

Decisions to be made in developing an adaptive testing system for K-12 education

G. Gage Kingsbury
March 9, 2012



Welcome and Introduction



Presenter

G. Gage Kingsbury

Vice President for the International Association for Computerized Adaptive Testing (IACAT) and Senior Research Fellow at the Northwest Evaluation Association (NWEA)

Decisions to be made in developing an adaptive testing system for K-12 education



The Idea

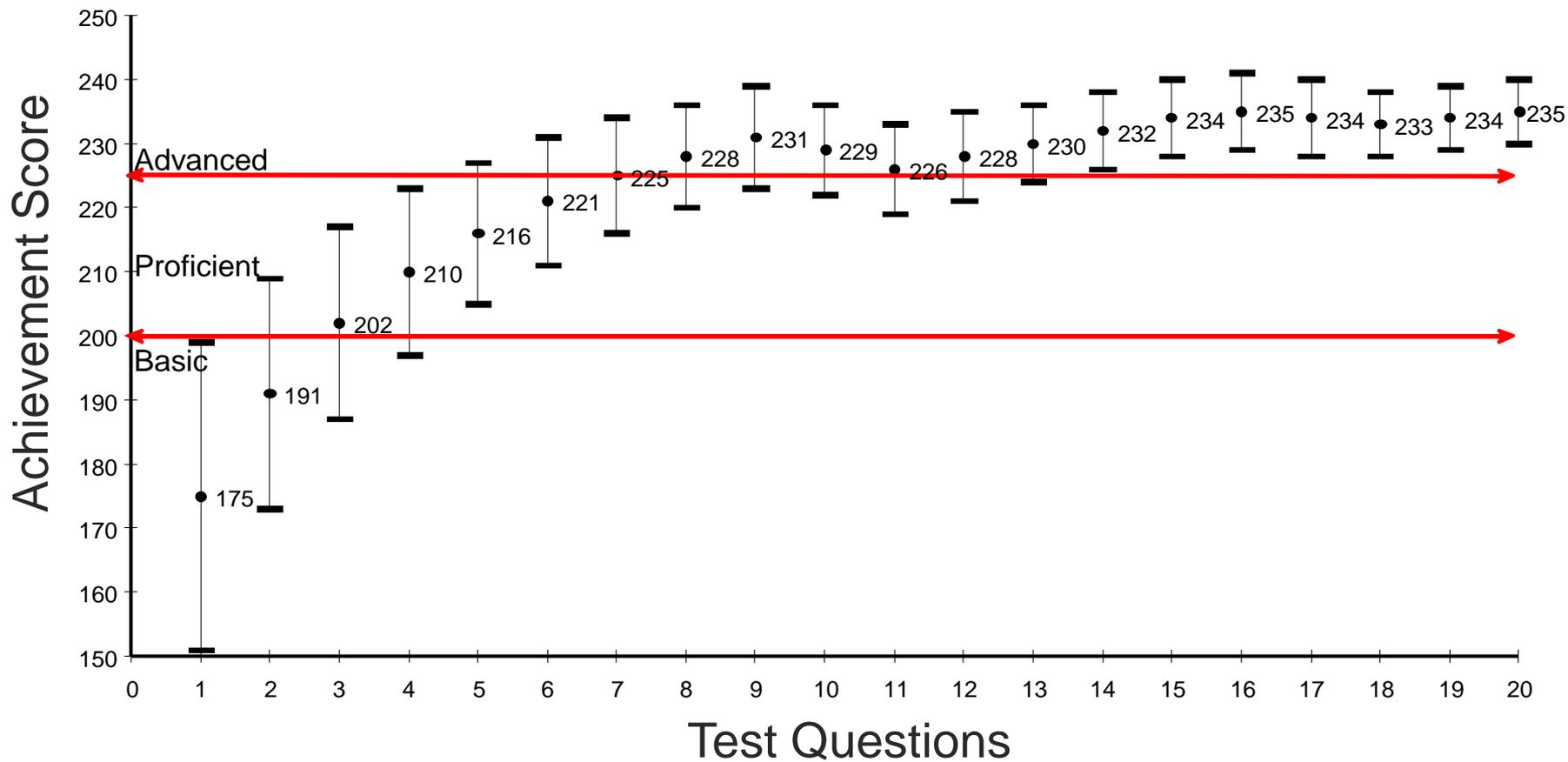
An [adaptive test](#) is a test that adjusts its characteristics based on the performance of a test taker.

