U.S. DEPARTMENT OF EDUCATION

* * * * *

NATIONAL MATH PANEL MEETING

* * * * *

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 MEMBERS:
DR. LARRY FAULKNER, CHAIR
DR. CAMILLA BENBOW, VICE CHAIR
DR. DEBORAH LOEWENBERG BALL
DR. A. WADE BOYKIN (NOT PRESENT)
DR. DOUGLAS CLEMENTS (NOT PRESENT)
DR. SUSAN EMBRETSON
DR. FRANCIS (SKIP) FENNELL
DR. BERT FRISTEDT
DR. DAVID GEARY
DR. RUSSELL GERSTEN
DR. TOM LOVELESS
DR. LIPING MA (NOT PRESENT)
DR. VALERIE REYNA
DR. WILFRIED SCHMID (NOT PRESENT)
DR. ROBERT SIEGLER
DR. JAMES SIMONS (NOT PRESENT)
DR. SANDRA STOTSKY
MR. VERN WILLIAMS
DR. HUNG-HSI WU

EX OFFICIO MEMBERS:
DR. IRMA ARISPE
DR. DANIEL BERCH
DR. JOAN FERRINI-MUNDY
MR. RAYMOND SIMON
DR. GROVER (RUSS) WHITEHURST

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

C-O-N-T-E-N-T-S

P R O C E E D I N G S

(8:30 a.m.)

I. OPEN SESSION - PROGRESS REPORTS TO PANEL

DR. FAULKNER: (Presiding). Okay. Let me convene this session of the National Mathematics Advisory Panel. Welcome to members of the public who are with us today. I am Larry Faulkner, Chair of the Panel. This is Camilla Benbow who is Vice Chair of the Panel. And we are principally receiving today, reports from the subcommittees and task groups of the Panel that have been carrying out a great volume of work outside of our public sessions and each of these bodies will be coming back in today to make a series of reports.

We are in the process of wrapping up the work of the task groups. The task groups have been assigned in particular areas of inquiry relative to our charge and those task groups will be making their main reports, by and large, today. There is one task group, Assessment that was started later in the process and is in the middle of its work. They will be giving an interim report.

The Panel will go from this stage

of receiving task group reports into a stage in its next meeting in Phoenix, Arizona, that will be largely focused on synthesizing a Panel Report, an over-arching report for the whole Panel’s message to the constituencies interested in this report. I wanted to highlight for the audience that we are at a major shift in our activity and are about to move out of a subcommittee-based activity, into a whole Panel activity.

Now, let me also ask if signing services are needed? Signing services are being provided right now and we will be happy to continue them if they are being used. If they are not being used we will discontinue it, with the understanding that it can be re-instituted if the need arises. Is there a need for us to continue the signing services?

[No Verbal Response]

DR. FAULKNER: If not, then we will discontinue them. And again, they are available if it becomes necessary.

The chairs and subcommittees and task groups will deliver their reports to the Panel from the testimony table in front. We will begin with a Subcommittee on

Instructional Materials. This is a group that has just been appointed in about the last month. And its job is to examine what can be said on the basis of strong scientific evidence regarding the effectiveness of instructional materials. Bob Siegler is chairing it. So, let me ask if the Instructional Materials Subcommittee will move forward and make its report.

There are, I guess some other comments that I might make about this group while they are taking their place. The Executive Order calls for the Panel to make recommendations based on the best available scientific evidence on instructional materials that are effective for improving mathematics learning. Originally this topic was included in the Instructional Practices Task Group that will be reporting later. However, because of some of the Panel members' professional involvement in this area, a separate subcommittee was constituted and officially cleared of any appearance of conflict of interest in order to address the Instructional Materials part of our charge. And that is one of the reasons why they have just begun their work. Bob Siegler is chair. Bob would you introduce the panel members.

II. INSTRUCTIONAL MATERIALS SUBCOMMITTEE

DR. SIEGLER: Yes. Sitting to my extreme right is Irma Arispe, Vern Williams next to her, Dan Berch next to him, and on my left, Bert Fristedt. And this is the Instructional Materials Subcommittee.

As Larry mentioned we have just started our work. We are not as far along as any of the groups that you will be hearing from, but we have accomplished a few things.

We are going to be looking at a variety of sources of evidence, much of it from the reports of the other National Math Panel groups. We also have other available materials, such as the National Opinion Research Center (NORC) Report of algebra teachers, the survey that you heard about yesterday, the NRC report on instructional materials, some mathematicians who have written about evaluating textbooks for accuracy and a variety of other sources.

We have decided to divide our task into three main parts. One is evaluating textbooks, one important kind of instructional material. Another is evaluating ancillary materials, and a third is evaluating knowledge creation mechanisms. And I will explain what each of those are and what topics we are going to be looking at.

So, on textbooks we are going to be looking at two main things. One of them is the mathematical accuracy of textbooks. The other is a cluster of concepts and dimensions that have led to the situation that we have heard about in Cambridge, Massachusetts, I believe, where textbook manufacturers were telling us that the average third grade math textbook is 750 pages and the average eighth grade math textbook is over 1100 pages, and there are a variety of reasons for that. We will be going into that and comparing these to textbooks in other countries to see if we really need to have textbooks that are this long. The length of the textbooks and the variety of topics that are involved, get into issues of coherence and sequencing. There are a variety of reasons why the textbooks are so long, and we have heard about many of them. We are going to be talking about whether the sheer length and diversity of topics interferes with the coherence and logical sequencing of textbooks.

The second kind of instructional material we are going to be talking about has to do with ancillary materials, materials other than textbooks that are used in instruction. Here we are going to be looking at calculators, computer software, teacher manuals and support for diverse students, including students of very low ability, and also, students of very high ability. We want to see what kinds of things are available for supporting these students. Finally, we are going to be looking a little bit at knowledge creation mechanisms and a little bit on the What Works Clearinghouse. We are also trying to identify areas that are particularly in need of greater research.

Now, this is all going to be very brief. We are charged with writing five to eight paragraphs, and we are looking at something in that range. So, we are only going to be able to touch very briefly on each of these, both due to considerations of length and considerations of time. Our calendar is that we start now, having identified the topics at this meeting. We are supposed to write something on each of these. Different people will be drafting different paragraphs by a week from today, when we will talk on the phone. A week after that we are supposed to have the report in and I wish us all good luck. That is all that I have to say. Would other panel members like to add anything?

DR. FRISTEDT: Bert Fristedt. My own inclination is to think primarily towards the future. So, on some of these early things where Bob mentioned that we are going to evaluate, that covers existing things. But we are not going to give such a detailed report that someone in the audience can come and say the National Math Panel says that this is a good resource and this is a bad one.

On the other hand, I think we can use the past to make suggestions. Maybe suggestion is exactly the right word for the future. For example, if we really do think that there is a preponderance of books that are too long, I think it is important for us to come up with suggestions for how they could be shorter without losing essential things.

DR. SIEGLER: Any other comments?

[No Verbal Response]

DR. FAULKNER: Thank you, Bob. Let me add something for the benefit of the audience, about the length of your report. Your group has been charged with developing language that might be effectively incorporated into the Panel Report. The Panel Report will be much shorter than the task group reports. The Panel Report, as a whole, is being targeted for something in the range of thirty published pages, which would have to cover of course, the activities of all the task groups and subcommittees, and deal with introductory material and so forth. So, there is a limited volume of space or limited space in the Final Report for any topic. They have been charged with going straight to the nominated language for inclusion in that report, rather than trying to develop a detailed study in this area.

There are various reasons why detailed studies I think, are very difficult for this Panel to carry out. I think we have limited the scope of what we are going to do to match up to what is possible. Do you want to add anything to that Bob?

DR. SIEGLER: No. No, that is exactly my impression too. Panel members any questions or comments, advice, for this set of colleagues?

DR. WHITEHURST: I am curious. Are we going to allude at all to the National Academy of Science's Report on curriculum and textbooks that came out several years ago?

DR. SIEGLER: Several of us read that and if it is what I am thinking of, it has a blue cover and is paperback? I do not know how recently you looked at it, but I have trouble keeping all these different reports straight to tell you the truth. We found it only a little bit helpful, actually. I frankly was disappointed in what I found there, as far as being able to help our group.

DR. FAULKNER: Russell, you look like you were going to say something else.

DR. WHITEHURST: No.

DR. FAULKNER: Deborah.

DR. BALL: There have been repeated references to the length of the textbooks, and I wondered if you were going to try to be analytic about what the sources of the length were? For example, one thing that our group has discussed is the potential of a new generation of textbooks that actually supported the range of capacities that teachers need to teach well. We will not be recommending anything about this, but it could be an intersection. Those would be found in the support materials. Or it could be length as in what students are expected to complete. Or there could be other sources of length, and there could be other things besides length in terms of usability. I am just curious as to what you mean by length and how you are thinking you might address this?

DR. SIEGLER: The variety of issues that the textbook manufacturers themselves brought up in Cambridge that we think are strong candidates for removal or minimal coverage, are extensive use of large color photographs, for example, that have little to do with the content that is being captured. There are materials that are required in some states, but not in the state where it is being used. For example, one of the things that I think we will discuss is whether given the current publishing capabilities could textbooks be created that just had the chapters that are going to be used commonly in a given state, for example. There are a number of states that have unified adoptions, in addition to the three very large ones that do have state specific editions. The textbook manufacturer said those books are 25 percent shorter in eighth grade or they are two hundred some pages shorter. So that this is a very large issue. Because when you have chapters that are not being covered that are strewn throughout the book, it has to interfere with the coherence and sequencing of presentation, because you cannot say in the last chapter we read (X), when you have no idea what the last chapter that student read was.

DR. BALL: Let me pursue the question of teacher support materials. Will you be examining and analyzing the quality, nature and content of the support guidance and so on, provided for the teachers?

DR. SIEGLER: Bert has been particularly interested in this issue and perhaps you should reply.

DR. FRISTEDT: My feeling is that the materials that the teachers get should have the following in mind. If the publisher thinks that this is an area where the teacher's knowledge might be somewhat shy, then they could focus on helping the teacher recall and get back to that particular aspect of mathematics. And so that would have to be there.

On the other hand, there are many things that are in some books for the teachers that really do not need to be there. They are anecdotal little extra comments, or some interpretations that one might make in a field outside of mathematics, such as things that fill up the margins where someone had an idea that it would be good for the teacher to say. That is not a good enough criterion for it to make it into the book in my way of thinking, but we will have to look at that in more detail.

But I think helping the teacher with the mathematics itself is the primary goal of the supplementary material for the teacher. Whether that should be in a separate little booklet or whether it should be incorporated in a teacher's edition, we don't know.

DR. SIEGLER: Just one additional comment to Deborah and to any of the rest of you. If you know of any articles on this topic that we should look at, please recommend them to us, because we can use all the help we can get. Skip.

DR. FENNELL: Skip Fennell. I think the challenge you have is trying to do what I have just heard in eight paragraphs or whatever that number was, because clearly the issue of curricular coherence has something to do with length, however you define that. And then you have this intersect between the mathematics and the pedagogy, and frankly the marketability of a program that gets into some of what Bert just suggested with regard to, if I can use the phrase “The fattening of the teacher materials.” So, this is going to be tricky for you, and I just want to be on record to say good luck trying to capture that in a couple of pages.

DR. SIEGLER: Tom.

DR. LOVELESS: Tom Loveless. Are you going to be looking at the assessment materials that come with textbooks? I know they vary a great deal in terms of the numbers of some books that have lots of quizzes and unit tests that come with it, others have very few.

DR. SIEGLER: I am certainly open to the idea. I do not know where the time is going to come from. Again, if there is something really good that is published that you could recommend and that we could look at and get something quickly, I think it is a very legitimate and important topic. But I am just a little daunted by the magnitude of the task relative to the time.

DR. WU: Hung-Hsi Wu. I wanted to just add a remark on the issue of length. And I certainly concur with the subcommittee's concern over length, and I would like to divert slightly from Skip about the need of length on account of coherence. The most coherent textbooks that I have seen are extremely thin and you can buy them from the American Mathematical Society. The Japanese textbooks of grades 10-11 are thin, to the point, and very coherent. I think the length is mainly, in my amateurish opinion, a function of marketability, commercial considerations, and to some extent, the level of the teachers who are going to use them.

There are various accounts from representatives, from publishers, that what is in there is because they found that those elements, the glossy pictures, the layout and so on, were those things demanded by teachers and therefore, they wanted to cater to that particular wish. So, I think that maybe that is an element. I do not know whether you have considered that, but I just thought that I would bring it up.

DR. SIEGLER: I am sure that you are right about the market factors that go into it. It is not the total reason. I think the fact that different states require different topics to be taught in different grades and cluster issues around that also goes into the picture. Certainly with things like these color photographs and inspirational stories about people who overcame obstacles to learn mathematics are probably market driven. I still think that it is important to bring them up as concerns.

DR. FAULKNER: Okay, anything else? Bert.

DR. FRISTEDT: One advantage of

what is now in the near future, in shifting to discuss the whole report, is that some things that are in different places can be brought together in a unified way. For example, there is our group talking about materials. There is already an assessment group, and we were just asked by Tom about the assessment materials that go along with the textbook. Where is the report by Russell on formative assessment? Are these materials part of formative assessment or are they just used for evaluating students at the end? So, the chairs of the three groups that are going to focus on the Panel Report as a whole, they have a chore too of bringing together these various things that our subgroups and some of the task groups are dealing with. So, I just wanted to say that we are not the only one that has a tough job ahead of us.

DR. FAULKNER: Sandy.

DR. STOTSKY: Just a quick question. I was not sure if I heard you mention cost in any way? And I know that I do not expect you to do great research on the cost of textbooks, but is there any possibility that you could have some sense of trends in terms of how the cost of school textbooks have changed over say thirty to forty years with respect to the increasing length and the photographs and whatever else, so we have a sense of how this is affecting school budgets?

DR. SIEGLER: With time permitting.

DR. STOTSKY: I understand.

DR. BALL: I just have a procedural question. At what meeting will we all get to discuss exactly what we are going to be concluding about textbooks? Would that be at our next meeting?

