

BRIEF #1: ALABAMA ASSESSMENT TASK FORCE

CONSIDERATIONS FOR PAPER PENCIL TESTING (PPT) VS. COMPUTER-BASED TESTING (CBT)

Juan D' Brot and Scott Marion, Center for Assessment

January 12, 2018

Considerations for choosing between PPT and CBT extend beyond the administration experience and technology capability in schools and districts. While technology capacity is usually a top consideration, the impact of the delivery mode also includes the following:

- Administration monitoring
- Test security and analyses
- Field test administration and design
- Scoring
- Comparability between modes

While this list is not exhaustive, these kinds of issues will dictate whether a state would support both PPT and CBT or choose a single administration method, which in turn will influence the cost of the assessment system. Generally, states should expect that dual mode administration (i.e., supporting both PPT and CBT) will be more expensive than supporting either mode alone. The remainder of this brief describes some considerations for each mode, mode issues and their complexity, and pose questions that will need to be addressed by the Task Force.

Paper Pencil Testing (PPT)

Paper Pencil Testing (PPT) uses paper for both the stimulus (e.g., test booklet) and response (e.g., score sheet). PPT offers an opportunity for easier administration, fewer technological considerations, and less perceived stress, the latter of which is typically attributed to those administering or supporting the test. However, it also sets up several challenges that limit the types of items that can be administered, reduce speed and efficiency of scoring, eliminate flexible approaches to field testing new items, and complicate the logistics of delivery, packing, and shipping. Furthermore, PPT usually eliminates the possibility of online manuals (e.g., test coordinator, test administrator, system administrator, etc.), ancillaries, graphics, and the like. This last point can be seen as either a benefit or a cost, depending on the case. Additionally, the cost associated with PPT administration is recurring and based on printing needs, pack and ship costs, physical scanning, warehouse space, and long-term storage.

Computer Based Testing (CBT)

The application of computer based testing (CBT) can vary, but for our purposes, CBT would be where both the stimulus (e.g., test item) and the response (e.g., item response) would be delivered and captured on an electronic device (e.g., desktop, laptop, tablet, etc.). While CBT has more startup capacity costs (e.g., infrastructure, hardware, and software), proponents of CBT argue that the tools for testing should be used for instructional delivery and student learning, mitigating some of this investment.

CBT offers several opportunities for efficiencies in delivery, administration, field testing, scoring, security, and reporting and supports the ability to leverage adaptive testing (i.e., adjusting test difficulty to the ability of the student). However, the primary challenges with CBT revolve around the number of available devices, scheduling time on those devices, system readiness (e.g., test administration system installation, system stress tests, etc.), and relevant training for educators (e.g., technology surveys, site readiness, administrator training, etc.). Like PPT, there are recurring costs. These costs are typically associated with help desk support and any annual printing of support materials. If, however, everything were supported through online administration and documentation, the initial cost would be associated with development of print-ready publication, potentially decreasing the overall costs for the life of a contract. It is important to note that these costs are usually shifted to districts and schools if they desire to have print-based resources.

Level of Complexity

As described in the two previous sections, PPT and CBT each come with their own issues and vary in complexity. We should keep in mind that supporting both modes can potentially negate the benefits of CBT- or PPT-only administration and increase the complexity of managing the program and assessment. The reality is that many states that have implemented CBT have supported dual mode assessment and address issues as they emerge. The following table outlines some of the issues and their corresponding complexity for each mode.

Table 1. Complexity of Testing Issues by Mode.

Topic	Paper	Online	Dual Mode
Administration	Low. A content area can typically be done in a single day, limiting the size of test windows.	Medium. Limited to the number of devices, often requiring larger windows	High. Requires managing both types of administration.
Design and Field Testing	Medium. Relatively restrictive field test designs, unless there is a heavy investment in administration time.	Low. Much easier to embed items to support comparability studies and linking item sets during field-testing.	High. Potentially eliminates the flexibility of online field test administration and could over-/under-represent certain schools or districts.
Scoring	Medium. Turn-around based on collection and pack/ship dates.	Medium. Turn-around based on size of testing windows.	High. Requires managing both physical and digital administration conditions and introduces examining differences in mode.
Item Types	Low. Limited to non-technology enhanced items.	Medium. Provides access to many more item types, but requires a strong rationale for their use (i.e., not using enhancements simply because they are available).	High. PPT operates as the driver for item type selection. Any dual mode items have to be examined for administration differences.
Comparability	Medium. Issues limited to accommodated forms, but it is difficult to track administration conditions without a strong policy.	Medium. Should be studied by device type, screen size, and peripherals. Administration choices (e.g., tools or accommodations) are easier to monitor.	High. Requires careful analysis of mode effects and how it may affect student scores.

Topic	Paper	Online	Dual Mode
Test Security	High. Test security can be more challenging due to physical test copies and fewer available analyses to monitor irregularities.	Medium. Test security can be less problematic due to electronically-mediated methods of delivery and more robust monitoring and analyses.	High. A strong rationale is required to select the appropriate analyses and monitoring techniques, especially if they differ by mode.
Cost	Medium. Typically front-loaded in delivery and spread over the cost of the contract; annually based on printing.	Medium. Similar to PPT, but also based on the level of support necessary and need for scalability of support (e.g., help-desk support).	High. Supporting both PPT and CBT will require factoring the challenges from both modes.

Questions to Answer

Based on the information presented above, the Task Force should be prepared to address the following questions:

1. Should supporting materials and documentation (e.g., test administrator manuals, test coordinator manuals, student ancillaries) be printed, regardless of decisions to support CBT or PPT?
2. Given the current state of online testing with the current vendor, is it worth supporting dual mode administration on a new assessment system that will be administered in the Spring of 2020?
3. If dual mode testing is supported, should the state of Alabama transition toward 100% online testing by some date in the future? Please note that any interim assessments, as described during our face-to-face meeting, would likely need to be online.