Questions and Answers



Computerized Adaptive Testing

20 Item Test



Pioneers of adaptive testing

- [Alfred Binet](#)
- [Frederick Lord](#)
- [David J. Weiss](#)
- Fumiko Samejima
- Mark Reckase

First implementers

- David Foster
- [Jim McBride](#)
- [Tony Zara](#)
- Gage Kingsbury

You have chosen to use an adaptive test because ...

- It can be more efficient than a fixed-form test
- It provides good information across a broader spectrum of student performance
- It can provide immediate scoring and reporting
- It can provide better security than a fixed-form test
- It can be designed to measure growth

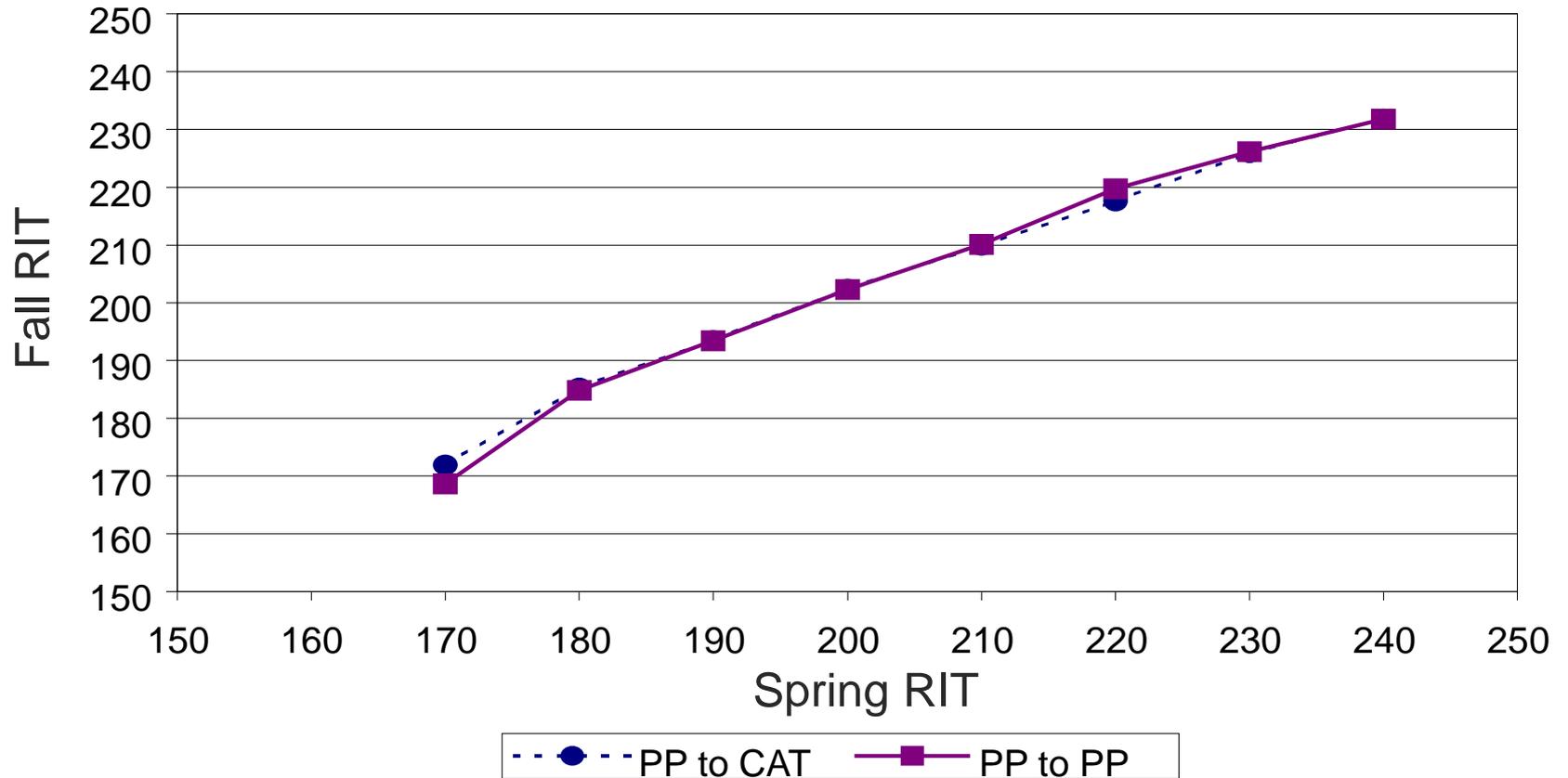
Since the first implementations

- We have seen international growth in the use of CAT for
 - Educational testing
 - Medical outcomes assessment
 - Certification and licensure

Accuracy of adaptive tests

- Compared to a fixed-form test
- As a function of test length
- Depending on termination procedure