DR. FAULKNER: Basically what they are producing is an analog to the working papers that will be coming out of the Task Group Reports. So, it will be working material for the synthesis groups to start putting into a Panel Report.

DR. BALL: So, the place in which the panel will consider what the end will really say about the instructional materials will be at the level of the Panel Report?

DR. FAULKNER: Not at the Panel Report, at the synthesis that will go on in

Phoenix.

DR. BALL: Right, right. Okay. Thank you.

DR. FAULKNER: The Vice Chair just commented that there are conflicts of interest in some cases that limit what people can say. And that is true, we will have to worry about the management of those conflicts. That is a major issue that has shaped the way we are having to go about addressing this topic and we will have to be cognizant of those even as we go forward. That is why we are not commenting on individual products. They are going to comment on the state of knowledge, really in some general way, but they are not going to be able to speak about individual products and that is because of various professional relationships that Panel members hold here. That is all going to have to be managed and it is being addressed quite carefully within the U.S. Department of Education and its oversight staff. That is probably the best I can say about that right now. But, to clarify Deborah’s question, I am going to repeat my comments.

We are in a process where each of

the subcommittees and task groups is producing a body of most important material to be sent into a synthesis process, where the Panel as a whole will be putting together their report. What they are doing is producing an analog to a working paper, all right? Not an analog to a Task Group Report. Okay. I think that is where we are, and I appreciate your coming up and telling us where you are.

There is also in this Panel, a Subcommittee on Standards of Evidence that has been working on that topic of standards of evidence for a year. It has been a highly collaborative enterprise that has involved the whole Panel at one point or another, through email and other kinds of communication. Initial guidelines were drafted and each of the task groups have developed additional criteria for their particular work. After using the guidelines in the review of the literature, basically test flying these guidelines, this subcommittee is now ready to present this document and after discussion, move for adoption.

The chair of the subcommittee is Valerie Reyna. Valerie, please take your place and make the presentation.

III. STANDARDS OF EVIDENCE SUBCOMMITTEE

DR. REYNA: Thank you, Mr. Chairman. Good morning members of the Panel, staff, members of the public. I want to thank the subcommittee to begin with, who worked very hard with me on this report, Camilla Benbow, Wade Boykin, who could not be here due to health reasons, and Russ Whitehurst. And also, our very special thanks are due to Mark Lipsy, who was invaluable to this effort and to the staff.

As the Chairman already mentioned, our task was to marshal the best scientific evidence in the service of producing an evidence base in mathematics instruction. This of course, leads to the inevitable question, what is the best scientific evidence? This is a challenging and difficult task and the document we have produced is not many hundreds of pages, although easily it could be. In this presentation, I am going to go over the highlights, but of course, naturally I will be willing to answer any questions that people may have.

Just to take a step back and give a

quick overview of the standards of evidence, we 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 many different diverse populations in different circumstances.

We also distinguished therefore, highest quality evidence, which is high in both internal and external validity from promising or suggested evidence.

One of the charges of this panel was not only to identify the very best evidence that could be marshaled in the service of the nation’s students, but also to identify areas that would benefit from further research, further development, scaling up, and the like.

So, in this category we were interested in studies for which there really would be some evidence of effectiveness, but that evidence was limited by some methodological shortcoming, lack of diversity of samples, and that sort of thing.

The third broad category really is opinion, and this is a catch-all term, it includes expert opinion. For example, questions such as what is the nature of Algebra are really a question of expert opinion rather than scientific evidence. So, this includes an assortment of things that really are not matters for which we have strong or suggestive evidence.

So, just to drill down a little bit into these categories. Again, we are still at the overview level. In our report we distinguished different kinds of questions; and this is very important, questions that involve survey methodology. For example, our subject of different kinds of methodological criteria that are experiments; and we differentiate that somewhat, but right now, again, we are still at the overview level.

So, our strongest confidence was reserved for studies that actually test a hypothesis. These are the kinds of studies where in fact you can disprove the opinion or belief that you started out with. These are very important and dis-confirmation, as we know, is a hallmark of science. Naturally we are also interested in studies that meet the highest methodological standards, as I have mentioned, and that have been replicated with diverse samples, again internal and external validity.

Also, it is not only the quality of the design that matters, but it is the balance in quantity of evidence in addition. So, we had to integrate the concepts of quality of evidence with quantity and balance of evidence. And here are some guidelines for how we decided that there would be strong evidence for a particular conclusion.

Things like, for example, that there are a number of high quality studies, three independent studies or more and these were all high quality. The directional differences were all in the same direction. They were consistent in other words. Or it could be a very large high quality multi-site study that would be in effect, a series of replications. And in this case there would be no negative evidence. So, this would be strong evidence at which all high quality studies would point to the same conclusion.

Now, I indicate here that there are a number of factors that affect the number of studies that we would take to be strong evidence. Again, we cannot do this very technical subject justice, by things such as error variance or just the natural variability in the measure, how sensitive the measures are. If the measures are not very sensitive, obviously you may need more than three studies. And the What Works Clearinghouse, of course, has been dealing with many of these issues and we cite them as a reference.

Moderately strong evidence would be one or two high quality study's effects, not necessarily independently replicated, and so on. So again, evidence, but not as much evidence, still all pointing in the same direction.

Suggestive evidence would be one of the things such as high quality studies that support a conclusion, but maybe other studies that may have a null result. Now, a null result as we know is the failure to detect a significant effect. It is not a negative effect or a contradictory effect that is covered under inconsistent evidence, which is below that.

I think the most important thing I can outline about inconsistent evidence is that results of high quality study designs trump inconsistent or null results of low quality designs. In other words, if you were to have three studies that say yes, there is an effect and three that say no, it is an opposite effect. Actually, perhaps the treatment group did worse than the control group. If the three studies that say that there is an effect are much more high quality, that is where the weight of evidence should be, that is where the strength of the conclusion is. Weak evidence, of course, is where there are only low quality studies available.

Again, as I mentioned earlier, standards of quality and the details of methodological rigor differ depending on the nature of your question. So, these are just three examples of different kinds of questions and examples of what we considered high quality. We get into more detail about different levels of quality for different questions in the document itself.

Effects of interventions are things that involve random assignment to condition. Low attrition is obviously a mark of high quality, valid and reliable measures. Valid and reliable of course, is very important. Sometimes valid can be a deep question, but it is one of the most fundamental questions in research.

A descriptive survey of course, has to have a representative sample, a low non-response rate and evidence that attrition was not biased. And many other standards are applied to that.

Tests and assessments are subject to a variety of psychometric standards, including some of the measurement issues that I have mentioned, such as validity, reliability and sensitivity. This document also comes with a set of references, despite the fact that it is not about empirical evidence. And I would direct people's attention to those. Some of the classic pieces on assessment and measurement have been covered in those references.

To conclude, all of the committees were charged to have some recommendations and our recommendation of course, revolves around standards of methodology. We noted, and this is not only the subcommittee but many other members of the Panel noted, that we 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. And therefore, the subcommittee recommended that the rigor and amount of course work in statistics and experimental design be increased in graduate training and education.

And to conclude, that kind of knowledge is essential to produce and to evaluate scientific research in areas of crucial national need, such as mathematics education. Thank you.

DR. FAULKNER: Thank you, Valerie. The subcommittee report is actually in the notebooks that are available to the panel members under Tab 8. We actually are at a stage where we need to carry out final discussion, if there is any, and make a formal adoption of this report as the basis for the Panel's activity. So, I would like to open the floor for discussion here. Was there additional discussion? There has been quite a lot of discussion of this in individual groups over a long period of time. Bert.

DR. FRISTEDT: I noticed that in some places the term scientific evidence is used and at other places the term evidence is used. I think it is important that both terms be used. I did not check exactly if I would agree where, but let us assume I would. It is important because there are many kinds. And as you commented in fact, there are many kinds of evidence that are not evidence based on the scientific experiment, but they are nevertheless, quite solid evidence. I think it is important that those words both appear, scientific evidence at some places, evidence at others. It seems that you have thought about that, where they should appear. Good.

DR. REYNA: Thank you very much.

DR. FAULKNER: Any other comments?

[No Verbal Response]

DR. FAULKNER: All right. I, the Chairman, entertain a motion to adopt the subcommittee report Guidelines for Standards of Evidence. Present a motion. Skip.

DR. FENNELL: So move.

DR. FAULKNER: Second?

DR. GERSTEN: Second.

DR. FAULKNER: Second. The mover was Skip Fennell, the seconder was Russell Gersten. Any other discussion?

[No Verbal Response]

DR. FAULKNER: Then all in favor of adoption please signify by saying aye.

ALL PANEL MEMBERS: Aye.

DR. FAULKNER: And those opposed?

[No Verbal Response]

DR. FAULKNER: There are none opposed. Valerie, we appreciate all the work that has gone into developing this report. And I should say for the benefit of the audience, that the reports that are about to come forward from the task groups have been developed using the standards of evidence that are represented here. So, I want to say that this report will be in operation momentarily here. Thank you.

DR. REYNA: Excellent, thank you very much.

DR. FAULKNER: Okay. We are now

ready to move into the Task Group Reports. And we will take the Task Group Reports in order, numeric order that we have used in the Panel for a long time. Each of the task groups will go forward and make their presentations at the testimony table. Each group will have around thirty to thirty-five minutes, twenty minutes for presentation of results, and ten to fifteen minutes for discussion.

Task Group 1, Conceptual Knowledge and Skills, is chaired by Skip Fennell. I am a member of Task Group 1 and will be going forward, so I will be turning the chair over to Vice Chair Camilla Benbow.

IV. TASK GROUP 1 - CONCEPTUAL KNOWLEDGE AND SKILLS

DR. FENNELL: Good morning, Panel colleagues and staff, who have essentially guided us through this effort. And importantly the public, who are going to review pretty much where we are to date, relative to the conceptual knowledge and skills, if you will, the math side of the panel's work.

Larry has indicated that as chair he is also a member of this Task Group.

Liping Ma, who is not here this morning is also a member of the group, as is Wilfred Schmid, who could not join us for today's meeting. Sandy Stotsky sits to my immediate left. For several meetings, we have referred to Hung-Hsi Wu as ex-officio to this task group because we lean on his expertise probably more than we should. But Dr. Wu allows us to do that. And finally, we could not do much without the able assistance of Tyrrell Flawn.

I am going to turn this over in a minute to Sandy to walk us through where we are in terms of how we have proceeded, and it will come back to me with regard to findings and recommendations. But our work has been guided by the three questions that you see in front of you.

What are the major topics of school Algebra? What are the essential mathematic concepts and skills that lead to success in Algebra and should be learned as a prerequisite, as preparation for Algebra? And then, does the sequence of topics prior to a formal Algebra or formal Algebra course work itself affect achievements in Algebra? Sandy will talk to us, to the group, about the methodological approach to our work.

DR. STOTSKY: Thank you very much, Skip. Let me just begin by pointing out that we have an introduction that also precedes this, in order to explain where these questions are coming from in the Executive Order. And then we have a methodological approach that we describe here so that you can see that we are using a combination of peer reviewed and published studies, as well as expert judgment. As you will see, this particular task force is going to be relying on a variety of different ways to address the three essential questions that derived from the Executive Order.

After we have an introduction that explains the background to the three questions and a description of our methodology, we also provide a context of student achievement in mathematics. Roman numeral III gives the actual contemporary context for looking at the problem that we are addressing in general, which is how best to improve mathematics education in this country. And here we rely upon needing some national kinds of information to inform us on what the problem is.

Then we move into what, like a Beethoven Sonata, we call the main theme, and its exposition. We have the major topics in school Algebra. These reflect the judgment of the mathematicians on the Panel and other mathematicians with whom they have consulted. These major topics are the main theme and then its exposition, as you see, is an overview of school Algebra. This is the centerpiece that launches the rest of the report. What is Algebra defined by? What are the topics? And the exposition, explanation discussion of the topics for school Algebra, which encompass both Algebra I and Algebra II.

From this main theme and its presentation or exposition, we then have secondary themes. If I can continue a little bit with my analogy, which is somewhat deliberate here, because I am trying to suggest the creative nature of part of this. We looked at where Algebra topics were in several different types of curriculum sources to bounce back and reflect on the main presentation of the major topics. We looked at state standards for Algebra I and II, in Algebra I and II textbooks, and in the leading country on the Third International Mathematics and Science Survey, Singapore; its mathematics curriculum for grades 7 to 10. We then looked at assessment sources as opposed to curriculum sources.

So, we looked at what were some major topics in school Algebra that were covered in the grade 12 National Assessment of Educational Progress (NAEP) test objects, and in the proposed American Diploma Project Benchmark and Test Objects for its Algebra II end of course test, which is now being piloted by a group of states on a voluntary basis.

We then have a comparisons section that shows how all of these different sources of topics reflect on our main intellectual objectives, which are these various major topics that we have listed before. And then we have a section that we call observations regarding rigor, which serves as a sense of transition to another section. But this includes some appendix material that is a focus on what are some of the errors that can be found in contemporary Algebra I and II textbooks. Earlier we had an appendix on what is in a number of contemporary Algebra I and II textbooks, and this is now a second appendix that should be of great value, we hope, that will indicate where there are some problems with accuracy. And we expect that the Instructional Materials Task Group will also be looking and using this kind of material.

Then we move into what we might call the development section, because once we have presented those major topics in school Algebra the central question is how do we help all students get to those major concepts and skills? What should they learn as preparation in order to arrive at formal Algebra that would be taught at the end of middle school or early high school?

We looked into international approaches to Pre-Algebra education in order to draw on what they had to offer from their research on curricula in other countries, particularly the work of William Schmidt and his colleagues on what was the curricula in what are called the A+ countries. Those are the six leading countries on the Trends in

International Mathematics and Science Study (TIMSS) international tests.

We also looked at material that had been developed, examining in greater depth what were curriculum approaches in these top performing countries compared to what appeared to be the approaches in American mathematics education.

