachers who have appropriate knowledge of mathematics for teaching, math specialists could help to address this need. But there needs to be a clarification of terms. So, when this Task Group is using the word 'math specialist,' the members mean teachers who have the requisite knowledge that is needed to teach mathematics and are responsible for teaching the bulk of mathematics in an elementary school. They do not mean pull out programs or math coaches. One thing the task group was lacking was evidence that could substantiate the lack or presence of mathematical knowledge among elementary school teachers. It is widely believed that elementary school teachers lack requisite mathematical knowledge. However, the studies that will allow the group to say that generally do not exist. They have phrased this carefully to say, in areas where it is clear that the shortage of such teachers exist; this could be a useful strategy.

The Task Group then presented recommendations for research. One is that it is quite clear that further research needs to be done to elaborate what mathematics teachers really do need to know to teach specific topics to specific students, particularly beyond what is in the curriculum.

The Task Group recommends the creation of better measures of teacher's mathematical knowledge that focus more squarely on what teachers actually use when they are teaching, instead of distal indicators such as certification or courses taken. They also recommend new research that identifies the specific features of teacher education, pre-service induction, and professional development that actually have an impact on teachers' effectiveness. These studies need to probe what it is that distinguishes the teachers who are making a difference with students. They also recommend studies of the effectiveness of elementary math specialists.

TASK GROUP ON TEACHERS: QUESTION AND ANSWER PERIOD

Dr. Loveless noted that the TIMSS data shows internationally that most algebra teachers have a bachelor's degree in mathematics. In the United States, most teachers of algebra have a degree in math education, which is quite different. He asked if there was research on the relative effectiveness of those two degrees. Dr. Gersten responded that the Schools and Staffing Survey, NAEP and other sources collect information on that, so they know the proportion of teachers who have degrees in mathematics or math education. However, to the best of their knowledge, there are no studies that examine the impact of those differences on students. It could be that the new high school longitudinal study, which will start with ninth-graders, will focus on mathematics or allow an opportunity to examine that more carefully.

Dr. Clements asked if the task group could say something about the differentiation between content knowledge and pedagogical content knowledge. Dr. Ball replied that it depends how far he would go into what is called pedagogical content knowledge. The task group was not investigating all the knowledge that teachers need to teach, but were focused on content knowledge. On the question of testing, they have some studies that look at knowledge used by teachers and they looked at that work.

Dr. Reyna mentioned that they could think of this in terms of a path analysis or cause analysis. For example, they can entertain the hypothesis that university courses in mathematics compared to not taking those courses in mathematics leads to higher levels of relevant knowledge. And relevant knowledge in turn affects student achievement. This could be a path for future research. Dr. Ball asked if Dr. Reyna had any evidence on effective math courses and relevant knowledge. The question of what is relevant knowledge is one of the concerns in the field that has been only in a very limited way unpacked. What Dr. Reyna is suggesting would

require answers to questions where there is a lack of literature. Dr. Reyna added that labor economists study incentives in a variety of fields outside of teaching, and there are some generalizations across studies of labor economics in terms of the effects of incentives and pay on a variety of work choices and occupations. She asked if any of those studies are relevant to teachers. Dr. Whitehurst stated that they cited some of that work. It is clear, for example, that salary differentials in different professions have substantially larger impact at the point of career choice than they do after that choice has been made. Salary differentials make less of a difference when someone is committed to the teaching profession, and interestingly more for males than for females.

Dr. Stotsky asked whether there was a body of research about student teaching as part of teacher preparation. Dr. Ball replied that they were not able to uncover research that showed differential effects of particular features of pre-service teacher education.

Dr. Berch asked for clarification on the distinction between the categories of what they do not know and what is not supported by research. Dr. Ball replied that the distinction they were experimenting with is between what they do not know, where there just is not research, but both could fall into the, “what is not supported by research” category. Dr. Berch said that it was still confusing. But it may be if there is mixed evidence, if it is not supported, it may mean that the evidence is inconsistent. In other cases, it is neither supported nor refuted. Dr. Ball stated that the task group was experimenting with these categories and they will take another look at them.

Dr. Wu asked if what Dr. Berch was saying is that when they say something is not supported by research, do they want it to mean it is refuted by research. Dr. Berch replied that he was wondering if that is what the implication is and it does not seem like that is consistently the case there. Dr. Wu replied that it is not their implication. Chair Faulkner stated that he takes it as research exists, but does not confirm the statement that is about to follow. Dr. Loveless asked that if they find something that is refuted by research and those findings are consistent, why they wouldn't put that under the “things we know” category. Chair Faulkner stated that they will talk about this further at another time.