Relationship between Spring and Fall Reading Scores



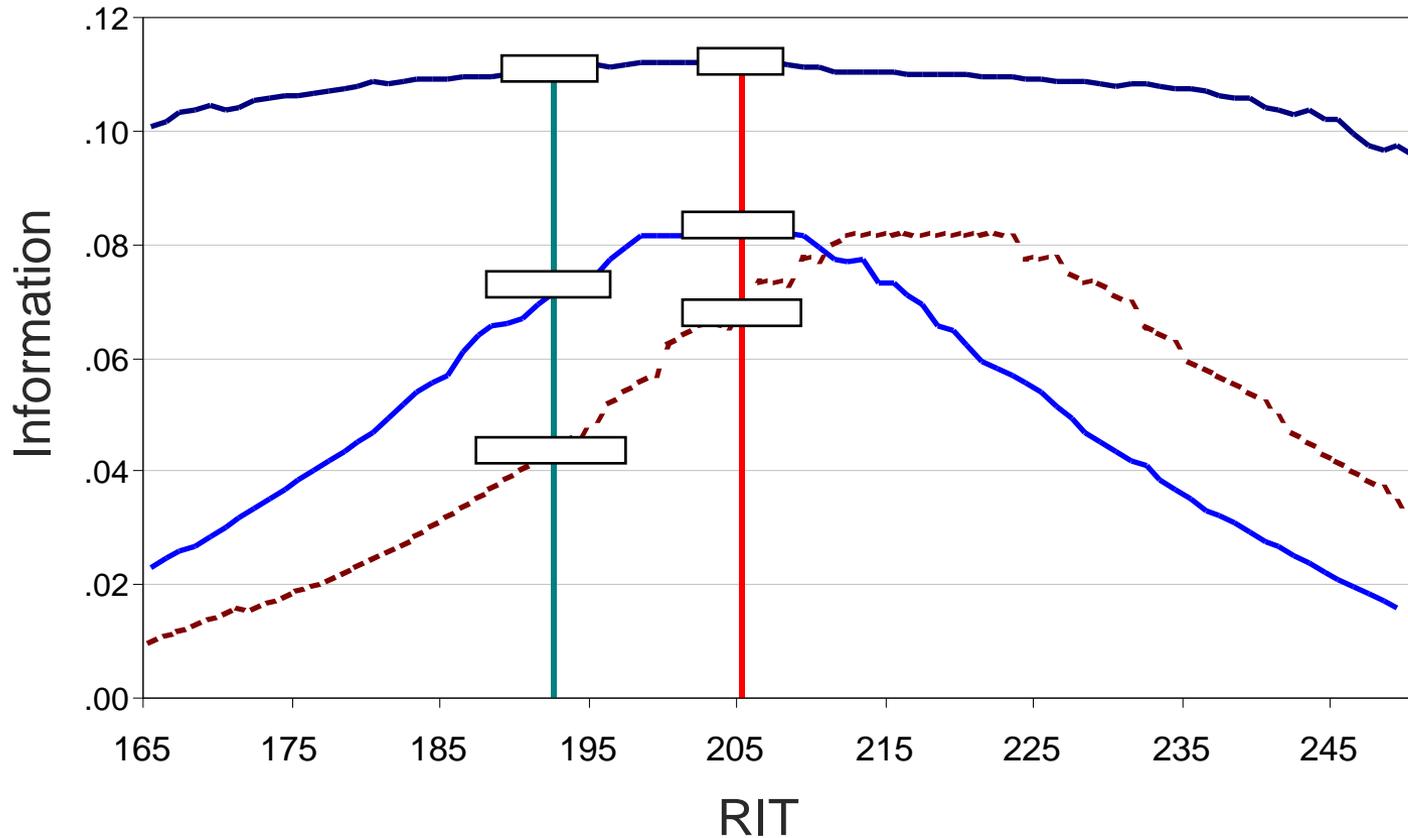
Students' Mean = **211.7**

s.d. = **11.11**

Proficiency = **205**

Basic = **192**

Test Information Functions for Grade 4 Mathematics



Choosing to use an adaptive test requires making a series of decisions in the areas of...

- Psychometrics
- Interface (including accommodations)
- Item designs
- Test designs
- Test distribution
- Item usage
- Item and test security
- Proctor training
- Reporting

Basics of a theoretical CAT

- IRT model
- Item pool
- Select first item
- Select next item
- Terminate test
- Score

Decision areas for an operational CAT for measuring student achievement

- **Before the test (Test stuff)**
 - How will we develop the measurement scale?
 - What mix of [item styles](#) will we need?
 - Which [IRT model](#) is appropriate?
 - What depth do we need in our item bank?
 - How will we choose an operational item pool?
 - What will our [test blueprint](#) include?
 - How will we QA everything involved?

Questions and Answers



Decision areas for an operational CAT for measuring student achievement

- **Before the test (School stuff)**
 - School, teacher, and student identification
 - Establishing a testing environment
 - Teacher training
 - Software/hardware setup
 - Proctor training
 - Student familiarization
 - Student scheduling
 - QA

Decision areas for an operational CAT for measuring student achievement

- **Test administration**
 - Student verification process
 - Test selection
 - Proctor throughout
 - Identify previously used items

Decision areas for an operational CAT for measuring student achievement

- **Test event**

- [Apply test blueprint](#)
