

Session B-1: Monitoring Test Administration, Including Special Populations

Panelists: Jan Sheinker, Heather Peltier, Sharon Hall

Moderator: Joseph Suh

Peer reviewers on the panel discussed the types of evidence that support a state's demonstration of effective test administration monitoring, both for general test administration and special populations.

Mr. Joseph Suh opened the meeting and read the text for critical elements 2.4 and 5.4 from the Guide. He introduced the panelists and asked them to respond to questions.

What have you seen states do well and not do well?

Ms. Jan Sheinker said monitoring is more than test security and more than following the design. Training for test administrators, while necessary, is often not sufficient. Training might not qualify as professional development. We need to know what training was done by whom, how often it will be provided, and who is responsible for ensuring that it is done. We need to know that everyone involved actually went through the training, either on-site or remotely. A significant problem for states is the gigantic task of monitoring the actual test administration on-site. There should be a set procedure, including what actions will be taken if problems arise. We often see a lack of evidence that any monitoring action was taken. The best submissions include a summary of the actions taken as a result of the monitoring.

Ms. Heather Peltier said it is good to see a lot of documentation. There is a top-down approach that assumptions will work. If monitoring is done at the LEA level, provide specific guidance, such as checklists. It is also good to see the strategy the state is taking—for example, by including past customer service reports or a statement of how the state targeted the sites for monitoring. It is important to cite evidence that action was taken if monitoring resulted in a finding.

Ms. Sharon Hall said you can't monitor without a policy. What would I see before, during, and after the monitoring? For special populations, that would include having documents in place. On-site monitoring is necessary to provide evidence or criteria to meet critical element 5.4 in the left and center columns. We also want to see that the accommodations shown on the students' plan are actually being provided. The teachers must be aware of the accommodations the students are supposed to receive. Then document that the test administrators are trained to monitor for this. Monitors should note whether there are anomalies in the data.

How does evidence of monitoring differ for different types of assessment?

Ms. Sheinker said evidence should not differ. Coherence should be evident across all the documents. They have struggled to get everyone trained in understanding what assessment is and how to make it coherent. Everyone has to be at the table, including those doing alternate assessments and monitoring ELP. A key way to avoid a lack of coherence is to have a common expectation—they expect that they are accountable to do "X." A lack of consistency in expectations is a disservice to test administrators.

Ms. Hall said the documents will be similar but will differ in some ways. Academic content assessments will be different from alternate assessments, so the resulting documents will be different. It is important that the training for monitors reflect those differences. The state must see the protocols for doing that.

Ms. Peltier said they want to see one monitoring plan, but the logistics will be different. Nuances are recognized, but they are included in one plan.

What evidence can states submit to show that tests are administered with fidelity?

Ms. Hall said a test administration manual should list the test, with directions for making it fair for the students who need accommodations. The protocol for monitoring explains how to read aloud and how

to scribe (two of the most common accommodations that can cause inadvertent indications of the answers). The monitor must be trained to listen for that. Afterwards, data analysis may reveal irregularities. For example, providing accommodations that are not in the individualized education program (IEP) would invalidate the test.

Ms. Peltier asked what happens when accommodations are not given, either when the student asked for one or declined one. They need to see a procedure for dealing with that situation. Ms. Sheinker said there is a difference between procedure and administration. It is critical, but they don't see it much. Everybody does their own on-site monitoring. To do comprehensive monitoring, you need to send a team for the three kinds of assessments you're looking for. It's labor intensive, but technology is enabling other ways of monitoring.

How might evidence for these critical elements differ for computer versus paper evidence?

Ms. Sheinker stated that the exact accommodations the student was supposed to receive should be entered in the pretest material. Technology can be used to map what actually gets turned on for that student. This creates a body of evidence for matching what was asked for with what happened. Checklists are one way to see what students actually received versus what they were supposed to get. These data are convincing. States also should ask whether the student refused the accommodation or whether it was not offered. Ms. Peltier said opportunities abound. Computer-based testing offers much more data (e.g., the length of time spent answering questions). Then it's a question of what states are looking at and how they are looking at it. Ms. Hall said to look at when the test was administered and the length of time the test was open on the computer. If a window was left open and unattended, security might be jeopardized.

What is the single biggest challenge for states for test administration?

Ms. Hall said there's no data validation that monitoring is actually taking place. Reviewers need to see a specific plan or protocol, rather than just read about the state's intentions. Ms. Peltier said sometimes it seems to be a "don't ask, don't tell" policy. The state should have plans, checklists, and procedures. In some states, this is done at the LEA level rather than using state teams. Reviewers need something that shows that the state is working with the test administrators. They need evidence that monitors were trained to use and apply the materials. Ms. Sheinker said to "trust but verify." Extensive planning goes into this. The real challenge is to provide evidence of the data collected. Realistically, a state will be monitored about every 10 years, so multiple ways are needed to verify the fidelity of testing.

