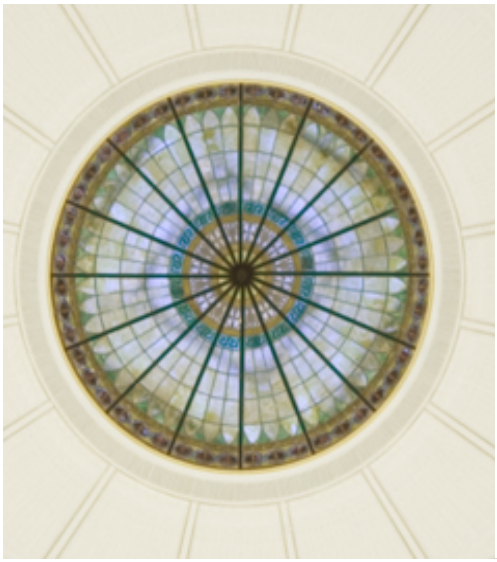
SMU



Establishing a Stopping Rule for an Instructionally Informative Diagnostic Assessment of Algebra Readiness

Deni Basaraba, Paul Yovanoff, & Leanne Ketterlin-Geller

Research in Mathematics Education

Southern Methodist University

SMU

Overview of the Presentation

- Rationale for diagnostic assessment of algebra readiness
- What are stopping rules and why are they important?
  - Rationale for stopping rules
- How do you establish a stopping rule?
  - Our definition of a stopping rule and criteria used to evaluate
- Methods
- Results

Rationale for Diagnostic Assessment of Algebra Readiness

- Recent student performance data indicate that 27% of 8th students are considered *Proficient* and only 9% are considered *Advanced* on the most recent NAEP (NCES, 2013)
- More states, districts, and schools are implementing multi-tiered integrated models of instruction and assessment to help identify students who may struggle to reach grade-level proficiency standards
- Successfully supporting these students requires access to theoretically-grounded, technically adequate diagnostic assessments to help teachers pinpoint *why* students may struggle with the content

What are Stopping Rules?

- Stopping rules typically specify a number of items that can be missed within a set of given items before administration of an assessment is discontinued
  - Premise: If items are ordered from least to greatest difficulty, stopping administration of the assessment after a child misses a certain number of items is unlikely to result in a loss of information
  - Employed on a number of achievement tests and, more recently, on formative mathematics assessments
  - Important because they are designed to provide an accurate estimation of student ability without requiring students to take all of the items on the assessment

Why are Stopping Rules Important?

- Rationale for including a stopping rule in an assessment is fourfold
  - Minimize test-taking burden placed on students and any fatigue they might experience
  - Maximize test-taking time efficiency
  - Obtain accurate estimates of students' current level of knowledge, skills, and understanding of the assessed content
  - May support the instructional utility of the results for teachers

How do you establish a stopping rule?

- Our definition of a stopping rule
  - The point at which administration of an assessment is discontinues that provides teachers with a reliable estimate of understanding about the assessed content *and* sufficient information to help him/her target instruction to meet students' learning needs

How do you establish a stopping rule?

- Consider the type of information about student errors you wish to collect
  - Slips: Random errors in students' procedural and declarative knowledge
    - Focus on mastery of content within the domain
  - Bugs: Persistent misconceptions about domain-specific knowledge or skills that consistently interfere with students' ability to demonstrate their understanding of the content

How do you establish a stopping rule?

- Potential criteria to consider
  - Efficiency
    - Administering only as many items as necessary to estimate ability reliably
  - Reliability
    - Administering enough items to have reasonable degree of confidence in estimation of ability
  - Relevance
    - Is information obtained from the assessment instructionally relevant for teachers?