TASK GROUP AND SUBCOMMITTEE REPORTS: TASK GROUP ON ASSESSMENT

Camilla Benbow, Chair; Susan Embretson; Francis “Skip” Fennell; Bert Fristedt; Tom Loveless; Wilfried Schmid; Sandra Stotsky; Irma Arispe.

Dr. Benbow stated that assessment is used in a variety of ways, to shape the content and format of instruction, to adjust educational experiences to meet the needs of individual students, for selection, and for evaluating student and school performance. Their review of the research focused on evaluating student and school performance, due partly to the impact of No Child Left Behind (NCLB) and its focus on using tests to hold students and schools accountable for performance through NAEP and state tests. State tests are designed to determine student proficiency in certain areas and all schools are required as part of NCLB to also participate in NAEP.

Because state tests and NAEP are such high-stakes tests and they are used in determining, for example, whether a school makes Adequate Yearly Progress and whether there could be consequences to schools as a result of their performance, the Task Group decided to take a hard look at these tests. They asked questions about these tests, such as are they appropriate; do they measure what is intended; are they biased; are their conclusions justified; what is their measurement quality; and do they measure what is deemed important for children to master.

Their methodology is quite different than the other task groups. They looked at the main NAEP test for the fourth and eighth grades, and the tests of states that were representative of the testing practices in this nation. They also drew heavily from the NAEP Validity Study report of

the mathematics assessment in Grades 4 and 8 as a reference point. They also reviewed the National Center for Educational Statistics' response to the validity study. In addition, the Task Group conducted their own search of the literature with the help of Abt Associates, and with the help of IDA/STPI, collected technical information from each of the state's websites in Grades 3 to 8 on framework, procedures, and release items. Finally, they carried out a case study analysis of released items in Grades 4 and 8 for NAEP and the state tests.

The Task Group has two main recommendations. The first is that NAEP and state tests must focus on the mathematics that students should learn. The Conceptual Knowledge and Skills Task Group has talked about what that knowledge should be. They believe that skills and assessments should be aligned, with scores on this critical content reported and tracked over time.

The second recommendation is that states and NAEP need to develop better quality control and oversight procedures to ensure that test items are of the highest quality, measure what is intended, and that non-content relevant sources of variance in performance is minimized.

Dr. Benbow then turned the presentation over to Dr. Loveless, who stated that they know at some point over the next 5 or 6 years, the NAEP framework in mathematics will be revised and a new framework will be adopted. There was an initial framework adopted in 1990 and it has been revised twice. What the Task Group wants to do is offer some principles for revision and reorganization of NAEP. He then presented preliminary principles that are still being debated.

The first principle is to disaggregate numbers in two separate areas. There are five strands in NAEP and the first strand is called "number" or "number sense." The first area would be looking at wholes and integers, and then, looking at fractions, decimals and percents. This would separate the assessment of those two areas as clusters of skills and knowledge in mathematics. The rationale for this is the Critical Foundations as laid out by the Conceptual Knowledge and Skills Task Group. Another rationale is that fractions are currently under-represented on NAEP. The final rationale for this is that NAEP scores are reported for strands so that national progress can be monitored in different areas. Currently we monitor national progress at the eighth grade level in the strand called number. They think it would be better to monitor progress at fourth grade with whole numbers and at eighth grade on integers.

The second principle is to combine measurement and geometry. This would make the NAEP consistent with the 12th-grade NAEP. The 4th- and 8th-grade NAEPs currently have these as two separate strands. By combining them, the Task Group believes the complexity of measurement items can be increased. Currently the measurement items are judged to be low in rigor.

The third principle addresses algebra. Reviewers of NAEP, especially mathematicians, have found problems with patterns. Many believe the pattern items are over represented on NAEP and that they are of poor quality. The Task Group recommends that NAEP make them more mathematical when they are used. There is also the question of whether kindergarten through fourth grade "algebra" is really algebra. The rationale again is the definition of algebra by the Panel's Conceptual Knowledge and Skills Task Group.

Dr. Benbow added that state assessments are heavily influenced by the NAEP, so when they make changes or recommend principles for the revision of the NAEP, indirectly they are also making recommendations for state tests.

