

Additional Context to Citation from **Associate Commissioner Ann Smisco**

Ann Smisco presentation on June 3, 1999: The University of California in Irvine, CA hosted by the National Academy of Sciences involving educational presentation from officials of the Texas and Kentucky regarding their accountability testing programs.

ANN SMISCO EXTENDED EXCERPTS:

“By law the exit level test, of course, has to be highly reliable and valid so that every high school student has the same standard to meet unless, of course, we do actually change the high school graduation standard. And, we do that by informing seventh graders of their high school graduation requirement...

“Test items are written by our test contractor. We do this through a bidding process. Happens to be National Computer Systems right now. They subcontract with Harcourt Brace Educational Measurement for the item writing. Those items are reviewed by the contractor, first of all. And then by staff internally. Our curriculum and assessment staff to make sure that they match to the essential knowledge and skills and to make sure they’re appropriate for the grade level and for the Texas environment.

“And then we have an educator – so have a bunch of items that are possible items for a test. Then we have the first of a series of educator review committees that are representative of the state as a whole. Both ethnically and geographically. And they are grade level subject area specific. In other words, there’s a third grade reading committee. Third grade math committee and so on and so forth. And we try to make sure that the representation is there on every single committee.

“That group is asked four questions. Does the item match the objective it’s supposed to match? Is it appropriate: that is. Should students have learned this information by the end of X grade level? The adequacy of preparation: that is. In your district, did you teach this by the end of X grade? Do students have sufficient information by the end of X grade to be tested on this kind of information? And then is there any potential bias that you can see in the item itself? That’s before we do any kind of field testing.

“Those committees have the duty to let us know whether or not the item should even be field tested. Sometimes they revise items. Edit items. Sometimes they do a little changing. Don’t call it this, call it that. Make this purple instead of green. Whatever the case might be. Or, if they just feel the item won’t work. They they tell us that and we don’t field test the item. Once they review those items, we go ahead and field test items...

“...We annually release every test that we give. So that once an item is given in a live test, it’s no good to us anymore. (ABSOLUTE HORSE MANURE... – EMPHASIS MINE) So, we have to build enough new items every year to totally revise the test or the have a totally new set. We have to have enough items to build a totally new test every year...”

“...The items (questions – emphasis mine) really are decided in terms of their level back here with whether or not it is – whether or not they think that’s an appropriate objective for that grade level. Remember, this is a grade level test. Are these items that kids ought to be able to answer and do well on by the end of the third grade?

“Now the question you may be asking is the last item on this sheet which is the cut score – the performance standard itself...We use the item statistics and difficulty levels to make sure that the items we pick to put together on each test each year are predicted to have the same difficulty level as the test the year before. We keep the level of difficulty of the test the same unless we decide that we’re going to make that jump to the next level which we made in 1990 with TAAS...

“So the level of difficulty is designed to be the same from year to year...the State Board of Education...makes the decision about what the passing standard will be. And they do that with data after a benchmark administration. In other words, every student took the test. And we provide them with the data about the population performance. Each population performance.

“As they are making the decision about what this passing standard ought to be so they they know at a 50 percent standard what the pass rate would be for all students, for all African-Americans, Hispanics, and economically disadvantaged. At a 60 percent, at a 70%. At each standard level they know when they’re making that decision how it would impact various populations given the benchmark test, of course...