

Recommendations for Reducing ISTEP+ Testing Time



EDWARD ROEBER

WILLIAM AUTY

Overview of the Presentation

2

- Introduction
- Review Principles
- Findings
- Recommendations
- Work to be Done

Introduction

3

- Our goal was to locate the sources of the testing time issue in order to enable us to make recommendations regarding how testing time for the 2015 ISTEP+ could be reduced.
- The Department and its contractor, CTB-McGraw-Hill, have been forthcoming and helpful in this review process.
- The purpose of this review was focused entirely on how to reduce the testing time without unduly affecting ISTEP+ this year or in the future.

Basic Principles

4

- **ISTEP+ testing time in all grades (grades 3-8) needs to be reduced**
- **Results of 2015 tests should be sufficiently reliable and valid to enable the intended purposes to be achieved.**
 - There are not fixed standards to determine whether a test is sufficiently reliable or valid. Instead, professional judgment is called for.
- **Changes made in the 2015 program should not unduly impact the 2016 program**
 - It is essential that this year's issues not continue next year and beyond.

Basic Principles

5

- **Our recommendations should not be restrictively prescriptive.**
 - IDOE has the responsibility for and understanding of the details of ISTEP+ design and implementation.
 - What we are proposing are parameters for how the testing time could be reduced.
 - We expect that IDOE and its contractors would use these guidelines to effect the suggested changes.

Findings

6

- Testing times in excess of 12 hours are currently scheduled for the mathematics, English language arts, science, and social studies tests in ISTEP+.
- The mathematics test contributes about 4 hours and the ELA test contributes over 8 hours of testing time.
 - The ELA tests are the real culprit
 - Steps should be taken to reduce the length of both the mathematics and ELA tests.
 - A major part of the extra testing time occurring in the English language arts assessment is due to policies on the release of open-end or constructed-response items.
- We believe that it is unrealistic to expect young children to take an assessment of 12 hours in length.

Findings

7

- **This said, states and the state assessment consortia are adding significant testing time to their programs, due to:**
 - more comprehensive standards
 - increased use of performance assessments to better gauge student achievement
 - Even reduced ISTEP+ tests may be longer than those used in the past
- **The lack of previously pilot-test or field-test items requires the use of more items than normal this year.**
- **The overage levels of “extra items” are not excessive (in the 50% range), given the structure of the tests and the nature of the intended score reports.**
- **We believe that there are ways the Department can reduce testing time and accomplish its assessment purposes.**

Recommendations

8

- 1. The Department not release 100% of open-ended (OE) items used in the 2015 ISTEP+. Instead, we recommend a far smaller percentage of item releases—perhaps 20%.**
 - Rationale—Our preliminary review indicates that the greatest contributor to the increased testing time is the need to administer enough OE items to be used operationally in 2015 as well as pilot items for 2016 tests.
 - We recognize that educators in Indiana rely on released items to guide instruction.
 - We would therefore include the recommendation that a few operational items be released and that these be supplemented with high-quality *example* items to be produced this spring and released publicly when the results are reported.

Recommendations

9

2. Some parts of the 2015 ISTEP+ be administered to only a sample of students being tested this year.

- Rationale: Another significant contributor to the test length is the requirement that all students take all items, including items that will be used as pilots for future testing.
- We recommend that items in test sessions be identified as “core” or “sample” items. All students would take the core items and half, a third, or fewer would take each set of sample items. Field testing of items usually can be done with many fewer student responses.
- Because test forms have already been designed, this change will result in somewhat less reliable reporting of standard information at the individual student level. Strand-level reporting at the school and district level can also remain high.
- If this recommendation to identify a core assessment and sets of sample items is carried out, this recommendation can apply to both ELA and mathematics.

Recommendations

10

- 3. Move the deleted OE items to fall pilot testing, and then matrix sampling in 2016 test, in order to use them to build forms for 2017 and beyond.**
 - Rationale—If the items are not field tested in 2015, they still need to be field tested in 2016.
 - They should be pilot tested in the fall, to make sure they are ready for field testing in 2016.
 - Then, they will be ready to be used in 2017 and beyond when the release of OE returns to a higher percentage in keeping with past practice.

Recommendations

11

4. The Social Studies part of the test be suspended for 1 year.

- Rationale: Since these tests are not required by NCLB, nor apparently used in school accountability, this change will reduce testing time by over an hour for student in grades 5 and 7.

Recommendations

12

5. Vertical scaling items should be removed from the online assessments.

- Rationale: There are other options for calculating growth in 2015 and the vertical scale could be constructed in 2016.
- However, the number of items taken by a student is small (5) and so testing time may not be significantly reduced.
- This would not affect test length by much.

Work to be Done

13

- **Recommendation 1** (Limit the amount of open-response items to be released)
 - How many items are needed to build this year's final form and the comparable forms next year, if items are reused that worked this year?
 - Which items can be eliminated?
- **Recommendation 2** (Identify the “core” and “sample” parts of the tests. Matrix sample the sample items)
 - Determine the core and sample parts of the assessments.
 - Determine how matrix sampling will be implemented in both the paper/pencil tests and the online assessments.

Work to be Done

14

- **Recommendation 3** (Pilot test the deleted OE items in Fall 2015 and field test them, using matrix sampling in 2016)
 - Determine how a fall pilot test can be added to the schedule and implemented so the deleted items are ready to be field tested in 2016.
- **Recommendation 4** (Suspend the Social Studies tests for one year)
 - Determine what changes would be required to implement this recommendation.
- **Recommendation 5** (Remove Vertical Scaling items from this year's test)
 - Determine what changes would be required to implement this recommendation.

Closing

15

- We commend the state for tackling this thorny issue and working together to resolve it.
- We believe that if these recommendations are followed, testing time can be reduced to more manageable levels.
- We remain willing to assist in efforts to implement these recommendations.