We then turned our attention from these figures for this information, on what were the features of the curricula in these A+ countries, to what were some national approaches to Pre-Algebra education in this country. We looked at the most recent offering by the National Council of Teachers of Mathematics (NCTM), called Curriculum Focal Points. We then looked at the curriculum profile of the six highest rated curriculum frameworks in this country.

These were the two major sources that we looked at for gathering some information on a comparative basis about what is in the curriculum profile for the A+ countries that is not in these national sources. What is in these national sources that is not in the A+ countries curricula profiles. We then go on to present, after some supporting material from a recent ACT Curriculum Survey, and from our own Algebra Teacher Survey, what we are recommending as the critical foundations for success in Algebra.

That then culminates this very active development section. And then we just narrow to the main theme again in Roman numeral VI, but we are approaching it with three sub-questions in a very different way.

We are looking at first of all, the question of does the sequence of mathematics topics prior to and during formal Algebra course work affect Algebra achievement? We looked to see whether there were indeed any studies, any research that could address that. And we found that there was no research whatsoever that we could draw on for answering that question.

So, we then moved to (A), What was the research on the benefits of an integrated or single subject approach for the study of Algebra? And here we found that even though there might be a large number of studies out there, no conclusions could be drawn from what was a very deficient body of research. We could draw no conclusion whatsoever on this question about how one should approach, from this perspective, the study of Algebra.

The third sub-question, which is (B), was then the question of when might formal Algebra course work be best addressed; what were the pros and cons? Here, there was a small body of research to draw on that met the criteria for the panel's standards of evidence. We could find some other information and sources of statistics that would support a recommendation that would address the question of the timing of formal Algebra. We then concluded this sort of different way of looking at this whole question of Algebra.

And going to the grand finale, which is our list of seven, I now believe recommendations, conclusions and recommendations for pulling in different elements that come from all parts of the document. This is what we hope will be useful and strong recommendations that can improve the education of all of our students and help them achieve much more success with Algebra whenever they do take it, at the end of middle school or beginning of high school.

DR. FENNELL: Thanks Sandy. Just as a point of information, and we will not be reading these, but just to have public record, at our last meeting in Miami, as a part of the public record, the major topics of Algebra were presented. I put them in this slot, if you will, solely for evidence of prior work.

We also read into public record in Miami the critical foundations for Algebra as noted on this slide, and there will be a bit more discussion of that within the recommendations.

Finally, we will move to the findings and recommendations. This slide talks about sort of the directions that the recommendations will move.

DR. FAULKNER: Let me just highlight the bullet points that are there. The task group affirms that Algebra is a gateway to more advanced mathematics and most post-secondary education. This was, of course, behind the President's Executive Order, in the fact that it has a focus on Algebra and charges this task group or excuse me this Panel, with principally addressing the question of how to get students ready for entry into and success at Algebra. This task group is affirming the important role of Algebra as a gateway in the educational process.

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. In other words, we are seeking a strong focus on this mission and the concept that basically it is a universal goal. It is much more important for our 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. But we stress that it is important that whatever courses are received, those students get legitimate Algebra courses and that they be well taught and that the students be prepared for them, rather than that the students get an early start.

To improve the teaching of Algebra the task group proposes the following six recommendations, which Skip, I think, is going to handle.

DR. FENNELL: Yes. Point of clarification. The word “finding” there should probably be thought of as a preamble not necessarily a finding. It sets up, if you will, these recommendations, and in fact, I believe there are seven.

The task group recommends that school Algebra be consistently understood in terms of the list of Major Topics of School Algebra (MTSA). This is the acronym of the morning provided in this report. The list of Major Topics in 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 by grade 11 at the latest.

Recommendation three. 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.

Supportive statements. 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 we define here 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.

Recommendation four. International studies show that high achieving nations teach for mastery in a few topics in comparison with our 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.

Recommendation five. 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 two; that is the topics that we have, and that students be mathematically prepared for such a course in the sense of recommendation three, meaning the critical foundations.

Anecdotal comment. Far too often we, if you will, push kids into a course called Algebra without the appropriate preparation. It is not good enough to say that we have (X) percentage of students doing Algebra at whatever grade level, it is more important to make sure that they are ready for that regardless of where that happens.

Six. 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.

Finally, recommendation seven. 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. That by the way, means not essentially the entire preparation of such teachers, but certainly fundamental in their background. Questions?

DR. BENBOW: We are open now for questions.

MR. SIMON: Skip, I have always felt from the beginning that your section, your task group's work was going to really be the heart of this whole report, because that is where a majority of the people that have to actually implement mathematics education in this country are going to go, right? What is it that we need to teach these kids? And I would give you guys an early grade, maybe an A minus on your section.

DR. FENNELL: We are all pretty competitive, so I want to know what I need to do to get an (A). So, what am I missing here, Raymond?

MR. SIMON: Yours, out of all the sections, needs to be an A plus. And in my opinion what you need to do to make it an A plus, and I would ask you to consider this and I would ask the Panel to support this. It is in the section dealing with essential concepts and skills that should be learned in preparation for Algebra.

DR. FENNELL: The critical foundation?

MR. SIMON: Yes. And I know you all have had a lot of debate within your committee on this, and I am not here to advocate for a grade-by-grade detailed listing of topics to be covered. I am not asking for that. But, I think the three grade cluster or the three clusters that you, --

DR. FENNELL: The foundations, -- critical foundations?

MR. SIMON: Those three cluster areas, in my opinion, do not go far enough to give guidance to the teachers in the state departments that are going to be setting standards and looking at this. So, I think you need a little bigger balance. We do not want a balance between being too specific to tell the states what they need to do, and you need to respect their curricula, and you need to respect individual differences, but you have to balance that against a growing number of schools who do not know what to do. We had some pretty impassioned comments yesterday, testimony from parents. We need to give these parents the best tools they have to go back home and say look, this is what we need for our kids here.

I would like you guys to consider at least putting in some benchmarks at certain grade levels that I think would also address the spiraling issue you talked about that keeps repeating and is never brought to closure. So, I do not know what the magic grades are in math. I know in reading it is third grade. By third grade you need to know X, Y and Z, or you are not going to be a good reader.

So, there has got to be benchmarks in my opinion, for the math. If you could put those in there I think I would give you an A plus. And it would be so much more helpful for teachers and for states as they set standards and as we move forward with the revisions to our math curriculum. I just think that it would be so helpful. So, I would ask you to consider that and I would ask the Panel to support that.

DR. FENNELL: Thank you, Mr. Secretary, that is a great comment and you know we are going to shoot for that A plus.

DR. BENBOW: Bert.

DR. FRISTEDT: I agree with so much that is in what you have written, but I have some severe problems with the organization and the messages that typical readers might get out of it. I can see focus is at certain places, but I am not sure from what I have seen here, that the message is going to be read by many the same way I would read it. So, I am in the position of thinking that some significant tweaking is necessary. At the same time as agreeing with practically everything, let me give you some examples and maybe that will suffice.

The combination of what I have heard verbally and some words I have seen on the screen, I saw the following words, Algebra, formal Algebra, authentic Algebra, legitimate Algebra, Algebra, Algebra I, Algebra II. I have a pretty good idea of what you mean by each of those, but this is a communication issue and I think that has to be dealt with in a very systematic manner. I noticed in the things leading up to Algebra an old fashioned word, arithmetic. I do not think I saw that once. So, I think that something has to be done on the communication side.

Also, I am concerned about being overly specific at places where you do not need to be. I, myself, am in favor of two high school Algebra courses, one called Algebra I and one called Algebra II. But what

I am really in favor of, regardless of how the courses are arranged, is that the list of topics listed under Algebra be fundamental and those are heavily calculational topics. That is the message that has to get through.

DR. BENBOW: This has to be the last comment so other people have an opportunity. I can come back to you, okay? Doug.

DR. CLEMENTS: Douglas Clements. Skip, and the rest of us, the sub-panel here would just like you to comment, because I think it is in your report, but maybe you have had limited time here to present it. Is this to be interpreted as critical foundations, the entire elementary curriculum for instance or are there more aspects of it that should be seen as just those foundations leading to Algebra?

DR. FENNELL: That is a great question and one that we want to be very clear about. Children from Pre-K through this opportunity called Algebra, regardless of what grade that happens, do a full curriculum in mathematics. And by a full curriculum what I mean are things that you did not see on any of the slides, i.e. measurements, i.e. some opportunity to analyze data, and the like. What we are suggesting are those critical foundations. We want children to know how to add, subtract, multiply and divide whole numbers. I think that is arithmetic. And also they should have a similar capacity with fractions, including decimals, percents and so forth. That whole area of work with whole numbers and work with fractions and particular aspects of geometry are actually foundational to that opportunity called Algebra. So, it is a subset of that experience we know as Pre-K through Algebra in mathematics, but a critically important subset.

DR. LOVELESS: Thanks. One more quick question if I could? You mentioned looking at these top six countries performing on Trends in International Mathematics and Science Study (TIMSS), looking at their curriculum. Did you also take the extra step that I do not think people always do, of looking at the bottom six or the middle six? Because if the bottom six indeed share certain characteristics of those same curricula, then logically those characteristics are not what is happening in terms of differentiating a high from a low scoring country.

DR. FENNELL: Good point and no, we did not.

DR. CLEMENTS: I think it is important to at least have a caveat about that.

DR. BENBOW: Tom then Wu.

DR. LOVELESS: Doug actually made my point. I was going to urge some caution in looking at the A plus countries. When you run regressions and use curricular variables to explain variance in Trends in International Mathematics and Science Study (TIMSS) scores you do not get a huge effect. So, he is quite right. The bottom scoring country, which is South Africa, has a classical curriculum. You will find many of the same characteristics in the South African math curriculum as you do in the A plus countries.

DR. FENNELL: Yes, thanks.

DR. BENBOW: Wu.

DR. WU: I want to agree with what Bert said a minute ago about our multi-faceted use of the word Algebra. That has to be clarified I think, to make sure that it is quite clear.

What I want to raise is a technical point and I thought the subcommittee had agreed to use the reference to find the probability in mathematics strictly as a simple application from the binomial theorems. That should make sense because we do not want the nation to misunderstand that finding probability and common economics is a major topic in Algebra. That was something that we agreed on.

DR. FENNELL: That was a good point. And the list that I just showed came from the presentation in Miami when it was stripped out as a separate topic. So that is a good catch. Our document, the actual report does have it folded in rather than listed separately.

DR. WU: Thanks.

DR. BENBOW: I have a comment then. I want to second Simon’s suggestion earlier, it is a comment that I just realized and I think will be very helpful to state departments in terms of developing curricula that we would all be proud of. I would also like to hear a little bit more. You commented on it, you were not opposed to it, for some students to even get Algebra before the eighth grade.

DR. FENNELL: I think that there are a couple of issues, Camilla. One is, we want to make sure that all children have access to Algebra when they are ready. And so it looks at both sides of that. I have this interesting part-time job and so I hear from people from all over the country. Last week I talked to a parent of a fifth grader who is doing Algebra II. That child had access to that. That child had all the prerequisites necessary for that opportunity. For legions of students, that is going to occur later in their educational background.

What I also want to be very sensitive to is the teacher who is receiving students for their first formal experience in mathematics. I think this teacher has the right to expect some prerequisite knowledge before that mathematics is begun.

I am looking very directly at Vern Williams who faces this issue probably everyday in his teaching career. So, we want to look at both sides of that. Opportunity and access for students whenever they are ready, regardless of grade level frankly; but also appending to those critical prerequisites that will allow them a level of success in mathematics.

DR. STOTSKY: I just wanted to add to your comment that the actual text itself before the recommendation, does talk about grades 7 or 8. It usually puts the two together because some of the studies do mention the possibility of offering Algebra I in grade 7 or beginning it in grade 7. So, we have concluded it there. It is not exactly in the recommendation directly.

DR. BENBOW: Thank you. Last question to Bob and if you have more comments I would just ask that you talk to Skip or the panel.

DR. SIEGLER: Yes, I would like to remind the subcommittee and also the Panel as a whole, about the previous discussions that we have had regarding the recommendations for age norms and grade norms for teaching particular topics. And Wilfred Schmid who is not here, and I and Doug Clement, and a number of people, have made the point that there is nothing really in either the empirical evidence from psychology nor in the logic of mathematics that says that topic (X) should be taught in grade (N). And we have gone round and round on this topic. And it would be great if there were some empirical evidence that would tell you when to teach these topics, but the fact is there is not. Given that the National Council of Teachers of Mathematics (NCTM) Focal Points does the best job I think possible at present to provide reasonable recommendations, I think the Panel might want to steer clear on linking specific grade levels to particular topics. I just do not think the evidence base is there.

DR. FENNELL: You have no idea what a difficult position you have put me in, in attempting to respond to that. So, I hear from the Deputy Secretary about thinking about, I am going to say grade bands. By grade 3 students ought to be able to do X, or Y or X or whatever. We have in fact, in numerous ways, had the discussion Bob, as you very appropriately outlined. I think for right now this task force needs to take all of this under advisement, but that is a very good point.

DR. BENBOW: Thank you very much. All right. Let us move on to the next task group's report, and that is Learning Processes, which is chaired by David Geary.

DR. FAULKNER: Before Learning Processes gets started, let me amplify for the audience where exactly we are in preparing these reports. Each of these task groups is coming forward to provide outlines of their task and their reports and their major findings and recommendations. There are significant drafts, big drafts of these reports and they are not in final form.

Each of the task groups has worked extensively here in St. Louis yesterday, most of the day, some the day before, and will be carrying away a need to complete some revisions. We will also be receiving comments here in this session and will produce what amounts to a final version or at least a reviewable version of each task group report by the 21st of September. That is the goal.

I want to indicate to the audience that very substantial, far along drafts of all these reports do exist, 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 and that is what were are in the process of conveying here.

So, let me turn it over to Dave Geary who is chairing Learning Processes.

V. TASK GROUP 2 - LEARNING PROCESSES

DR. GEARY: All right, thank you Larry. I want to begin by acknowledging my very able colleagues, Bob Siegler, who you have heard from, Dan Berch and Valerie Reyna, who you have also heard from. Wade Boykin was unable to make it to this meeting. Susan Embretson is also on the committee and does not seem to be here, and Jennifer Graban has been just a tremendous help in preparing this report and keeping us on task.

