

**Testimony before the Minnesota House Education Policy Committee  
January 24, 2005**

**Comments on the math test specifications, January 25, 2005, by  
Lawrence Gray, Head, School of Mathematics, University of Minnesota**

**Brief history:** Two years ago, Commissioner Yecke assembled a group of 40 people to write the new Minnesota State Math Standards. The process was very intense, with widely varying points of view. The final product was a carefully worked out balance between conceptual and computational approaches to K-12 math, and there was surprisingly broad acceptance from all sides.

Although there was plenty of input from the 40 people on the standards committee, the final version of the standards was the work of 6 people, who spent several long days at the MDE, working under an extremely tight deadline. I was one of those 6 people, along with Ellen Delaney, Barb Stofflet, Lesa Covington Clarkson, Nancy Nutting, and Ellen Hoerle. Three of these (Ellen Delaney, Barb, and Nancy) are current or former K-12 teachers, two of them (Lesa and I) are university professors, and Ellen Hoerle represented parents. All of us expressed a willingness to participate in the next critical stage, which was the writing of the test specs. We realized that in spite of our efforts to make the standards as clear as possible, our familiarity with the intent behind those standards would be essential in the process of producing well-aligned test specifications. I was under the clear impression that a commitment was made to me and the others that we would all be consulted during that process.

In the latter half of September, 2004, a draft of proposed test specifications for the math MCA's was made public. This was the first time that I had any chance to see what had been going on. In particular, I was not consulted during the production of that draft. I was extremely concerned when it became apparent to me that the proposed test specifications were significantly out of alignment with the standards that we had written 18 months earlier (more specifics about the alignment later).

Within a week or two, I met with Commissioner Seagren and staff from the MDE. In my view, the meeting did not go well. But I had a followup meeting with Commissioner Seagren in which she called for patience while the period of public feedback was going on. In the end, this feedback period was much longer than originally announced. It appears that in the interim, staff from MDE and volunteers from K-12 (including Ellen Delaney and Barb Stofflet) were hard at work producing a revision of the proposed test specifications.

On January 12, 2005, I had a first chance to look at the latest revision. I was pleasantly surprised to see that significant improvements had been made, and that the Commissioner seemed genuinely anxious that the test specifications be in alignment with the standards. Since then, MDE has been very cooperative with me, and I have had a chance to study the revised drafts. This resulted in a large number of questions, which I have discussed at length with MDE staff and with Ellen Delaney. There is much left to be done, but it

appears to me that MDE is working in good faith to bring about alignment where possible (I will say more about this later). Promises have been made to address several of my current concerns, and reasonable explanations were given regarding other concerns that I have. I am optimistic about the final outcome, but cannot be definitive in my level of support until I see the end result.

**Alignment and related issues:** The lack of alignment in the September draft of the test specifications consisted of two different problems. First, the percentages assigned to different parts of the standards were skewed away from the more computational benchmarks, in favor of the more conceptual benchmarks. These percentages determine the number of items on the test that are associated with each benchmark. The shift was particularly dramatic in the Algebra strand of the Grades 9-11 test specifications. That strand is divided into two substrands, designated as III.A and III.B. The first substrand is primarily conceptual, and it contains 5 benchmarks. The second substrand is primarily computational, and it contains 12 benchmarks. In the September draft of the test specifications, the split between these two substrands was 70% for III.A and 30% for III.B. Since substrand III.B contains almost all the algebra items that are considered critical on many math placement exams for college, I was alarmed. In the most recent draft, the percentages have been changed to 40% for III.A and 60% for III.B. This still represented a slight shift in favor of III.A, but it is one that I can accept. There were many other places in the September version of the test specifications where I had similar concerns. I also had concerns about the balance between “calculator problems” and “non-calculator problems” in the specifications. But in the latest draft, the percentages seem to be in better alignment with the standards throughout, and I believe the calculator issue will also be satisfactorily resolved, although there are some tricky issues with the calculator in Grades 9-11 that I need to discuss with the MDE staff. Also, I still need to see exactly how the percentages for substrands will be translated into actual test item counts for individual benchmarks, before I can be comfortable with this issue.

The second source of misalignment in the September version was in the wording of test specifications for individual benchmarks. I noticed many cases where a benchmark was rewritten, or “content limits” were added, that significantly weakened the benchmark, relative to its original intent. In a few cases, a benchmark was weakened to the point where it no longer made sense. I won’t give examples here, because they would involve us too much in technical details. I have had a chance to discuss all of those individual items in detail with MDE staff and Ellen Delaney, and we have at least come to some sort of understanding in each case (more on this later), although it will be very important for me to see the final version before I will feel comfortable on this issue.

There was another serious problem with the September proposal, which had nothing to do with alignment. The September draft contained a large number of errors, ranging from simple typos to serious mathematical blunders. Many of these errors are not yet corrected, and other errors have crept into the revisions. I have not had a chance to discuss all of these in detail with the MDE staff, but my understanding is that they will welcome my help here, since these errors have little to do with opinions about appropriate content or pedagogy. In this case, I hope that deadlines do not prevent us

from working together on this aspect of the test specifications. While these types of errors are not always fatal, they can lead to unnecessary difficulties, and in any case, they give a bad overall impression when the test specifications are reviewed by outsiders.