- Select first item or set of items
- Check for effort
- Update item selection $\hat{\theta}$
- Update constraints
- [Select next item](#)
- Terminate test

Decision areas for an operational CAT for measuring student achievement

- **After the test**

- Calculate final score
- Calculate growth
- Terminate test session
- Store data
- Identify student as completing test
- Compare to norms, growth norms, content, etc.
- Create individual student report
- Add information to teacher/administrator reports

Measuring growth and adaptive testing

- Measuring at multiple points in time
- The standard deviation of growth
- The standard error of growth
- Reduction of uncertainty
- Growth and instruction

Adaptive testing and idiosyncratic knowledge patterns

- Can there be multiple thetas without multidimensionality?
- Selecting items to reveal knowledge patterns
- [A simple algorithm](#)
- The impact on instruction

Field testing within an adaptive testing system

- Calibration differences from paper to CAT
- Random sampling for calibration in CAT
- Using provisional calibrations in [CAT field tests](#)

Cautionary notes

- Adaptive testing needs to be well tuned to avoid bad tests.
- The item pool must support the stakes.
- Adaptive testing changes, but doesn't eliminate, security issues.
 - Brain dump sites
- Limit desire. No test can do everything.
- Adaptive test development is never done.

Have fun

- The decisions to be made should consider the good of the students for whom the test is designed.
- Don't try to build the perfect test—it won't be.
- Consider a “dry eye” policy—making kids cry isn't the purpose of the test.

Thank you

Gage Kingsbury

gagekingsbury@comcast.net