Our report covers general principles of learning, including cognitive processes and learning outcomes; working memory; social, motivational and affective influences on learning; mathematic knowledge children bring to school and mathematics learning and cognition in the areas listed there. Within those areas we focused on content topics that were deemed critical by Skip’s task group that you just heard from. So, we focused on some things and not on others.

We cover individual and group differences, specifically race, socioeconomic status, gender, learning disability and gifted students. And finally, we reviewed the research on brain sciences and math learning.

The methodologies used in the conclusions of this task group were based primarily on studies that test explicit hypothesis 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. Conclusions are qualified when the evidence is not strong, and suggestions for research that will strengthen the ability to draw conclusions is provided. There were multiple approaches, procedures, and study types reviewed and assessed with regard to convergent results using a variety of methodologies shown on the screen.

With respect to the literature search we looked at key mathematical content terms linked with learning and cognitive processes. Our first search looked at peer reviewed learning, cognition and developmental journals. We then conducted a second search that supplemented the first and included other empirical journals, Social Science Index and Psych Info and Web of Science.

Criteria for inclusion are as follows: published in English, participants aged three and older, published in peer reviewed empirical journal or review of empirical research and books or annual reviews. The research was experimental, quasi-experimental or correlational.

Turning to just a brief overview of some of the types of things we cover under general principles, from cognition to learning. There is a great deal of scientific knowledge on learning and cognition that could be applied to improve student achievement, but it is not currently being used in the nation’s classroom.

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 new problems and better long-term retention, higher order thinking and problem solving. This presumes acquisition of basic skills is not only necessary for entry into the scientific and technical workforce, but also becoming increasingly important for achieving success in other kinds of occupations. 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.

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 U.S. do not solve simple arithmetic problems as fast, as 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 practiced using algorithms combined with instruction or related concepts.

Moving to social, motivational and affective influences. We want to note that Vygotsky’s Socio-cultural prospective has been influential in education. His theory treats learning as a social induction process through which learners become increasing able to function independently through the tutelage of more knowledgeable peers and adults. However, due to shortage of controlled experiments, the usefulness of this approach for improving math learning is difficult to evaluate at this time.

We do have 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. 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. However, intrinsic motivation declines across grades, especially in mathematics and the sciences as material becomes increasingly complex. There are educational interventions, which are part of the educational environment that can influence students' intrinsic motivation to learn in later grades.

Relative to children in nations with high mathematics achievement, children in the U.S. 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.

Turning to 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, at least through high school and probably 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.

Let’s turn now to the mathematical content area. We reviewed quite a bit of research in these areas and we organize our recommendations around classroom practices or research needed to facilitate these practices, training of teachers and future researchers, curriculum, including content and textbooks and basic and applied research in these areas.

The task group cannot review all the basic findings in these areas of all the corresponding recommendations here, although of course they will be in our final report. The task group here highlights core points. 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.

Beginning with whole number arithmetic, 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, such as poor knowledge of the base-10 system.

By the end of elementary school the majority of children do not appear to understand many basic concepts, including the distributive property, 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, development of mechanisms that facilitate the translation of basic research into knowledge useable in the classroom.

Moving to fractions. Fractions are formally introduced in elementary school, yet remain difficult for many adults. Twenty-seven percent of eighth graders cannot correctly shade 1/3 of a rectangle in the 2005 National Assessment on Educational Progress

(NAEP) Assessment. 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, vis-à-vis for example medication adherence.

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 on a physical and ultimately 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 in this crucial domain of mathematics. 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.

Moving to estimation. Numerical estimation is an important part of mathematical cognition, both because it is frequently used in everyday life as well as in scientific, mathematical and technical professions and because it is closely related to overall mathematics achievement. Poor estimation performance often reveals underlying difficulties in understanding mathematics in general. In some classrooms estimation has been equated with rounding to such an extent that children do not know that the purpose of estimation is to approximate the correct value.

Children’s estimation of the magnitudes of fractions is especially poor. Structural programs for helping children accurately estimate fractional magnitudes are urgently needed.

Moving to geometry, and I note here that we focused only on those aspects of geometry that were highlighted by the Conceptual Knowledge and Skills Group. This is not to slight other areas of geometry, but this was our charge.

Of the five mathematical content areas assessed by the 2003 Trends in International Mathematics and Science Study (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 Trends in International Mathematics Science Study (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 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 such as geo-boards, dynamic software and so forth, evidence regarding their usefulness in helping children learn geometry 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. Thus, leading to a mathematical understanding of linearity.

Moving to algebra. Cognitive studies of algebra focused on linear equations and word problems have revealed that many students in high school algebra courses are woefully unprepared for learning the basics of algebra. The errors students make when solving algebraic equations reveal that many 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.

We turn now to individual and group differences, beginning with learning disabilities. The empirical evidence suggests that between 5 and 10 percent of students will experience a significant learning disability or learning difficulty in mathematics before completing high school. This is above and beyond instructional or other factors. 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.

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

Turning to gifted students. The few cognitive studies of the sources of the accelerated learning 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.

Turning to gender or sex differences. For national 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 U.S. is poor.

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

Our final content topic is focused on brain sciences in mathematics learning. We 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 domains 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.

There are some general conclusions with respect to research. For all areas, and as we noted earlier, a pipeline of research must be funded that extends from basic science of learning to field studies of classrooms. We recommend incentives to encourage partnerships between basic and applied researchers. The many interventions demonstrated to be effective in experiments should be scaled up and evaluated in classrooms. Research is essential to discover mechanisms that contribute to emergence of formal competencies in schools linking their earlier intuitive understanding with later formal mathematical problem solving.

Educational research must be integrated with basic research in cognition, motivation, neuroscience and social psychology. Educationally relevant research need not be conducted in classrooms. Research conducted in laboratories under carefully controlled conditions can often be directly applied in classrooms. Incentives are needed to encourage more scientists to perform instructionally relevant research and participate in research teams that will translate basic science findings into instructional interventions. More research is needed that specifically links cognitive theory-driven research and classroom context.

At the same time, cognitive research on learning needs to take into account more facets of classroom settings if it is to eventually have a greater impact on instruction.

We will also make recommendations regarding teacher training and curricula as I noted, a few of which are noted here. We recommend instruction in scientific method in evaluating research evidence, 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. Questions?

DR. FAULKNER: Questions or comments? Russ.

DR. WHITEHURST: There is certainly a lot of information in your report. I wonder if somebody on the Panel would venture an answer to this question. Given this voluminous research on learning with respect to math, what are the three most important things that policy makers or educators should do to translate that research into changes in current practice?

DR. GEARY: The first thing is to read our report, I think. Well, we obviously have a lot to say in here. There was a lot of work to be reviewed and covered and to be made relevant to the topics and content identified by the concepts, knowledge and skills group. There are many messages that need to be heeded, one of which just in a basic learning perspective is that you cannot separate conceptual and procedural learning. Much of the math wars has been kind of based on this false dichotomy that you teach children concepts or you teach them procedures, when in fact we have empirical evidence that they bootstrap and interact with one another. That needs to be made clear. And we need to understand better how those interactions occur and particularly for the core content areas in algebra and leading up to algebra.

DR. SIEGLER: Yes, so I provided two and three. Dave's number one actually was on my list when Russ posed the question, too. So, two others that I think are well grounded and important implications are first of all, that programs for improving low income preschoolers' mathematical knowledge are at the point where scaling up is really appropriate. They have met all kinds of criteria as the What Works Clearinghouse has certified for a couple of them just recently, and I think that it is essential to provide ways of preventing these early, relatively small, though still substantial deficits, from growing into the huge intractable ones that we are all too familiar with in later grades.

And the other kind of strong policy recommendation has to do with fractions. I think all of us have been somewhat surprised when we looked at the literature, just how bad understanding of fractions really is. And we have been very influenced by the Conceptual Knowledge and Skills (CKS) group’s analysis of the absolutely essential quality of understanding fractions for learning algebra. It makes a lot of sense when you think about it, but it certainly was not something I had thought about a lot before. So that’s two and three on my list.

DR. BERCH: First, I’ll just

comment that we are in the process of trying to distill the major recommendations for that next step and for the new groups that are forming in terms of the Final Report, and thank you for pointing out the need to do that. But, I think one of them is clearly right there under the next to the last bullet, 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: I want to underline our recommendation about teacher training which is in front of you. The theory here is that if we increase not only the teacher training, but also all personnel throughout education, from superintendents to principals who have a conversant familiarity with the essentials of children's learning, that it might create more demand for the kind of research we have been talking about at the level of practical implementation in the classroom. The supply of course of basic research is essential to that equation as well.

DR. FAULKNER: Deborah.

DR. BALL: This goes back to a question that I asked when we began this work in our second meeting I think, and I sort of warned us that one of the problems I thought we would encounter. How do we sort out whether what we see in literature is evidence of how kids develop and learn or whether we see the effects of instruction? So, how do you sort out when you make claims about children learning, what they can learn, what they typically have trouble with from the instruction they have received? To me this seems an absolutely fundamental question for the Panel to consider. How do you as people who do research in this area sort this out? Kids are not just like in the wild developing, so?

DR. GEARY: You obviously have not met my children, but that is for another committee. Yes, that is an excellent question, and it is a very difficult one to answer, because we do not have random assignment to classrooms versus the park for x number of years. It is very difficult because children have kind of a natural development of their abilities to learn as well as being placed in a context in which they are expected to learn certain types of things.

I think the most important point that we can make is that the assumption about readiness in stages and so forth we know is not correct. We know children are capable of learning much more than they have been learning. The extent of that, how far down it can be pushed or how far down it is beneficial to push it or how far we can accelerate it and so forth, these are really important questions that we do not fully have the answer to. With great respect, how do we know if it is instruction or general development?

There are experimental studies on having children practice problems or solve algorithms or do facts or whatever, and we know that things like repeated presentation will have certain effects on their learning and will improve their learning in certain ways and so forth. So, we know something about their learning, but we do not know fully the interactions between brain maturation, natural cognitive development as that is embedded in a classroom. Those are great questions. And hopefully someday we will know more about them.

DR. SIEGLER: So, your question is a version of the very general heredity and environment question that pervades all of psychology and lots of other social science disciplines as well. One way that people are addressing it in the area of mathematical development is through cross-national studies.

We learn things, for example, that Chinese preschoolers, unlike preschoolers in the U.S. and a number of other Western countries, come to school knowing not only a lot more arithmetic and other skills like counting that some families teach directly, but also skills that no families in either culture seem to teach directly, such as number line estimation. These are studies before the kids ever set foot in formal schooling.

There are influences that are environmental but that are not part of schooling that influence children's learning.

In other cases we learned just what is possible. Again, studying East Asian versus Western, in particular U.S. achievement, on things like fractions. East

Asian kids are way more advanced at the same age level than kids in the U.S. So, this points to some combination of the educational system, the general culture, the values that are embodied in the culture and so on, as making it possible. It is not like these problems are just part of the human condition; they are things that can be changed.

DR. REYNA: There is also the notion of random assignment here that is very important and counter-intuitive. The idea here is that if you have children who range in instruction prior experience and you randomly assign them into two groups and then do an experiment on their learning, you can try to separate these influences. There may be interactions but they’re distributed in both groups. So, that is one way we were able to make these kinds inferences.

I would also say there are some very interesting studies, and I won’t go into details due to lack of time, where people have looked at, Fred Morrison, for example, at first graders who are the same biological age, but because their birthday falls before or after enrollment in formal schooling. You can compare brain maturation, physical and cognitive maturation and compare that to the effects of formal instruction. And bottom line, of course is instruction makes a big difference.

DR. WU: Now, well I guess what I have to say, it pretty much has been said, but let me just make it more specific, on the subject of fractions. In my opinion, the problem with evaluating the non-learning of fractions is that I think largely it is not so much because children cannot learn fractions, but rather they have been taught so badly in schools to judge by what goes on in the textbooks of all kinds. It doesn't matter if it's formal or traditional. When you are taught so badly, then I think it is very hard for them to learn things. I was wondering whether there is any possibility at some point, even if it is expensive, that you run an experiment where one control group is taught the usual way and then you have some separate class of students taught more correctly and then evaluate their learning achievement or whether that is even remotely possible.

DR. REYNA: Not only remotely possible, we strongly advise that it be done.

DR. WU: But expensive?

DR. REYNA: Not even that expensive.

DR. FAULKNER: Bert.

DR. FRISTEDT: I observed this one point you made about the decreasing interest in mathematics as students increase in grades. I was wondering if you have any feeling or evidence about what is the cause. I have a couple of speculations, and I am wondering if they match.

One speculation is that the book in the previous grade was so long that the teacher did not finish it and then in the next grade the student feels lost. Or another speculation is the book in the next grade took too many pages of straight review from the previous grade, and the student gets bored. And of course, I am asking these because they are related to this other job that I have of being on the materials subcommittee.

DR. GEARY: Well, that is a good question. I will offer some speculation as well. Those may contribute to that, we do not fully understand that change yet. In getting back to Deborah’s point, I mean you are also dealing with kids who are maturing, going through puberty, social peer issues, other sorts of things may become more prominent for them. In fact, there are studies of that--looking at what kids prefer to do and when they are the happiest with what they are doing. It tends to be lowest when they are doing things like mathematics homework; unbelievable to me, and all of you I am sure, and highest when they are hanging out with their friends.

So, there are multiple issues going on there, not only with the curriculum, but just the general biological development of these children leads to a focus on these things. Now, it does not mean that going to the mall is more important than mathematics homework, it is not, it is just they do not understand that.

DR. FAULKNER: Russell, you get the last question.

DR. GERSTEN: This is more a suggestion than a question. That as you begin to crystallize your recommendations, and I think this process has been a good one, it would be good to continue to focus. There are many, many interesting things, but the idea that the need for rigorous serious experimental study of teaching fractions given its importance would be one.