What piece is most often missing?

Ms. Hall said critical element 5.4 provides several types of evidence in the second column. She would like to see states break it down by each subcomponent of the critical element, with an associated page number. Ms. Peltier said the naming protocol used by the state should use the critical elements so they naturally cluster together. That makes it easier to tell if something is missing. Ms. Sheinker noted that it is okay to borrow strategies and methodologies, but they are meant to be models and should be adapted for the individual state. It all needs to be coherent with what the state is doing, which means more than changing the state name on the document. Ms. Peltier added that the states don't have to reinvent the wheel. They just need to be sure that any models they're borrowing actually fit the state's activities.

Questions and Comments

- Mr. Mark Weinstein, Caveon Test Security, asked if there is a "magic number" in terms of the quantity of schools that should be monitored under these standards. In a state with 4,200 schools, what percentage should be monitored? Ms. Sheinker responded that there is no magic

number for critical mass for on-site monitoring, but you need multiple strategies, so you can adapt. Ms. Hall suggested asking the TAC about what would be reasonable to do.

- Ms. Nancy Godfrey, Maine Department of Education, noted that the assessment team in Maine consists of two people. She was not sure how she would do the monitoring and rely on the truthfulness of people at the schools. She asked if there was a template for training monitors. Ms. Hall said such a template would be based on the particular protocol for monitoring (i.e., based on the test administration manual). Ms. Sheinker said it's more effective to involve education people and testing people in the monitoring. Ms. Godfrey said that if the state has regional offices with peer review, there may be a vehicle for states to share procedures. Ms. Hall said local accountability coordinators would be helpful. Ms. Godfrey noted that there are none in Maine.
- Ms. Maxine Moore, North Carolina, said they have a percentage of schools they monitor, but how the schools get on that list varies. For example, it may be because of data irregularities. They have looked at a variety of ways and have considered involving other people. They try to go across regions so that people are not monitoring people they trained.
- Ms. Yvonne Fielding, Montana, asked, "What about cheating?" There are three people in their division. They have support from leadership because of peer review. She asked if they could provide any guidance on what an anomaly report should look like. Ms. Hall said their technology vendor can run those reports. Ms. Fielding wanted to know if there was a general guideline for what should be on the report. Mr. Lou Fabrizio stated that those are quite standard. The big thing they want to see in peer review is what you did when you identified anomalies (e.g., did you revoke a license)? Whether it's a state or a consortium, did the test administrators actually receive training? Mr. Weinstein added that a vendor can report anomalies in the test results. However, the statistics only indicate that something unusual occurred; it's up to the states to determine what the anomalies were. He said it's illuminating to watch administrators give a test. In many cases, monitoring is the issue. Also, talking to students about what happened during testing can be very helpful.
- Mr. Steve Ferrara, Washington, DC, noted that states have to decide whether they need professional help with test security or whether someone in the LEA can handle it. However, because no one wants to have trouble in their own LEA, they have an apparent conflict of interest. Knowledge, skills, and experience are needed.
- Mr. Pietro Semifero, Michigan, said the threat of state people showing up inspires improvement. Without the IEP plan, you don't know what the accommodations should be. Random monitoring is done in Michigan, but it is still hired out, as is the data analysis. Every school that "messed up" the previous year is monitored the next year, and professionals do it (a "cop" and an academic).
- Ms. Godfrey noted the issue of accommodations students should have had but didn't get, versus the accommodations students could have had but declined. She asked for opinions. Ms. Sheinker said at least one state checks to verify what happened in these cases. Ms. Hall suggested that the teacher document the student's decline of accommodations and enter it in the student's file or give it to the principal. Ms. Peltier said Tennessee has clear guidance on handling these situations. A record of the event is put in the student's cumulative folder, so students can't say later that something was denied. Ms. Sheinker noted that if it can't be demonstrated that the student declined the accommodation, the test is invalidated.
- Ms. Rae Clementz, Illinois, said various facilities have different philosophies on what it means to assess their students. Ms. Hall noted that some of that is in required state regulations, and Mr. Fabrizio added that also there are federal regulations.

- A participant from Florida asked about ways to avoid causing confusion for peer reviewers who are providing feedback. Ms. Peltier said reviewer feedback is highly dependent on the preferences of a particular review panel, which can vary. Ms. Sheinker said an example that might cause confusion is when a discussion about monitoring shifts to a different topic (e.g., test security).
- Mr. Weinstein noted that some states have challenges with alternate assessments. That frequently involves one-to-one test administration. If the student qualified to take the test has many challenges, having another person in the room to monitor could invalidate the exam. How do schools deal with that? Ms. Hall said the monitor could be placed where they are not visible but can hear, or a one-way window could be used. Ms. Sheinker said to be careful about generalizations. Case-by-case considerations are needed for the students.