Development of Cognitive Complexity Measures for PARCC

April 5, 2014

A collaboration and presentation by PARCC, ETS, and Pearson in the Cognition and Assessment SIG Business Meeting at the annual meeting of the American Educational Research Association, Philadelphia

Overview for the presentation

- Cognitive complexity team
- PARCC assessment goals and role of cognitive complexity
- Cognitive complexity: Definition and intended uses
- ELA and mathematics cognitive complexity sources
- Validation, refinement, and potential uses

PARCC cognitive complexity team

- PARCC and Achieve (PARCC's support contractor)
- Educational Testing Service and Pearson (development contractors for PARCC)
- Development of the cognitive complexity framework a collaboration involving Enis Dogan, Steve Ferrara, Nancy Glazer, Joanna Gorin, Jeff Haberstroh, Bonnie Hain, Kristen Huff, Patricia Klag, Jay Larkin, Steve Lazer, Ric Luecht, Paul Nichols, Carrie Piper, and Kathleen Sheehan

PARCC assessment goals

- Determine whether students are college and career ready or on track
- Assess the full range of the Common Score State Standards, including standards that are difficulty to measure
- Measure the full range of student performance, including high and low performing students
- Provide data during the academic year to inform instruction, intervention, and professional development
- Provide data for accountability, including measures of growth
- Incorporate innovative approaches throughout the system

Partnership for Assessment of Readiness for College and Careers

Ultimate goal: Validity of intended score interpretations, evidence for validity argument

- Primary interpretation is the performance level into which students are classified
- Task models and cognitive complexity framework help to link items to intended scale locations
- As such, intended inferences about student proficiency, represented by the PLDs, are supported by item locations on the scale—their difficulty and their complexity

Adapted from K. Huff, 2013

PARCC Partnership for Assessment of Readiness for College and Careers

ETS

PEARSON

What is cognitive complexity?

- An item's cognitive complexity accounts for the content area, cognitive, and linguistic demands required to understand, process, and respond successfully to that item

- PARCC cognitive complexity measures account for those things
  - Multidimensional
    - Other indicators of cognitive complexity (e.g., Bloom's Taxonomy, Depth of Knowledge) summarize (or obscure) that multidimensionality
  - Conceptually, not the same thing as item difficulty or discrimination
  - Item response demands are moderately correlated with item difficulty and discrimination in several studies (e.g., Ferrara et al., 2011; Huff 2003)

PARCC Partnership for Assessment of Readiness for College and Careers ETS PEARSON

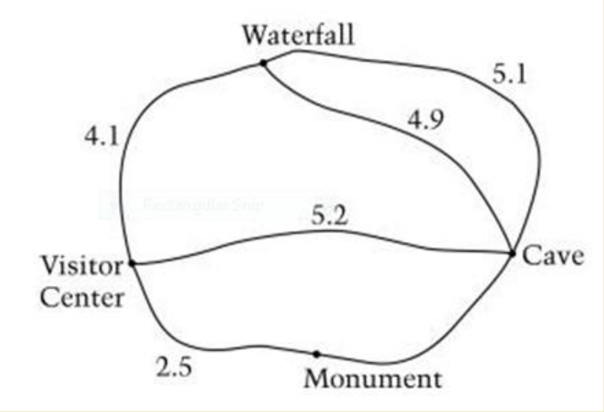
What is it (cont.)?

For each item or task, high, medium, or low complexity

ELA	Mathematics
Text Complexity	Mathematical Content
	Mathematical Practices
Command of Textual Evidence	Stimulus Material
Response Mode	
Processing Demands	

PARCC Partnership for Assessment of Readiness for College and Careers ETS PEARSON

Illustration



The map above gives the distances, in miles, between various locations in a state park. Teresa travels the shortest possible total distance from the visitor center to the cave, waterfall, and monument, but not necessarily in that order, and then returns to the visitor center. If she does not retrace her steps along any path and the total distance that Teresa travels is 14.7 miles, what is the distance between the cave and the monument?

Content: L Practices: M Processing: H

Judgment of overall cognitive complexity: M

Adapted from J. Haberstroh, 2013

ELA/L Cognitive Complexity Measures

- Three main sources of item complexity:
  - Command of textual evidence
  - Response mode
  - Processing demand
- Each rated as high, moderate, or low; the three are then combined into a single score for Processing Complexity
- Processing Complexity is combined with the fourth source, Text Complexity, to produce an overall cognitive complexity score

PARCC Partnership for Assessment of Readiness for College and Careers ETS PEARSON

Command of Textual Evidence

- Amount of text examinees must process in order to respond correctly to an item:
  - Low complexity associated with items targeting a single piece of information in a single text
  - Moderate to high complexity associated with items requiring synthesis of ideas and details, either from a single text or across multiple texts
- Other contributing item features:
  - Ease or difficulty of locating needed information
  - Type of transformation needed (e.g., whether there is a close correspondence between wording of key and expression of requisite textual evidence or a conclusion/interpretation is needed in order to key the item)