But this is, I will tell you, the missing link in our report, and I am sure others will see it. The fact that logically knowledge, deep knowledge proficiency with fractions, seems critical for success in algebra as do the others. I have asked Dave and they have searched and no one has found any empirical evidence, to document that kids who in the fourth and fifth grade do not know fractions are tending to do badly in algebra? The kind of work that Wade was alluding to that Connie Jewel and others did twenty years ago in reading. I think that is critical. If there is any way we can call for use of a national representative data set, correlational work could be done. It seems to be critical for us to really have an empirical basis for this logical analysis and also to learn more about what are the patterns, what are the things about fractions that are critical for success in algebra.

So, I would urge that as the number one research bit that needs to be done or through secondary analyses.

DR. GEARY: Two points. The first is that when we began this process we asked Abt and others to look at national databases to look for longitudinal studies that would allow us to look at early predictors of outcomes in algebra. And we were very surprised that there is no appropriate data available in these large national studies to allow us to really make those links.

The other point is if we look at students solving of linear equations and we look at error patterns, many of the error patterns, and even more sophisticated studies look at how they are actually tracking and processing the equations, many of those patterns strongly imply that the students do not really understand ratios, fractions. If they did understand them, they would not make the kind of errors that they make.

DR. FAULKNER: Okay, let me express appreciation to the Learning Processes Task Group who have worked collaboratively and have presented a lot of information here, more than we can fully discuss on this occasion.

We are going to break right now. I am going to ask people to come back in ten minutes. We have got three more reports that we need to receive before we finish today.

[Whereupon, at 10:40 a.m. this meeting was recessed to resume the same day at 10:50 a.m.]

DR. FAULKNER: Thanks to the audience. During the break people indicated to me that there have been several questions among members of the audience about availability of materials. I would like to indicate that the draft reports are not available because they are still in draft stage. The materials that are being presented here as these PowerPoint presentations will be posted on the website of the Panel, which is at the U.S. Department of Education website, which is ed.gov I think.

So, the U.S. Department of Education website will host a Panel site, and at that Panel site you will have the postable materials from every meeting not just this meeting. Also, there have been questions about the Miami meeting and so forth. But each of these presentations that you see here will be available on the website. Also, the adopted report on standards of evidence will be on the website.

The actual Task Group Reports will not appear until they get to an approved stage, and that is not where we are. So, I think that answers that question. We now go to the presentation of the Instructional Practices Task Group, and Russell Gersten chairs that. Russell.

VI. <u>TASK GROUP 3 - INSTRUCTIONAL PRACTICES</u>

DR. GERSTEN: Our topic is Instructional Practices, which is how to teach math well or effectively, which is a massive intricate topic, which we approached. This is the members of the group, and they are all here with me. There is Doug, Bert, Camilla, Tom, Vern and Irma and also Marian Banfield has been invaluable in terms of our Department of Ed. support person. Joan is not here because she had to be in Washington.

This is what we took to heart from the beginning, the charge and what Valerie presented to us a couple of hours ago. And you notice how rigorous the standards were. That came a lot from the National Academy of Sciences report and other work from the Institute of Educational Sciences, and we took that to heart.

Now, this is what we know. We were going to not look at a lot of studies. And the issue we had was there was part of, I think, each of us who thought it might be interesting to look at descriptions of effective practice, qualitative studies, correlational studies. The idea was, though, that once we did that we opened things up to large, large degrees of discretion. By sticking to strict standards, and these are very, very strict, they are very similar to the National Reading Panel, except they are more rigorous due to advances in methodology from 1998 to 2007, where we use some state of the art statistical techniques. So, it limited the number of studies we found.

But there is a sense that these are the kind of studies, as Val talked about a couple of hours ago, from which if there is a pattern of findings three or more of these high quality studies, we can draw inferences about effective practices. Now, it does not necessarily answer problems, but this is the way we decided to go.

We have a list of topics and when you see them they won't necessarily see how they fit together, and they did not come from one clear framework. We really tried to look at two types of things in particular. These would be things like, in some schools, in some states, in certain areas people have said real world problems are critical. Another thing is guided inquiry. We heard in the testimony yesterday there are some districts who say this is the way math will be taught to all students, or all students but the honor students. The other thing is the idea of enrichment programs for gifted, is there any evidence there? So, that was one reason for picking things, because they are these hot button issues.

We also chose some subjects, or at least I made some strong suggestions, because they seemed topics of importance and have come up in the National Council of Teachers of Mathematics (NCTM) surveys, but also that I knew that there would be some quality experimental research there. Because we did not want to come up with a report that said we do not really know much about these seven things, goodbye and lead a good life. This is one topic there was a lot of research. Twelve high quality studies met all these quite rigorous standards.

What these looked at is, if teachers use a formative assessment system weekly, bi-weekly and there are other randomly assigned classrooms in the same school, do their kids learn more math on an array of math achievement measures than the teachers who don't use formative assessment? And the answer is yes, and that is consistently replicated. The effect size is 0.2. I am not going to use any adjectives to describe it, but the fact that it borders on significance is very good. We have this technical statistical issue. We have effect sizes at the student level, in the class level. They are higher at the class level. There is no known way to put the two together. But we do have a nice picture there, that use of formative assessment. The picture is one with perhaps implications for No Child Left Behind, that in mathematics, as well as what is increasingly done in reading, when teachers use this information to alter and differentiate instruction, students learn more math.

The second finding was the effect size actually doubled if they had something that we call enhancement. By enhancements we mean a whole array of things. One would be after the formative assessment, you have a set of four kids who are in the fifth grade but they still cannot add fractions or they do not know what they are doing with improper fractions or long division. They do not have the rudiments of long division. So, the teachers have some information for tutoring the children themselves, providing interventions. Sometimes expert math teachers or math coaches came up with ideas. So, here are kids who did not learn this the first time, here are some ways to go through this material in more depth with the kids and move at a more deliberate pace. So, there is a whole array of things. But basically if teachers get this additional information and not just the numbers, the effects are doubled.

There are some limits here, I am going to go through them quickly, and these slides will be available. All studies but one were done at the elementary level. Most of the enhancements, not all, were done with special ed. students.

This is the most important limit. All of this research was done with one type of formative assessment. It is one that is widely advocated but does not appear in most texts . This is a sampling from the year’s major objectives from the state standard so kids have items. So, if about 20 percent of the objectives deal with operations involving decimals, one-fifth of their bi-weekly tests deal with decimals. So, basically it is one type of assessment. We do not know for now about other types of formative assessments at this point in time. There are suggestions that it would be useful there, but that would be where we were going beyond the hard data. So, there is definitely a type that will help at the elementary level.

This was one that Tom Loveless was the lead on. Teacher-directed versus student-centered. This has been a debate that has gone on probably since Socrates at least. Basically what Tom found when looking at studies that pit one against the other, that the only finding that emerged in this literature search was something called Team Assisted Instruction. Bob Slavin and his colleagues developed and studied in the '80s and '90s, and then became infused in Success for All, something very similar to Team Assisted Instruction. It was the only thing with consistently positive effects in the area of computation only not in concepts.

This is something that really is a hybrid. There is a teacher-directed part. There is a very explicit instruction part and then there is a way where kids work with each other and teachers use formative assessment.

Probably the most important finding was there is no data that supports in a consistent way from high quality experiments, student-centered instruction as the way to go, direct or teacher-directed instruction as the way to go or any other instructional regimen for the average student or the high ability student.

So, if somebody says we know for

sure that direct instruction is the only way to teach math or somebody says we know for sure that child-centered inquiry is the best way to teach math, there is no scientific evidence there. That’s our conclusion.

Real World Problems has been basically a huge theme, and Joan Ferrini-Mundy was the lead research analyst on this along with the Abt researchers. She really tried to look at two things. The first one is does it really help kids learn math if part of their instruction involves the real world type problems that are in many texts. Does it help kids learn mathematics by including them? There are many who advocate that.

Second question. Are there better ways to teach kids so that they can solve these more complex real world word problems? So, there are two questions. To cut to the chase, not enough is known yet about question two. There are some promising ideas that are discussed in a very tentative way in the report.

For question one we have five high quality studies that put together have a pooled effect that is significant when you look at all the measures. These were specific topics sometimes involving geometry and fractions and sometimes they were just kind of multi-step problems. When you looked at all of the measures that those researchers used, some of which were invented, they all dealt with applications of mathematics. None of them were wacky kind of things. When you looked at all the measures you got an effect. However, what Joan and the analyst did is they also looked only at typical math achievement, the typical word problems or occasionally computation things involving fractions. When you looked only at those measures, you no longer have a significant effect.

This is in Valerie’s criteria. There was only one of the five that had any negative outcome. So, it is a sign it is in a good direction. So, what it says is kids can learn to do real world problems. We do not get an effect in achievement. It does not mean if they had more of them or if they were part of a curriculum that there is more success. So, there is some promise there. How it gets integrated and infused, there is a long way to go. But that is one of our few robust findings along with the formative assessment and cooperative learning.

Camilla was the lead researcher for the section on gifted students. The consistent finding was there were no known negative impacts as the studies accrued. There were no known impacts. It was also noted that enrichment, despite the wide advocacy for it, there is virtually no research, save for one study. The other thing that the enrichment study did, and it is a point that the Panel discussed and I think it is an important one, that in reality a good or a mathematically sound enrichment program is likely to include acceleration. As kids dig deeper into the math that underlies third grade material or fifth grade material, they start to dig into number theory and algebraic concepts or reasoning. So, that is the gifted.

Low achieving students. We went with this though it was very hard to operationally define this group. We are basically talking about the lowest third of the population in math, and there needed to be some objective measure of that. But there were not a large number of studies. We did look at school-wide reform, because if it was a school-wide reform there would be a math component, but there is also a reading component, professional development et cetera, et cetera. So, we did not look at those. Other folks in other areas are looking at that in terms of the Institute of Education Science (IES) practice guide and other things.

We categorized them either as explicit instruction or other strategies. We have a sense of what explicit instruction is, but it is a construct that is very hard to unpack. And I see a lot of the future of work in instructional practice in this area. For explicit instruction for low achieving there were five studies.

When they were pooled the effect size was 0.97, which makes many of us happy. It is a nice hefty effect. Most of the studies focused on word problems, not real world problems, whatever exactly that means, but word problems. There was an array in these five studies. And basically in the learning disabilities theory, the same kinds of themes emerge. Some of them were the approach we heard about a little bit from one of the public testimonies, the connected math concepts where there are clear models. Kids were taught these steps to do and told when they do them and when they are appropriate. There is a lot of unison response scripted.

Some of the other models have parts where the teacher models things but they are much more interactive, probing for misconceptions. It is a mixture. They are kind of explicit components with other kinds of things added into them. So, they have a very different feel if you would see a video of some of the interventions that John Woodward has done and Von Rueden and Holland has done, versus the kind of ones that his colleagues have done. They just look different and feel different to kids, they are just different, but both types seem effective.

One thing that was noteworthy, and this is again both in lower achieving and the studies, with children with learning disabilities is careful sequencing of examples was almost always stressed as they described and gave examples of the instructional approaches.

One thing, I mean it is just one study, so we know it is promising it was a fairly large-scale study, nice effects in both concepts and calculation. One reason we chose to highlight this a little bit as a promising thing is given the interest in Tier 2 interventions with the re-authorized Individuals with Disabilities Act (IDEA). This was a study of taking first graders using a valid screening measure, those kids who were lacking in foundational skills. Even the basic kindergarten pre-k foundations were weak. Teachers' assistants worked with them in a very structured intense way a half hour a week every other day and you got nice effects with these kids. So, it is a sign that at least we have one approach and some promise as a possible model or prototype for Tier 2 interventions.

Learning disabilities basically has the same themes as low achieving and has a larger number of studies. Here probably the two unique things were 1) more of an influx from cognitive psychology in some of the newer studies and 2) several of the best designed and more interesting studies with significant findings really looked at how to build this quick retrieval of arithmetic facts and combinations both in multiplication and addition through kind of intense multi-faceted instruction. We know that that is kind a hallmark of a kid with a math disability. They do not have this retrieval and it is not easy to build. So, they are this kind of nice models and prototypes of practices that could work for this population.

The last thing was with regards to a strategic move from concrete objects to visual representations and then going back and forth between visual representations and the abstract equations or math notation. There was a promise when it was carefully orchestrated and sequenced for teachers. The kids with learning disabilities did a lot better in terms of learning algebra in middle school. So, again this is a promising direction or theme.

Technology, Doug Clements was the lead here. He did two things, he and the group. One was to look at graphing calculators, because everybody is interested in if they are they good or bad. There was no evidence of harmful effects. The studies are limited. They are older studies. And there seems to be some facilitative effects on word problems. So, there was no harm being done by use of calculators which were much more primitive compared to the ones in use now.

Doug also did a meta-analysis of the meta-analyses of technology use. And it is very, very complex. There is no clear clean message there. So, it depends on the software and goals. And that really is something we cannot do a crisp summary of right now, other than to say there is no clear signal that emanates from that.

A couple of cross cutting themes, when we are trying to put this together, and one thing that we found by sticking to these rigorous standards, and I think everybody learned the amount of work is phenomenal in the analyses and reads and re-reads and cross validation in responses to peer reviews in digging into studies.

The other thing I think the group did experience though is this liberating sense that it is not a judgment or a personal professional decision as to which studies and which findings are included and excluded. There are external objective criteria that are worked out both with the researchers at Abt in Cambridge and our group around the country.

So, with that a lot of these interventions are multi-faceted and some of these labels we used like Team Assisted Instruction has formative assessment, it has explicit instruction from the teacher, and it has cooperative learning. There is a nice incentive motivational structure for kids. Almost all the explicit things have example sequencing. It could be that other example sequencing, careful sequences, that make mathematical sense and that have kids practice the kinds of problems they really need to practice and have guided practice with those, may be the critical key. There are a lot of themes to unpack. And with that, I think I will just quickly repeat the only three robust findings we have are in formative assessment. There is enough evidence that I think we can recommend that it be used with the caveats, which are no more so the caveats here than the reading findings of Connie Jewel and others about success in reading in third grade. It is enough to get people knowing this is likely to be a real good direction.

