

Comments on Issues Related to Indiana's Transition to a New Student Assessment

Derek Briggs, PhD

University of Colorado

July 8, 2014

Issues I Was Asked to Address

1. Growth Options (Comment on Betebenner Memo)
2. Why Not Use NWEA's MAP?
3. Adaptive vs Fixed Form Tests
4. CTB's CCRTA "Operationalized Field Test"
 - Owning vs. Leasing Test Items
 - Comparability of standards set on tests provided by two different vendors

1 Year Projected Approach

- Strictly speaking requires two years of the same assessment system.
- Betebenner proposes an equipercentile concordance mapping as the next best thing.
- Details on how this would proceed are not yet available.
- Should be possible to evaluate whether this approach is valid using real data...

A Concern about Any Growth Approach

- One reason why growth metrics have been introduced into school and teacher accountability is that they have the potential to communicate something different than a status metric.
- Status (e.g., % of kids passing test in one year) is highly correlated across years.
- In contrast, growth and status are usually only weakly associated—many low performing students are still capable of demonstrating growth.
- The more strongly a growth metric is correlated with status, the more that it may communicate redundant information.

Interpretation?

	A3	B1	B3	C1	C2	SGP
Prior Achievement	0.55	0.55	0.55	0.38	0.34	0.28
FRL%	-0.38	-0.38	-0.38	-0.28	-0.25	-0.17

- Different ways of defining “growth” can lead to different conclusions about teacher and schools that are successful at raising achievement.
- Categorical growth models A3, B1 and B3 give schools more credit for students who stay in same achievement level, models C1 and C2 give less.
- SGP approach is always comparing achievement for students who were comparable in prior test performance.

NWEA and the MAP Tests

- The MAP tests pretty consistently attract a strong following among schools/districts that use them. Seem to be doing something right!
- However, the MAP was designed to be an interim assessment for use as a complement or supplement to a state exam written for high-stakes purposes.
- While it might be possible that the MAP could be validly used for such purposes, to date no such evidence has been made publicly available.

Questions for NWEA

1. How large is the MAP item bank? How often are new items introduced and old items retired?
2. How are you able to assess depth of knowledge in math and writing ability in ELA with all MC items?
3. How would you establish that MAP items are aligned to the Indiana's content standards? How has this been established for other states?
4. What has happened in other settings when the consequences associated with MAP scores go from low to high stakes?

Adaptive Tests

- A computer-based adaptive test can provide many advantages:
 1. Can minimize measurement error
 2. More efficient use of testing time
 3. Immediate feedback

But a Lot Can Go Wrong

- A large item bank is needed.
- Since every test event is different, hard to establish test alignment to standards
- Meeting specifications of test blueprint can compromise key advantage of adaptivity.
- Technical demands of underlying psychometric model are stringent.

Given a tight timeline and resource constraints, a fixed test may be the more conservative and sensible approach.

CTB's College & Career Readiness Transition Assessment (CCRTA)

- Appears state will only be leasing these items.
- Problematic if there is a desire to create scale and set standards for 14-15 test that can be connected to 15-16 test.
- Tests can only be equated if there is overlap in either items on the tests or kids taking them.

“Operationalized” Field Test

- Defined as such because, according to CTB the test will contain both operational and field test items.
- But which proportion will be “operational” and which won’t?
- To quote Inigo Montoya from *The Princess Bride*: “I do not think that word means what you think it means.”

<http://www.youtube.com/watch?v=G2y8Sx4B2Sk>