Response Mode

- How an examinee is required to respond to an item; in general, selecting a response is less demanding than constructing a response:
  - Low complexity associated with selecting a correct answer from a series or list of options
  - Moderate to high complexity associated with selecting multiple correct answers (from an expanded list of options), ordering response options, and writing an extended constructed response
- Other contributing item features:
  - Closeness of response choices

PARCC Partnership for Assessment of Readiness for College and Careers ETS PEARSON

Processing Demand

- Linguistic demands and reading load in item stems, item directions, and response options;
- Three contributing features, with values ranging from low to moderate complexity:
  - Knowledge of words and phrases (e.g., whether generally common, relative to grade level, or unusual, idiomatic, or abstract)
  - Grammatical complexity (e.g., whether or not item requires processing of dependent clauses, complex verbs, relative pronouns, and/or prepositional phrases)
  - Reading load

PARCC Partnership for Assessment of Readiness for College and Careers ETS PEARSON

Text Complexity

- Qualitative and quantitative score assigned to each text in separate process prior to item writing and cognitive complexity scoring

- Three possible ratings:
  - Readily accessible
  - Moderately complex
  - Very complex

Combining the Sources of Complexity

- Eight item-level complexity ratings are aggregated into overall processing complexity rating and then combined with text complexity rating
- Each contributes 50% of final item rating
- Rater judgment required to adjust "split"ratings (low/moderate or moderate/high)

PARCC Partnership for Assessment of Readiness for College and Careers ETS PEARSON

Preliminary and Final Ratings of Cognitive Complexity

- Item writers
  - assign preliminary ratings during initial item writing/reviewing stage
  - keep sources of cognitive complexity—command of textual evidence, response mode, processing demand—in mind as items are written and reviewed
  - use holistic, judgmental approach to estimate ratings
- Expert raters
  - rubric-based ratings
  - assign final rating

PARCC Partnership for Assessment of Readiness for College and Careers ETS PEARSON

Mathematics Cognitive Complexity Sources

Three main sources believed to be predictive of mathematical
cognitive complexity

- Mathematical Content
- Mathematical Practices
- Mathematical Process (stimulus material, response mode, processing demand)

Each rated separately as either high, moderate, or low according to
the criteria in a rubric, and then ratings combined into a single estimate.

PARCC Partnership for Assessment of Readiness for College and Careers ETS PEARSON

Mathematical Content Complexity

- Relative to the typical mathematical knowledge expectations at the grade level, the extent to which an item or task requires the content to be accessed and applied
- More routine applications associated with low to moderate complexity; less routine with moderate to high complexity
- Other contributing factors
  - Types of numbers
  - Expressions and equations
  - Figures and graphs
  - Problem structures

Mathematical Practices Complexity

What the student is asked to do with the mathematical content , relative to the four sub-components below. This source reflects the level of mathematical cognitive demand in the item or task.

Sub-components of Mathematical Practices

- Prompting (more directed vs. less directed)
- Level of integration of knowledge
- Mathematical modeling process
- Explanations, justifications, proofs; degree of scaffolding

Mathematical Process Complexity

Other sources believed to contribute to overall mathematical complexity

Sub-components of Mathematical Process Complexity

- Stimulus materials
- Response mode (SR; multiple; online tools; extended CRs)
- Processing demand (includes linguistic demand and processing steps)

  - Linguistic demand (increased amount of text; vocabulary)

  - Processing steps (multiple non-iterative steps, concepts, processes)

Combining Sources of Complexity

Nine different complexity ratings are assigned to each item.

- Mathematical Content Complexity
- Mathematical Practices Complexity (prompting, integration, modeling, argument)
- Mathematical Process Complexity (stimulus material, response mode, linguistic demand, processing steps)

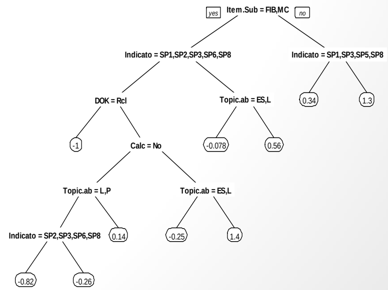
Ratings are then aggregated. Content and Practices each contribute equally, and slightly less than Process to the final overall estimate.

Final holistic check: Does the estimate seem "right" relative to the mathematical criteria described in the individual sources?