For part of the day, having some type of explicit instruction for some of their math for the lower third of students. It may be better if that is part of their math instruction, but some explicit instruction or guidance for both your students with math disabilities and low achieving. Also, there is some promise to the serious use of real world problem solving on an array of mathematical tasks, but it has not panned out in terms of traditional achievement.

And now we will see if there are comments from my colleagues.

DR. FAULKNER: Anyone else from the task group want to say anything?

DR. LOVELESS: I just want to correct that TAI stands for Team Assisted Individualization.

DR. FAULKNER: Thank you. Okay. Bert.

DR. FRISTEDT: A comment on real world problems. Joan's essay on real world problems acknowledges that there is a definitional problem of what real world means and it treats that quite nicely. But, I did want to bring it up. If it is interpreted in the broadest sense, I think it is motherhood and apple pie that real world problems belong in part of mathematics. But then there are other words that are used, too, story problems and word problems. Are they all synonyms or are there distinctions? And unfortunately the researchers themselves seem to use these terms and other terms somewhat differently. That is one comment I wanted to make.

The second comment, I am actually quite disappointed that our group did not get into sort of what might be called the details that everyday teachers confront in the classroom about which things should come first. Now, of course they are guided much by the textbook, but then we do not have much to say to the textbooks other than the obvious things, organize things in a logical way. But, it is not true that there is just one logical way to organize materials, and some might be significantly better than others, and we did not touch on that. And I must say I am disappointed about that.

Russell just mentioned a series of

well thought out examples. If they are well thought out, I would like to see examples of them, because other people might come up with a series of examples that are not so well thought out. It would be nice to really get some handle on what it means to organize well.

And I have one other comment back to the real world problem. I think a big issue that is not really confronted in the real world problem and maybe there is just not enough evidence, is to what extent should a topic be introduced by a real world problem? For example, suppose you are going to introduce simple algebraic expressions line (2x + 7 = y). Should the first introduction of that be, say, conversion from Fahrenheit to Celsius temperature, or should you hold off on a specific instance of it until you have nailed down the concept? And I do not think any of the research focused on that particular issue, but certainly textbook writers do have views on that.

DR. FAULKNER: Doug.

DR. GERSTEN: I am just going to add one thing before Doug goes. Joan and I, for our two sets of studies, will ask mathematicians to take a look at the example sequences and the mathematical correctness or what mathematicians and math educators think about those. That is something we will do in the very near future.

DR. CLEMENTS: Now, just a comment because this came up in the Learning Processes group, as well as a new addition to our group on this concrete to visual. I would just recommend that we are very careful about feeding into what is a long standing Buehnarian type of idea that learning precedes from concrete, to visual, to abstract. When I think that if visual is meant as needing visual supports, it is very different from the creation of visual mental structures or visualization skills that kids have. Also the research has not been kind to Buehner's original sequences of absolute kinds of steps. I just think that we need to be careful about the nuances of supporting that when there is not a lot of research that indeed abstract thinking is non-existent until you go through the other two stages.

DR. FAULKNER: Okay, let me go to questions. Vern.

MR. WILLIAMS: I just have one thing to say about the real world problem thing. That we found evidently that there was some small positive effect on some things. What I have been hearing as a teacher over the last five years is that a child will not understand a topic unless it is introduced through a real world problem. I want us to make sure that there is no research to support that, loud and clear.

DR. FAULKNER: Thank you, Vern. Sandy.

DR. STOTSKY: Actually the recent comments of the Panel members have helped to address to some extent this question about what you actually found in terms of the problem with real world problem solving. I think that it would be helpful if there were a clarification that it may be promising, but the issue has been that there is no evidence to support the focus or the emphasis that has been placed on it. That is what I think the evidence is suggesting, that no one is suggesting it should be excluded from further use in the curriculum or explored in research, but that the emphasis that has been placed on it has been misplaced and cannot be justified. That, I think, is what needs to be clarified for educators, particularly, in the field.

DR. GERSTEN: We tried to be clear that if your goal is raising achievement using the more typical word problems, calculations, things like shading in parts of a fraction, you know, which one is 7/8ths or 9/8ths, that real world problems is not going to get you there. It is not going to lead to growth there. That is one of our two findings.

DR. STOTSKY: Yes, but you said that was for assessment, I am just saying in general this has been emphasized for instruction in textbooks. So, I am just suggesting that if the research evidence does not support a robust use of this everywhere, that somehow I think there should be an indication that its use should be simply more limited than it may be.

DR. GERSTEN: We can flesh that out.

DR. STOTSKY: Bob.

DR. SIEGLER: Yes, I was a little concerned about grouping together all computer software and technology under the same heading. Asking is computer software good is like asking are chairs good. There are good chairs, there are bad chairs, and there are chairs that are in between. And there are some computer software programs such as the Algebra Tutor that have received the What Works Clearinghouse (WWC) imprimatur for being effective. I think just saying we do not have enough research to meet this criterion of three or six, I don't know how many studies have to document it, may not be enough. Maybe when you realize how different these programs are, maybe it is not the right criterion to apply to this particular area. I am sure Doug, who did a very nice report on this, has an opinion on this issue.

DR. CLEMENTS: No, I agree completely that it was only time constraints and the notion that there are various types of software, and then within those various types varying substantiations of those types and various software. But we were just basically saying there are some things that are very promising but you have to guide us in the details, here and you really have to go to the report. We did not put it in this presentation, just due to time constraints, so, well said.

DR. FAULKNER: Other questions or comments? Dan.

DR. BERCH: Yes, I want to ask you something. In several places you mentioned that explicit needs to be unpacked in a better way. I also wonder whether you can make, or maybe you will be more explicit about the distinctions between direct instruction, explicit instruction, guided inquiry, et cetera. Where there seems to be a message here, at the very least, that while you can make distinctions here as well as distinctions between more of the extremes of, let's say, certain kinds of direct instruction and unguided inquiry, that in many cases there is a more eclectic use of these sorts of things. Perhaps, the notion of always pitting these against one another, at least in practice and arguing is it this one or that one, won’t get us much. In the words of Chase, over thirty years ago in a classic paper, "You cannot play twenty questions with nature and win." I am wondering whether you are going to bring that out more in terms of a particular kind of recommendation?

DR. LOVELESS: Let me take a stab at that. I reviewed the literature on direct instruction, teacher-directed instruction is the way we defined it. Now, there is this term Direct Instruction with a capital D, capital I, which is the Engelmann/Carnine scripted model form of instruction. Teacher-directed instruction refers generally speaking to a larger pool of interventions where the teacher is the center of instruction or the dominant actor in the instruction. To be honest with you, and this is one of the recommendations that we make, is that we do have more research that looks at different forms of teacher-directed instruction, typically in the research that we looked at. Again we applied these rigorous qualifications to the research to screen out research that we could not rely on. The research that we looked at typically the teacher-directed group was the control group. It’s also sometimes just called the traditional instruction group and it was not explicitly described or described with as much specificity as the experimental group was, and that occurs a lot in the scientific literature in education, but it is unfortunate. Because we do know that teacher-directed instruction is a very popular form in classrooms, lots of teachers use it, but we 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.

So, one of our main recommendations is that we have more experimental studies that look at different kinds of teacher-directed instruction so we can learn more about it.

DR. FAULKNER: Liping.

DR. MA: Now, we saw yesterday afternoon you have this teacher survey about student regimens for algebra. They mentioned word problems. I was wondering do we have any evidence between real world problems and word problems? Did I make it clear? How does serious work on real world problems help them learn word problems in algebra?

DR. GERSTEN: The analysis does not look at algebra, but what Joan found was no, there is no significant impact on the more typical word problems from these real world experiences. So, that does not help student learn. Many of the studies on low achieving and learning disabilities worked on more traditional word problems, the kind of ones that we encounter all the time, and there are ways to help students learn how to do those through these various things called explicit instruction. But the real world, to get back to Sandy's point, no.

DR. MA: So, was this made clear in your statement?

DR. GERSTEN: It will be, yes, yes. We are drafting the summary. It will be.

DR. MA: Thank you.

DR. GERSTEN: Yes, yes.

DR. FAULKNER: Skip.

DR. FENNELL: I am wondering, Russell, if there can be some way for you to say something like the following perhaps, not using these words. I suspect that there is not a person in this room who would not acknowledge that problem solving, the ability to solve problems, is important for anybody learning mathematics at any level. And we banter around phrases like more typical word problems, word problems, real world problems and the like. And the issue is, certainly problem solving and the ability to solve problems is important, but trying to sort of ferret out the difference between a quote, unquote, "typical word problem and/or a real world problem," is part of the mess we have all found ourselves in. So, I do not know what to do. I do not know quite how you would say that. But I just heard you say that there is some evidence that says this comparison to real world versus typical, which questions in my head okay, what is typical? Because when a classroom teacher or a publisher or anybody else crafts a problem for a student to solve it is contrived by that person, for that moment, for that class, for that mathematics. And so how do we draw the distinction between something that is typical or real world, and are there gradations of interest, of difficulty, of relevance, of input?

DR. GERSTEN: This is an issue that Joan and all of us have been grappling with. Basically, and we will get input in the next several weeks from our guests who are in the audience and others. Most of the real world problems, just because of who did the studies, are the Jasper Woodbury work that the Branford Group did. And if you just operationally define word problems such as those that appear in the various state assessments and whatever, the Jasper Woodbury are really multi-stepped kind of problems that way. But that is something we are going to continue to probe. I do not think we can find the answer to it because of how they are operationally defined. And including taking a look and having Joan and other mathematicians look at Jasper Woodbury and see what they think of the mathematics there, because so many of them are that or related videos.

DR. FAULKNER: Vern.

MR. WILLIAMS: Skip, I would like to help answer your question this way. Real world problems at times, really do not concentrate purely on the mathematics. Now if you open up an algebra book and see a distance, rate, time problem, the student may not be motivated by distance, rate, time, but the problem will involve the concept that is attempted at that moment. But real world problems sometimes are put in for motivational purposes.

So, they get into these other topics that might be related to something that has nothing to do with the mathematics involved. And I think that could be a distinction. When we generally discuss word problems, we are thinking more of the typical algebraic, typical mixture distance, rate, time et cetera. In real world problems, we want to talk about kangaroos in India and I guess they are in Australia really, but see I am already confused.

DR. LOVELESS: How long are you talking about?

DR. FENNELL: I think that is a good point. And Vern as I think both you and I share the same frustration. Real world, whose world is it? I mean is it the world of the child? Is it the world of the teacher who crafted that thing about I guess it was kangaroos?

DR. FAULKNER: All right, we need to bring this one to an end. Is there a last burning question that has to be addressed, because we need to move to the next group?

Thank you very much, Russell and

team. Now, we go to the task group on teachers and teacher education and Deborah Ball chairs that.

VII. TASK GROUP 4 - TEACHERS

DR. BALL: All right, I am Deborah Ball and on my right I have Russ Whitehurst and Hung-Hsi Wu, and on my left Ray Simon, and not with us today is Jim Simons, and Jim Yun is our staff associate. We will be presenting sort of the synthesis of the work we have been doing on teachers, and as Larry said it also filters into teacher education.

We have been working across the questions that we are investigating that our fellow panelists are quite familiar with. We realized that it was important at this stage of our work to try to describe for you what we think is the logic of why our group has been investigating the questions we have. And we actually think this may be important to the overall story of the panel report. And we hope that in the coming weeks this narrative that I am about to explain or perhaps another one might help us as a Panel to be thinking what is it exactly that we’ve been doing as we tried to address the charge we had.

So, I am going to try to explain where we started in trying to express to you what we have been doing. We start with the assertion that comes through from research and I think also from plenty of other sources that teachers are crucial. Teachers make an enormous difference. And one way to see that is to look at studies that have examined the contribution made by the teacher to the sort of achievement gains of students. There are studies that show that a large portion of the variability in student achievement gains is due precisely to who the teacher is. You think of that as the teacher effect.

Unfortunately, in this research, that substantiates something that many people would already hold to be important. We have learned very little from these studies about exactly what it is that these teachers who are making these differences to students know or do specifically that makes them more effective than teachers who are less effective with students.

So that leads to a question that actually is the province of the previous two groups at least. We wanted to frame that for you to show a kind of gap as you think about the logic of our work. It seems that what you would want to know next then is what in fact do these teachers do, or what is it they know or something about them that would help explain why some teachers make greater gains for kids than others do.

So, we put IP there to refer to instructional practices, your task force, but it is also LP, I think learning processes, and I think that showed up to some extent this morning as Dave and his colleagues summarized their work. But that is clearly one of the crucial things to understand is what is it that teachers are doing when they make greater differences for students than others? But our group was not charged with looking at instruction, or at learning or the interplay of those two but rather more about teachers.

So, we still have a reasonable line of work to pursue given that essential finding that teachers make a difference. The way we organized our work is just a different way for you to understand our four questions. First of all, there are already strong hypotheses about teachers' mathematical knowledge and that lead us to want to investigate what evidence there is about the relationship of teacher content knowledge to student achievement, and how states' teacher assessments can rigorously measure this kind of content knowledge.

Second, it would make sense for us to learn as a Panel what is it that is known about how to train, recruit, retain, reward teachers in such a way that we have more teachers who can produce consistent achievement gains in students.

Finally, given the huge number of teachers there are, it made sense to us that we would examine whether there might not be a useful way to consider what we have been referring to partly as the scale problem. There is a huge need for teachers who know enough and are skillful enough to produce achievement gains in students. We also looked into whether there is anything known about the specialization of elementary or middle school teachers that might help to address this enormous need to have more teachers who can produce these kinds of achievement differences for kids.

So, think of that as a proposal for

one way to understand the teacher task group work and possibly more of what we have been up to as a Panel.

So, now I will just remind you without going into them that these were the four questions that we have reported on each time that we have made a public session that you have been reading and the work we have been doing. We looked at these four questions. They grow precisely from what I have just showed you on the previous slide.

Briefly, the methods that we used echo things that you have all been saying about the methods you used in your task groups. We had a few challenges possibly that differentiated our work from some of the other groups. But, essentially we followed a lot of the same procedures. We tried to identify the available scientific evidence for the questions that we were investigating. We figured out search terms that allowed us to search these different databases. We also looked manually based on recommendations we got from people, from testimony, from other things about what else might be out there that might meet our criteria and might supply evidence on our questions.