Dr. Benbow then presented the second part of the presentation that covers quality control issues. They asked what do these tests measure and how well do they measure whatever it is that they are measuring. One of the issues that concerned the Task Group was the issue of non-construct relevant variance and contamination, which can come from verbiage that is unnecessary, excessive, or unfamiliar; or from confusing visual displays. Contamination and non-construct relevant variance means that test scores may be determined by things other than the mathematical skills they are supposed to measure. Specifically, the Task Group reviewed excessive verbiage as this can attenuate the performance of some groups. They did a case study analysis of state test items, and they did find many instances of test items with problems of this

type. They also reviewed situated mathematics problems (real-world or word problems) due to their concern with excessive verbiage.

The Task Group recommends that if tests use real-world or word problems, they should meet the following conditions. They need to be concrete and serve to clarify mathematical relationships in the problem. They need to draw on math that has been taught. The items need to be well written. There need to be enough items and depth to address the entire range of student ability. In addition, scientific and logical evidence, as well as content expertise, need to guide the test design. The members also felt that item content should be carefully examined to understand performance.

Dr. Benbow addressed the need for communication by those who develop frameworks on what is intended by test items. The group was disappointed by the lack of high-level research on the design of mathematics items. The Task Group also made recommendations about quality control.

They reviewed proficiency standards and how states and the NAEP set them. There are several ways this is done, and the Task Group recommend that the methods follow the best scientific practice. Their review of the literature determined that the modified Angoff method probably has the most support. NAEP should also draw on international data on student performance to help in that process.

Another recommendation the Task Group is considering is whether NAEP should conduct a special study of algebra involving students who have completed or are about to complete one or more courses in formal algebra, as defined by the Conceptual Knowledge and Skills' Major Topics of School Algebra.

The Task Group will also review the research on calculators and their role in assessment. In addition, they will look at different item types, for example, multiple choice versus constructed response. They have found already that calculators in the early grades were not used very frequently.

Dr. Reyna asked if they are looking at preparation for Algebra and what an appropriate assessment would be. She stated that it would be useful given the current discussions to have an appropriate instrument that would assess, particularly in a diagnostic way, adequate preparation for algebra. Dr. Fennell replied that there are at least two standardized tests that do that now and the group should compare them to the Major Topics of School Algebra.

Dr. Fristedt stated that some of the things that have been said are actually related to assessing preparation for algebra. There is some concern that NAEP fails to test the things that are identified as critical for algebra.

Chair Faulkner thanked the public for attending, and announced that the next National Math Panel meeting will be hosted by Arizona State University in Phoenix, Arizona, October 23rd and 24th. He adjourned the meeting at 12:25 p.m.

I certify the accuracy of these minutes.

Chair Signature _____ Date _____

Vice Chair Signature _____ Date _____

ADDENDUM: PUBLIC PARTICIPANTS

First Name	Last Name	Organization
Rick	Armstrong	N/A
Hyman	Bass	University of Michigan
Jack	Beers	Metropolitan Teaching and Learning
Anilkumar	Belvadi	SLPS
Julie	Bierach	KWMU
Jarrold	Brown	SLPS
Jennifer	Chintala	Pearson
Lori	Craig	SLPS
Cathie	Dillender	Pearson
Matt	Frueh	McDougal Littell
Joan	Gallagher	Parent
Lisa	Gill	Texas Instruments
Jeanne	Golden	Clayton Math Matters
Anita	Greenbey	Compass Learning
Laura	Grizzle	SLPS
Heather	Gunsallus	Words & Numbers, Inc.
Pamela	Hamped	PD
Mischelle	Handley	America's Choice / NCEE
Ingrid	Iskali	SLPS
Audrey	Jackson	National Council of Teachers of Mathematics
David	Kilper	Washington University, Photo Services
Elizabeth	Lehnerlz	Pearson
Andrea	Maddox-Dallas	Clayton Math Matters
Judy	Martin	Washington University
Michael	McCrary	SLPS
Lynda	McKelvey	Sopris West Educational Services
Chris	Morrison	CTB / McGraw-Hill
Karessa	Morrow	SLPS
Robin	O'Callaghan	The College Board
Leigh	Palmer	Houghton Mifflin Company
Pamela	Powell	Uteach, The University Texas at Austin
Debra	Powell-Childress	SLPS
Rick	Rickhoff	SRA McGraw-Hill
Diane	Schaeter	ASSM Rhode Island Dept. of Ed
Ines	Segert	University of Missouri
David	Whiting	McGraw-Hill
Christine	Willis	Cambium Learning, Inc
Evelyn	Woods	SLPS
Mark	Wrighton	Washington University
Susan	Zareh	Forsyth School