**Why alignment can be tricky:** I mentioned earlier that the MDE staff and I had come to an understanding on certain benchmark specifications. This means that in some cases, they were able to give me reasonable explanations why a benchmark needed to be weakened. These have to do with the fact that some of the prescriptions in the math standards were identified by a large number of K-12 teachers as being inappropriate for testing at the grade level where they appear. In most cases, it was deemed that they came a year too soon for testing purposes, especially in the context of “No Child Left Behind”, and so they were moved to the next grade level in the test specifications. Fortunately, I did not find any significant issues of this sort in the most recent version of the Grades 9-11 test specifications. I had a few concerns in each of Grades 3, 4, 5, and 6. It appears that I have reached agreement with the MDE staff for those grades, although I reserve final judgment until the end result appears.

We had the most difficulties with Grades 7 and 8. The problem is compounded here by the fact that the variation in content in 7<sup>th</sup> and 8<sup>th</sup> grade math texts is much higher than at other grade levels. There are sharply divided opinions about timing in the introduction of various topics. These are not even partially resolved until 9<sup>th</sup> grade. When we wrote the standards for 7<sup>th</sup> and 8<sup>th</sup> grade, we tried to find a reasonable pathway, one that would provide guidance for appropriate middle school content. But when test specifications are written, one cannot ignore the wide variation that currently exists at this level. Some compromise is inevitable. I am not completely comfortable with this compromise, since I think there is a general weakness in many middle school math programs, but I am afraid that the writing of test specifications is not the appropriate place to try to address those weaknesses. So I believe that there is an outcome that we can all live with here, although as I have said before, I will need to see it before I make my final judgment.

There is also the issue of how different benchmarks are most appropriately assessed. Some of the benchmarks were primarily intended to be assessed in the classroom. Such benchmarks will necessarily be underrepresented on the MCA tests. Some benchmarks are best assessed with a calculator, others without a calculator, and others work both ways. I was pleased to see that the MDE is planning to require that 50% of the items on the 3<sup>rd</sup> and 4<sup>th</sup> grade tests be non-calculator items. The non-calculator items will include those that require knowledge of the basic arithmetic facts, with or without a real-world context. The calculator items will include those that require arithmetic with larger numbers. This seems to me to be a sensible approach, consistent with the original intent of the standards. It is not obvious to me that this approach was so clear in the September version of the test specifications.

**Some recommendations:** I believe that much of the tension during the past few months could have been greatly alleviated if I (or someone like me) had been consulted during the writing of the September version of the test specifications. I understand that a very large number of K-12 teachers were consulted, and that both Ellen Delaney and Barb Stofflet were heavily involved as well. I value the expertise that was represented, and the hard work that was expended. But university level mathematicians bring a different perspective that is vital to the whole process. The relationship between mathematics in K-12 and mathematics at the college level is critical. Furthermore, we have a number of university level mathematicians who have been deeply engaged in the process of writing standards for several years now. It was a mistake to leave them out when the test specifications were first drafted.

From the very beginning two years ago when the standards were written, unrealistic deadlines have been imposed. Writing standards and test specifications is an incredibly complex process. The standards were written in a very short time period, primarily by volunteers who did it in their spare time. Apparently, volunteers have also played a significant role in the writing of the test specifications, and we again find ourselves facing a tight deadline. Under the circumstances, amazing things have been achieved. But in spite of this, the quality has suffered.

Once the test specifications are completed, we have a chance to step back and look at the whole picture. In retrospect, it is clear that the standards would have been written differently if more had been known about the process of writing test specifications. The actual content in the standards would not have changed much (although a few individual items would have been moved around), but certain things would have been done differently. Here is one example. When we wrote the standards, we were under pressure to keep the length down. Therefore, we treated the topics in an accumulative fashion. That is, if a certain topic appeared, say in Grade 4, we did not repeat it in Grade 5 unless there was something substantially new to be added regarding that topic. We fully expected that the topic would be continued and further tested in Grade 5, and we explained this in the introduction. This feature caused some problems with the test specifications. For example, suppose a topic is only mentioned in a 4<sup>th</sup> grade benchmark, but it is deemed appropriate to first test the topic in 5<sup>th</sup> grade rather than in 4<sup>th</sup>, or to test it more completely in 5<sup>th</sup> grade than in 4<sup>th</sup> grade. The legal requirements of alignment between standards and test specifications make it difficult to put test items on this topic into the 5<sup>th</sup> grade test, because each test item at a given grade level must be explicitly associated with a benchmark at that same grade level. A 5<sup>th</sup> grade test item cannot refer back to a 4<sup>th</sup> grade benchmark.

There are many other such “loose ends” that need to be addressed. We should not wait until the next deadline (whatever that may be) before getting people together to make improvements. As soon as the test specifications are completed, we should start to work on a better version of both the standards and the test specifications.