We then organized the evidence that we collected into categories based on study strength using the agreed upon criteria that we developed as a Panel. We noticed that of the four research questions that we have been investigating the strengths of the available evidence varied quite a bit. You will see the way that we have organized today's presentation reflects that variability and strength of evidence.

It is also the case, although this is now hidden behind several people's heads, that we gathered information from different sources depending on the question. We looked at what states' assessments looked like or what the PRAXIS exam is like and what it measures. Or for math specialists as you will see we did not find research on math specialists, but we were able to gain a lot of information about different things that go under the heading.

By the way, just as an aside, I do not know Valerie, sometimes it seems like we should not be referring to this as methodology, but maybe that is another discussion. It seems like it is the methods of our work, not a study of methods. But I was just wondering about that.

So the structure of what we want to present today is organized as follows. For each of the questions we are going to organize what we tell you today into what we think we know, based on what I just showed you about the criteria we used, what we would now say we do not know, and, additionally, what is not supported by the research that we found.

We also found that there were studies around certain aspects of our questions where there were consistent non-effects. And we think that it is useful to propose to the Panel that we consider those as things that we ought to stop saying. It is almost like yesterday we were saying myth bunking, myth debunking. But things that often get said to be known, but not only is there not knowledge, it is suggestive that there possibly isn’t substantial evidence to continue to claim that certain things are true. So, we want to put that out today to see what you think about this way of organizing our findings.

Overall, we then present some draft recommendations for policy and for research based on what we have been learning.

So, I am going to launch into the teacher content knowledge findings organized into those three categories.

So first, what we know about teacher content knowledge is very summarized now from what you have read in our report. One is that the signal across the studies we reviewed is that the teacher's mathematical knowledge is a positive factor in student achievement. Second, what we are calling proximal measures, that is tests of the relevant knowledge that teachers actually use to teach mathematics, show a stronger signal than more distal indicators like certification status. That makes sense, but it also shows up in the research that we reviewed.

So now I am going to turn to what we do not know. And I am sure the public would like to know exactly what teachers need to know to teach particular topics to particular students. That does not show up in the research that we reviewed. And we do not know very much about how teachers' mathematical knowledge affects instruction in student achievement. So, if you think about trying to fill in the dotted lines between teacher content knowledge as a predictor of student achievement and everything that happens in between, the research does not say very much about what aspects of instructional practice or student learning are particularly influenced by the nature of the teacher’s knowledge of mathematics.

What is not supported by research is the belief that elementary teachers who take more university math courses are more effective. If someone believes that, there is pretty substantial evidence that that is not supported by research. We do see, however, some sign of this for secondary school teachers. But the idea that more math is better for the effectiveness of elementary school teachers is not supported by research. Similarly, we 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.

So if you try to understand the

syntax of these sentences those are claims that people might make, and what we are saying is, these are not supported by the research we reviewed.

Okay, I am going to move on now to teacher education. And we defined teachers education here to include teacher preparation both conventional and alternative pathways, induction programs, and professional development. So we are using that word inclusively as an umbrella for all forms of professional training or education.

Okay, so what do we know? That is it. We actually do not have anything that we can claim that we know about teacher education from the research that we reviewed.

So, I am going to move on to say what we do not know. We do not know what features of teacher preparation or professional development produce changes in teachers' knowledge or in their students learning. There are many claims that are made about structure, collegiality, and content and so of the like, we cannot actually say that from the studies that are done. That does not say that they are not supported, but we just simply cannot find that out from the research we reviewed that met the criteria that we used.

What is not supported by research? Not supported by research is the belief that different pathways into teaching at entry produce differential effects in teacher effectiveness. So, so far the evidence does not show differences among these different pathways into teaching.

I think a disappointing area of our work was the last one. It was the one that you have not heard from us yet. There were an enormous number of studies that turned up into the search terms. And very, very few of them met the criteria that allowed us to make claims that would lead to the sorts of questions we had, like what produces what kinds of changes for teacher capacity? We were not able to determine that from the research.

I want to turn now to teacher incentives. Some teachers produce more learning achievement gains than others. What is known about what they know, how they are trained or what they can do, and what attracts people into the profession, into certain locations where we especially need teachers, or might reward them for producing achievement.

So, here are some of the things that we know from the research we reviewed. First of all we have the salary differential between teaching and other technical fields for which teachers who are qualified, who were educated well enough to be math teachers. The differential between those other fields they could enter and teaching is quite large. But interestingly it is not large at entry it is that it increases dramatically across the first ten years of someone's work life.

Second we know 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.

We also know that location-based pay, which is pay used to attract teachers to high need areas, where kids particularly are not getting highly qualified teachers. We see that location-based pay can keep or retain experienced teachers in such high need areas. We also see that performance pay for teachers can enhance student achievement. And in our report you see the more subtle aspects of these kinds of programs.

What do we not know? We do not actually know from this research how best to design these sorts of pay schemes in ways that would reliably enhance student achievement. We do not know for example, whether it is better to have these schemes be at the level of the individual teacher or the school. We do not know whether it is better if they are competitive or not. We also do not know what levels of compensation make a difference. And we would urge that knowing that would make quite a difference in pursuing this for policy purposes.

We do not know whether and how location based pay helps to attract teachers to high need areas. What I said a few moments ago is that we know something about retention of teachers in high need schools, but not necessarily as attractors.

I am going to move on now to math specialists. So, what do we know about math specialists? We know that this term we have continued for the moment to be using this term, but that the term is being used for at least three different models of types of specialization at the elementary level. One of the things we are going to recommend is that continuing to use this term to cover different models is actually obscuring a conversation that is probably worth having.

What we do not know is whether math specialist in any of these models using them at the district level in schools leads to greater gains in student achievement. There is just not research that would allow us to say that. What is not supported by research and I think this is important given what often gets said, and again a kind of thing that gets said but is not actually supported. It is not the case that most high performing countries use math specialists at the elementary level. So, that is an example of something that is not supported by the investigations we did.

I am going to move now to our preliminary recommendations based on the work of our task force. They will follow across these four areas.

First, we said that given what we are able to substantiate, given the evidence that we had, that we should at least say that teachers should be required at least to know the mathematics they are teaching. Certification and licensure examines should at least test well the content that teachers actually teach. We think that it is worthwhile at this point developing alternative pathways into teaching, exploring whether those can be used in ways that could make a difference. We think that it is worth pursuing alternative salary schemes, including differential pay for teachers of mathematics, pay based on location, performance. These should be pursued with appropriate investigation of the questions I raised a moment ago about what we do not yet know about those schemes.

We think that where there is a shortage of elementary school math teachers who have appropriate knowledge of mathematics for teaching and that math specialists could help to address this need. But we need, as I said a moment ago, to clarify terms. So, when we are using the word math specialists here we mean teachers who have the requisite knowledge that is needed to teach mathematics, who are responsible for teaching the bulk of mathematics in an elementary school. We do not know about pull out programs. We do not mean math coaches. We are talking about a kind of math teacher who has the kind of level of mathematical knowledge needed for the work. This is one way of handling the enormous scale problem of the numbers of teachers needed who would know math well enough to teach it.

One thing that we were 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 us to say that in general really do not exist. So, we have phrased this rather carefully to say, in areas where it is clear that the shortage of such teachers exist, this could be a useful strategy.

Now, recommendations for research, we have five. One is that it is quite clear to us, that we need further research to elaborate what mathematics teachers really do need to know to teach particular topics to particular students, particularly beyond what is in the curriculum. It seems rather simple minded to say teachers should know what they are teaching. And we believe that there is more to understand about what else about mathematics teachers need in order to effectively deploy it when they are teaching kids.

We think that we need better measures of teacher’s mathematical knowledge, that focus more squarely on what teachers actually use when they are teaching, instead of the kind of distal indicators such as certification or courses taken. We say this because we found the strongest signals when these sorts of measures were used. And were we to have more measures of that type we would have made greater gains in this area as a field.

We need to have studies that identify the specific features of teacher education, pre-service, induction, alternative, and professional development that actually have an impact demonstratively on teachers' effectiveness with kids. In the case of these studies that examined the differential effects that some teachers have compared to others, we need what might be referred to as epidemiological studies that would allow us to probe what is it that is distinguishing the teachers who are making a difference with kids? What are they doing or what do they know? Or how are they relating to kids? What is it that they are doing that would permit us to know something about what is happening to explain why some of them are producing more than others?

Finally we think that we would be well served by having studies of what grows from specializing more at the elementary level in teacher assignment? Including questions such, as is it practical to do this and does using such sorts of arrangements at the elementary level actually produce greater student achievement?

I am going to ask right now if my colleagues have things they would like to correct or add to this report.

[No Verbal Response]

DR. BALL: Okay, so we are ready

for questions.

DR. FAULKNER: Thank you, Deborah. Tom.

DR. LOVELESS: I have a question about teacher preparation of Algebra teachers, teachers that teach Algebra. In the Trends in Mathematics and Science Study (TIMSS) data, one of the things that is noted is that around the world most Algebra teachers, and I cannot remember those percentages, but it is quite high, it is at least two-thirds, have a bachelors degree in mathematics. In the United States most of our teachers of Algebra have a degree in math education, which is quite different. Is there research on the relative effectiveness of those two degrees?

DR. GERSTEN: The Schools and Staffing Survey and National Assessment on Education Progress (NAEP) and other sources collect information on that. And we can tell you the proportion of teachers who have degrees in mathematics or math education. But to the best of my knowledge there are no studies that examine the impact of those differences on kids. It could be that the new longitudinal, high school longitudinal study, which will start with ninth graders and focus on mathematics will allow an opportunity to examine that more carefully and it is possible with the current data bases.

DR. FAULKNER: Doug and then Valerie.

DR. CLEMENTS: You did not say very much today about what I still take as a useful differentiation between content knowledge and pedagogical content knowledge. Was there nothing that you looked at, I am thinking off the top of my head of Cognitively Guided Instruction (CGI) research and some of the more recent elaborations and delineations of the whole notion about this kind of pedagogical content knowledge that for instance the learning processes report would have talked about? I believe there are studies that were indicative. I am not sure if they meet the criteria because I have not read them for a long time, but did you run into those or dismiss those studies or not look at that kind of distinction?

DR. BALL: It depends how far you would go into what is called pedagogical content knowledge. Our group was not investigating all the knowledge that teachers need to teach. We were focused on content knowledge. So, studies that would have fallen inside that search term would have included studies that looked at the kind of more specialized knowledge of mathematics, which could be construed as one part of pedagogical content knowledge. Cognitively Guided Instruction (CGI) would not have cut in there because that was knowledge of student’s mathematics, and we were not looking at that. We were looking at teacher’s knowledge of mathematics. And there are some studies in that third group of our question where we looked at certification and course taking and then testing.

In the testing group we have some studies that look at that more, sort of subtly closer to the knowledge used by teachers, which I think fits your questions. We did look at those.

DR. FAULKNER: Valerie.

DR. REYNA: I was waiting for my cue. Valerie Reyna. I was thinking as you were speaking about a kind of path analysis or cause analysis. For example, we 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. I can imagine if you were looking up key terms involving teaching, the first two terms in that causal path would not necessarily appear as a relationship. But if you have the resources to do or you may want to set this as a question for future research is the differential effect of course taking, like university math courses, on relevant knowledge, and then working through that as an intermediate step toward student achievement.

DR. FAULKNER: Sandy.

DR. BALL: Did you have any evidence on that?

DR. FAULKNER: I am sorry, Deb.

DR. BALL: Was there any evidence on the effective math courses on relevant knowledge?

DR. FAULKNER: I am really eager, I did not want them to answer.

DR. BALL: I mean the question of what we are partly saying is that the question of what is relevant knowledge is one of the constructs in the field that has been only in a very limited way unpacked. So, what we are left with is studies that look directly from course taking to student achievement. We do not even have things that look very carefully at the relationship between content knowledge and instructional effectiveness, which is where we should link with instructional practices. So, the notion that one could look from course taking to relevant mathematical knowledge would be wonderful, but that would require answers to question one up there which we were saying we did not find in the literature at least very emergently only.

DR. REYNA: Yes. And I would also add just quickly, 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. Are any of those studies perhaps relevant to teachers?

DR. WHITEHURST: Yes, and we have cited some of that work. I mean 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. A student who is a sophomore or junior in college thinking about being a teacher or thinking about being an engineer and having the mathematical skills to be a math teacher or engineer, when we look at differential payoffs that will affect decisions. Salary differentials make less of a difference when someone is committed to the teaching profession, and interestingly more for males than for females. So, men are more likely to change jobs and move to other locations or drop out of teaching for salary reasons than women are. And that is interesting. It also presents a conundrum for policy makers in that the amount of money that it would take to generate men is substantially lower than it would need to be for women to achieve an overall impact.

DR. FAULKNER: Sandy.

DR. STOTSKY: Just a quick question. I understand that you would not have had time to look at every single body of research. I was just curious given the importance of student teaching as part of teacher preparation. Is there a body of research to refer to on that? Is there something that could be looked into eventually? I just do not know what the extent of that work would be on something that has been such an important part of teacher preparation.

DR. BALL: So, that fits I think well with our comment about number three and our blank slide which we showed you earlier. We were not able to uncover research that showed differential effects of particular features of say in this case, pre-service teacher education. We were not able to uncover studies that showed that particular features had an impact on a teachers’ ability to produce achievement in students. There were no studies of that kind.

DR. STOTSKY: Okay. There was nothing specific to student teaching per se?

DR. BALL: Right. Or other features for that matter. But student teaching as you say is one of the key features one might choose to study.

DR. STOTSKY: Thank you.

DR. FAULKNER: Dan. This will be the last question.

DR. BERCH: Yes, thank you. Dan Berch. At first I thought I understood your distinction between the categories of what we do not know and what is not supported by research. While I cannot remember the specific examples, one of the first examples seemed to suggest that the results in the latter category, the results were inconsistent with the claims being made. And yet, in another example it seemed like the results neither supported nor refuted the claim. So, am I correct that when you say what is not supported by research you are suggesting evidence that is inconsistent with, not evidence that is not clear cut one way or the other?