Method

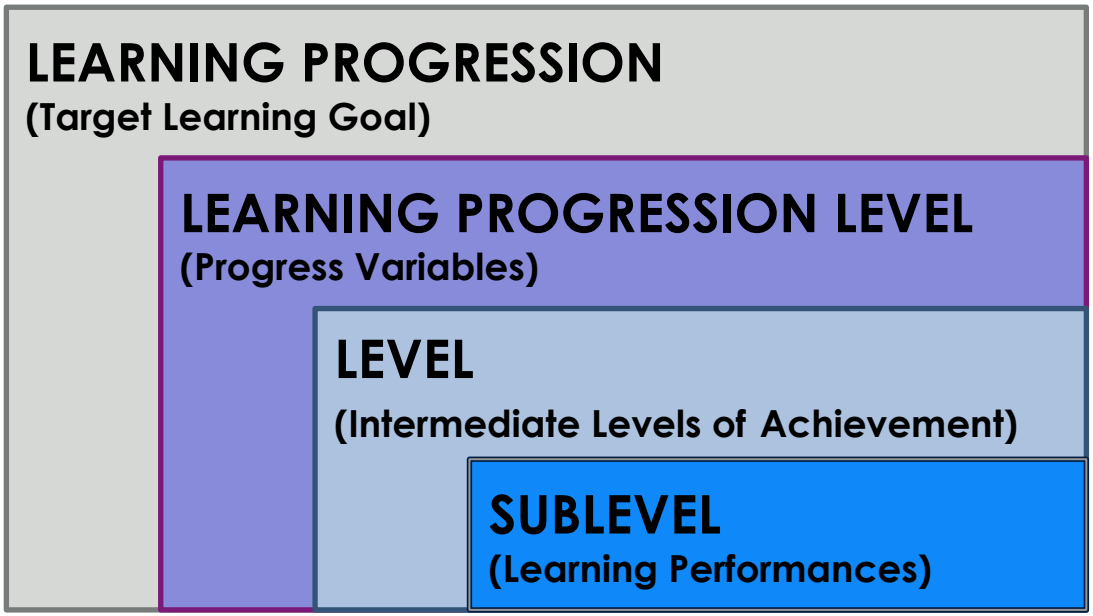
Participants

- Full Sample: 270 students in Grades 5-8 from 3 middle schools
- Analytic Sample: 55 students
  - 18 5th grade students
  - 20 6th grade students
  - 11 7th grade students
  - 6 8th grade students

Measure

- Diagnostic assessment of algebra readiness designed using mathematical learning progressions as the cognitive model
- Complex structure
  - Learning Progression (target learning goal)
  - Learning Progression Level (progress variables)
  - Level (intermediate level of achievement)
  - Sublevel (learning performances)
- Stopping rule: 3 consecutive items incorrect *within* a Level

Structure of MSTAR Learning Progression



2 Learning Progressions

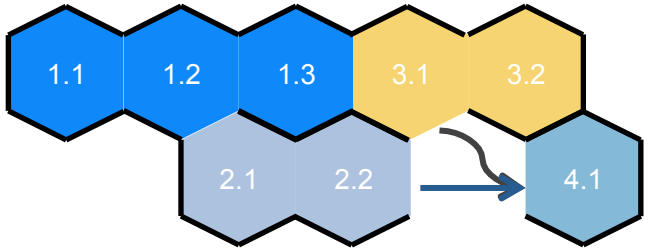
One test for each of 5 LP Levels

Multiple Levels comprise each LP Level

Multiple Sublevels comprise each Level

Items within a test form (LP Level) are ordered *across* Levels by item difficulty from easiest to hardest

Items are also ordered *within* Levels and Sublevels from easiest to hardest



Analyses

- Two types of stopping rules are proposed
  - Comparing three consecutive incorrect responses to two- and four consecutive incorrect responses
  - Comparing 80% proficiency to other, less stringent percent proficiency criteria

Analyses

- Efficiency
  - Use 2 PL item parameters to estimate (a) student ability and (b) probability that student will respond correctly to next item (conditional on the ability estimate and known item parameters)
  - Use logistic regression to treat correct responses on the next item as a dichotomous DV and number of sequential incorrect responses (e.g., 1, 2, 3) as IV
- Reliability
  - Use 2 PL item parameters, estimate student ability and overall measurement reliability after each item response

Results - Efficiency

Probabilities of responding incorrectly to the next test item conditional on a sequence (1, 2, 3) of incorrect responses

	Observed Probability				Logistic Regression		
Level	1	2	3		1	2	3
4	0.50	0.52	--		0.50	0.52	0.54
5	0.39	0.30	--		0.38	0.30	0.23
6	0.29	0.53	--		0.29	0.53	0.76
7	0.47	0.68	--		0.47	0.68	0.83

Apart from Level 5, the probability of selecting an incorrect response is greater than the probability of selecting a correct response after 2 consecutive incorrect responses

Stopping rule of 2 or 3 consecutive incorrect responses may be defensible

Results - Efficiency

Probabilities of responding incorrectly to the next test item conditional on meeting a set percent proficiency criterion for all items in the Level (e.g., 80% of items within a Level correct)

Level	80% or higher	70% - 80%	60 – 70%	Less than 60%
4	0.31	0.40	0.37	0.47
5	0.31	0.40	0.34	0.44
6	0.21	0.26	0.36	0.27
7	0.68	0.48	0.51	0.54

For Levels 4, 5, and 6, probability of selecting an incorrect response was relatively low (~0.30) when students were held to an 80% proficiency criteria

As the percent proficiency decreases (e.g., 60%) the probability of selecting an incorrect response increases

Results - Reliability

Using 2 PL item parameters and computer-adaptive testing (CAT) psychometric modeling, estimated ability and overall measurement reliability

Optimal stopping rule will be response at which neither ability nor reliability change by some specified amount

	Number of Consecutive Incorrect Responses		
Level	1	2	3
4	0.72	0.77	0.78
5	0.23	0.32	0.48
6	0.39	0.58	0.68
7	0.45	0.64	0.76