DR. BALL: The distinction we were experimenting with is between what we do not know, where there just is not research, like what Sandra just asked, from a case where something gets said an awful lot about professional education or teachers or something like that and yet, in the studies either there is a mixed signal or there is counter evidence. So, both could fall into the, what is not supported by research. I am confusing you further?

DR. BERCH: Well, I do not know. It depends on how others react to that. To me that statement is a little confusing to say what is not supported by research.

DR. BALL: Well, it is not supported.

DR. BERCH: No, but it may be if there is mixed evidence, again in some cases if it not supported it may mean that the evidence is inconsistent with it. In other cases it is neither supported nor refuted.

DR. BALL: Well, let us look more closely as we experiment whether these were useful categories. Today we were trying, -- we worked yesterday to try to organize our work into these three. And we noticed that in some of the other task groups this could be useful to like in Instructional Practices. So, I think it is something for us to talk more about as a Panel. Treat it as an experiment today and we can look at it more carefully.

DR. BERCH: Okay. No, it is

interesting.

DR. BALL: No, I think it is a good question that you are asking.

DR. WU: I think that what Dan wants to say is that when you say something is not supported by research you want it to mean it is refuted by research? Is that what you want to say?

DR. BERCH: Well, I am just wondering if that is what the implication is and it does not seem like that is consistently the case there.

DR. WU: That is not our implication.

DR. FAULKNER: What I understand it to mean is that research exists, but does not confirm the statement that is about to follow.

DR. LOVELESS: If we find something that is refuted by research and those findings are consistent, why don't we not put that under the, we know category?

DR. FAULKNER: Okay. Well, we are going to have to talk about this later. All right thank you Deborah and team, we appreciate it.

Okay, we are going to the last presentation, which is the Task Group on Assessment. All right, the Task Group on Assessment is one that was constituted later in the process and got a start in the late spring. It is not at the same stage of maturity as the other task groups. Our Vice Chair of the panel, Camilla Benbow, chairs it.

VIII. TASK GROUP 5 - ASSESSMENT

DR. BENBOW: As they said, we got started rather late. We started in April and these are the members of our task group. And we have been working very, very hard over the summer. Now, just let me give you a little bit of background before we get to our sort of tentative findings.

I think, as most people know assessment is used in a variety of ways. It could be used to shape the content and format of instruction. It can be used to adjust educational experiences to meet the needs of individual students. Assessment can be used for selection and of course, for evaluating student and school performance.

When we looked at assessment we really focused in on the last bullet, evaluating student and school performance. And this I would guess primarily because of the impact of the No Child Left Behind (NCLB). Assessment is a huge part of that NCLB, where basically we are using tests to hold students and schools accountable for performance. In particular we use the National Assessment of Educational Progress (NAEP) and the state tests. State tests are designed to determine student proficiency in certain areas and all schools are required as part of the No Child Left Behind Act to also participate in the National Assessment of Educational Progress (NAEP).

State tests and the National Assessment of Educational Progress (NAEP) are such high stakes tests and they are used in determining for example whether a school makes AYP, Adequate Yearly Progress, and so on, and whether there could be consequences to schools as a result of their performance. We decided we needed to take a hard look at the National Assessment of Educational Progress (NAEP) and state tests. And we focused our work on those two tests. The kinds of questions that we pursued were about the National Assessment on

Educational Progress (NAEP) and state tests are, are they appropriate? Do they measure what is intended? We hope they are not biased. Are the conclusions drawn from test results justified? And basic issues of measurement quality. Also in terms of the National Assessment of Educational Progress (NAEP) and state tests we want to determine do they measure what is deemed important for children to master.

Our methodology is quite different than the other task groups. We did not look at all of the National Assessment of Educational Progress (NAEP) tests. We just looked at the main NAEP test for the fourth and eighth grades. We did not look at long-term trends. In terms of state tests we could not look at all fifty state tests so we focused our efforts on these states. They are supposed to be representative of these testing practices in this nation.

Here were the foundations for our report and for the conclusions that we make. Lucky for us there was an on-going Validity Study of the National Assessment on Educational Progress (NAEP) mathematics assessment in grades 4 and 8. There is a report coming out from the NAEP validity study panel. We were very fortunate that we were able to look at the results of the NAEP validity study, however that study is embargoed. It should be released in a few weeks. So, we cannot quote from that study, but we can say that we drew heavily from it. Also the initial report was drafted by the National Center for Education Statistics (NCES), which issued a response to the validity study. We were able to look at that, but again that is also embargoed. But we did not just limit ourselves to the validity study or the National Center for Education Statistics (NCES) response. We also conducted our own search of the literature with the help of Abt Associates.

We also collected additional information. With the help of IDA/STPI, we collected technical information from each of the state’s websites in grades 3 to 8 on the following issues -- the framework, procedures and release items.

Because of the approval process and so on, we were not allowed to conduct a survey to collect the kinds of information that we might have wanted. So, we were limited to information that was available in publications, on websites, and so on. So, again what we found as we walked down the road was that we could not answer all of the questions that we wanted to answer because we could not get access to the information.

We also did do a case study analysis of released items in grades 4 and 8 for NAEP and the state tests.

There are two main recommendations and I am going to have Skip and Tom talk about the first one and I will try to portray what we found in the second one.

But the two main recommendations are that the National Assessment of Educational Progress (NAEP) and state tests must focus on the mathematics that students should learn. The Conceptual Knowledge and Skills group has talked about what that knowledge should be and we believe that they should be in some ways aligned, and with scores on this critical content reported and tracked over time.

The second one is that states and

the National Assessment of Educational Progress (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-constant relevant sources of variance in performance is minimized.

I'll move on the first one and turn it over to Skip or Tom.

DR. LOVELESS: First of all to look at the NAEP which stands for the National Assessment of Educational Progress, it is also known as the National Report Card, we know that at some point over the next five or six years the NAEP framework in mathematics will be revised and a new framework will be adopted. There was an initial framework adopted in 1990, it has been revised twice.

So, what we want to do is offer some principles for revision and reorganization of the National Assessment of Educational Progress (NAEP). And let me state that in advance these are preliminary principles. We are still debating within our task group exactly what the correct wording would be.

The first one is to disaggregate numbers. There are five strands in NAEP and the first strand is called number or number sense. The first principle is to disaggregate number into two separate areas. The first area would be looking at wholes and integers, and then (B), looking at fractions, decimals and percents. Assessing those two are clusters of skills and knowledge in mathematics. The rationale for this is the foundations that were laid out earlier that you heard today from our Conceptual and Knowledge and Skills group.

The second rationale is that fractions are currently under-represented on the National Assessment of Educational Progress (NAEP). There are several estimates of the total percentage of NAEP items. But, the eighth grade, people who have examined NAEP and categorized items, formulate less than 20 percent of the items address fractions and decimals.

And then finally the rationale for this is that on the National Assessment of Educational Progress (NAEP) the scores are reported for strands so that we can monitor the progress, national progress in different areas. Currently, we monitor national progress at the eighth grade in the strand called number. We think it would be better for us to be able to monitor progress let us say at the fourth grade with whole numbers and at the eighth grade on such ideas as integers. And by the way, let me add our idea was the number strand for instance that category (A) would emphasize whole numbers at fourth grade and then on the eighth grade that particular strand would be emphasizing integers.

The same thing with the number strand (B), fractions, decimals and percents, we would expect that at fourth grade to be emphasizing more elementary ideas of fractions, decimals and percents, as opposed to the eighth grade version. So, the rationale again is to produce a score to track progress over time.

The second principle is to combine measurement with geometry. This would make the fourth and eighth grade NAEP consistent with the twelfth grade NAEP. The fourth and eighth grade NAEP currently has these as two separate strands. By combining them we believe we can increase the complexity of measurement items, and the NAEP has found that the measurement items have problems in terms of being low in rigor.

Next slide. Principle 3 addresses Algebra. And we are still discussing this. Many people who have reviewed NAEP have found problems with patterns, especially mathematicians. A lot of mathematicians believe the pattern items are over represented on the NAEP and that they are also of poor quality. We would like to make them more mathematical when they are used. But then looking at the second point under (B), there is this question of whether K through 4, -- fourth grade Algebra really is Algebra? If it is dominated by patterns and if those pattern problems are not mathematical in nature, then we ask the question is it really Algebra? So, we want to take a good hard look at that.

The rationale again is the definition of Algebra by our Conceptual Knowledge and Skills (CKS) group and then also the criticism of pattern problems as being non-mathematical. Skip, do you have anything to add to that?

DR. FENNELL: I think you did very well. Tom.

DR. LOVELESS: Okay.

DR. BENBOW: Let me just add that state assessments are heavily influenced by the National Assessment of Educational Progress (NAEP), so when you make changes or you recommend principles for the revision of the NAEP, indirectly you are also making recommendations for state tests.

Now, let me go on to the second part of our presentation in our work and this is looking at quality control issues. And so the first part was really "what do we measure?" And the second part was "how well are we measuring whatever it is that we are measuring?"

I think one of the things that concerned us was this issue of non-construct relevant variance, and contamination. And I think contamination can come from two sources. It could be from verbiage, which is maybe unnecessary, excessive, unfamiliar, or there could be other things that could contaminate a test item and one example is confusing visual displays.

And when we talk about contamination and non-construct relevant variance what we mean is that test scores, may be determined by things other than the mathematical skills that we are trying to measure. So, it may be a less robust measure of mathematics and maybe measuring reading ability for example. I mean minimized to the extent that it is measuring reading ability as compared to mathematics.

So, one of the things that we really drilled down deeply about was excessive verbiage, because it was felt that excessive verbiage can attenuate the performance of some groups, and so we really looked at items to see if this was a problem in the NAEP or state assessments. We particularly looked at state items. And it was a case study analysis, but we did find many instances of test items with problems of this type. So, this was an issue we want to bring forth is that excessive verbiage seems to be a problem.

The other thing that we looked at was situated mathematics problems or real world or word problems, and because of the excessive verbiage concern we came to these following recommendations. If you are going to be using real world or word problems, these test items should meet the following conditions. We actually felt that it was important to have word problems. And when you do have word problems they should be focused on deciding what mathematical knowledge and skills to draw on.

We felt that language needs to be concrete and serve to clarify mathematical relationships in the problem. Of course the knowledge has to have been taught. The items need to be well written. And of course we need to have enough items and depth to address the entire range of student ability.

Another kind of recommendation as we looked at the test items and we looked at the literature on measurement and quality and item quality, is that we really felt that scientific and logical evidence, as well as content expertise, needs to guide the test design. We also felt that item content needs to be carefully examined in order to understand performance.

So, as a result, one of the things that we looked at is that sometimes, perhaps the communication or what is intended or what those people who develop frameworks intend, sometimes are not played out in a specific item. We felt that there needs to be much more detailed item specifications coming from those who designed the test to the ones who actually carry it out and construct items.

And one of the things that we felt very disappointed about was, I guess almost every group who has reported on has noted, is the lack of research on high level analysis on the design of mathematics items. There really is not much research out there to talk about those issues.

Here are some quality control recommendations. We did not just want to criticize. We wanted to provide some recommendations for quality control. I am not going to run through all of these, because there is a bunch of them, but here are some suggestions for how we can have better quality control.

We also looked at proficiency standards and how do states and the NAEP set proficiency standards? There were a lot of different ways in which states or the National

Assessment of Educational Progress (NAEP) went about in terms of deciding when is a student proficient and when is a student not. We felt that the methods should follow the best scientific practice. And our review of the literature came down that the modified Angoff probably has the most support. So, we recommend that states and NAEP use this procedure as setting proficiency standards.

If many of you perhaps were reading the report in June where you could see that states and the National Assessment of Educational Progress (NAEP) have very, very different criteria for what is proficient. We felt that perhaps we ought to draw on international data on student performance to help in that process.

Another recommendation that we are discussing, and again all of these recommendations we are discussing, is that the National Assessment of Educational Progress (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. We believe that they should assess the Algebra objective endorsed by the National Math Panel.

And we have some more work to do. Calculators, what is the role of calculators in assessment? And also different item types, for example, multiple choice versus constructed response. And we have just not completed that work. But we did find that calculators, in the early grades were not used very frequently. And that is it. Does anybody have anything else to add?

DR. FAULKNER: Thank you, Camilla. You have gone an impressive distance in a short time and I appreciate all the effort that has gone into making that happen. Questions or comments? Valerie.

DR. REYNA: You mentioned assessment of Algebra and some of those slides went by really quickly, but I do not know if you mentioned preparation for Algebra and what it would be. Obviously it would be very useful given the discussions we have had today to have an appropriate instrument that would assess, particularly in a diagnostic way, adequate preparation for Algebra. Would that maybe form a recommendation or possible area of focus?

DR. FENNELL: There are at least two standardized tests that I am aware of that do that now. What we might do is look at suitability of those compared to what Conceptual Knowledge and Skills has suggested.

DR. FAULKNER: Any other questions or comments?

DR. FRISTEDT: If I can respond to Valerie's question also. I think some of the things that have been said here are actually related to assessing preparation for Algebra. Some of Tom and Skip's concerns about NAEP really reflect a feeling to various degrees among us, we are not all in exact agreement, that the National Assessment of Educational Progress NAEP fails to test the very things that were identified in earlier discussion here as critical for Algebra. It does not take it off the table, but the weight of it is just at the wrong place.

DR. FAULKNER: Anything else?

[No Verbal Response]

DR. FAULKNER: All right, well then that takes care of it and I want to thank the Task Group on Assessment.

That brings this meeting of the

National Mathematics Advisory Panel to a close. I would like to thank the public for attending. And I would also like to announce that the next National Math Panel meeting will be held at Arizona State University in Phoenix, October 23rd and 24th. And I would also like to remind the Math Panel members, those who are still here, of the importance of September 21st. Thank you.

[Whereupon, at 12:25 p.m. the meeting was adjourned.]