PARCC Partnership for Assessment of Readiness for College and Careers ETS PEARSON

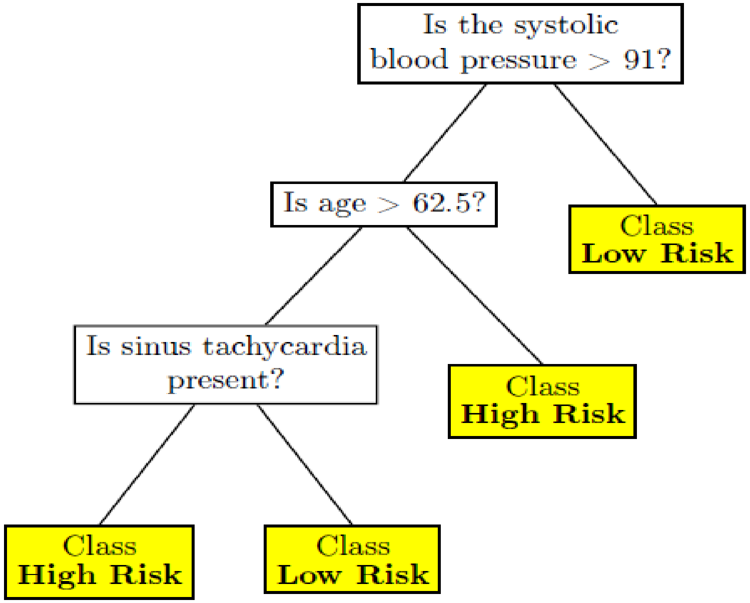
Validation and refinement

- Rater agreement: qualification and validity check sets
- Validation
  - Identify individual sources, empirically, that contribute to complexity
  - Efficacy for hitting difficulty targets, improving discrimination
  - Think alouds: Does examinee thinking correspond to judged complexity levels
- Model, empirically, the relative contributions of each source and the interactions/dependencies among the sources: OLS regression, classification and regression tree (CART) analysis



Regression Tree analysis

- Classification and Regression Tree analysis (CART; Breiman, Friedman, Olshen, & Stone, 1984) is used to “grow” binary decision trees that can predict outcomes.



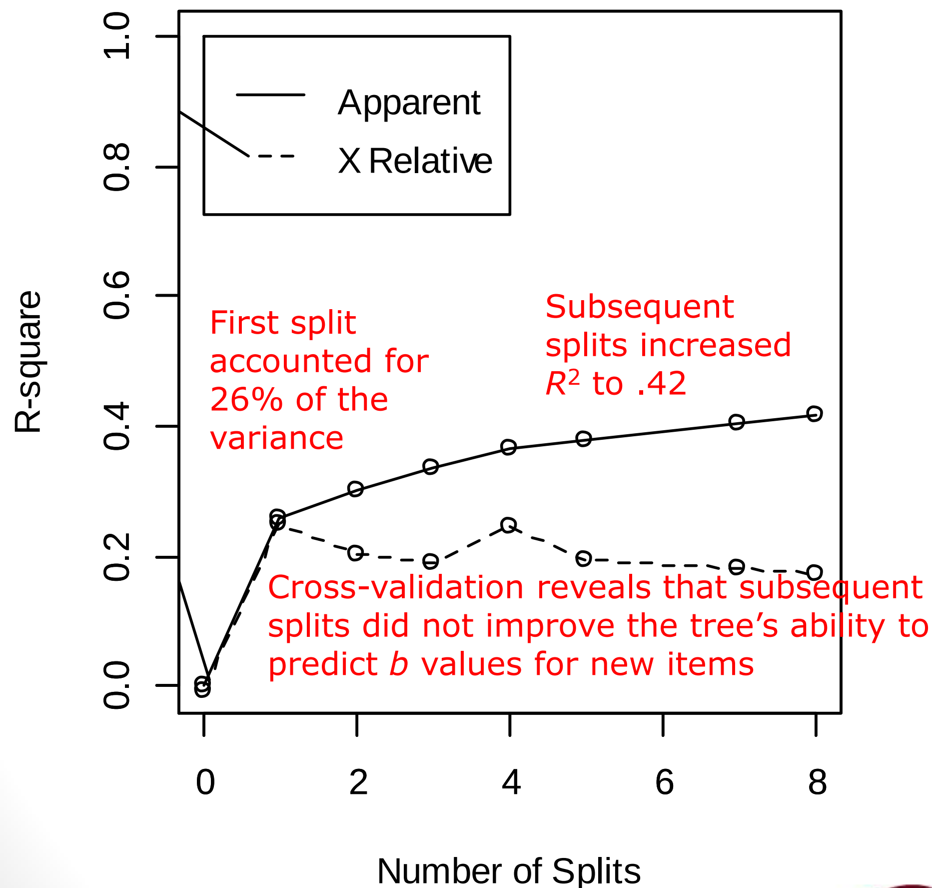
Is a person who survived a heart attack within the last 24 hours likely to die in the next 30 days?

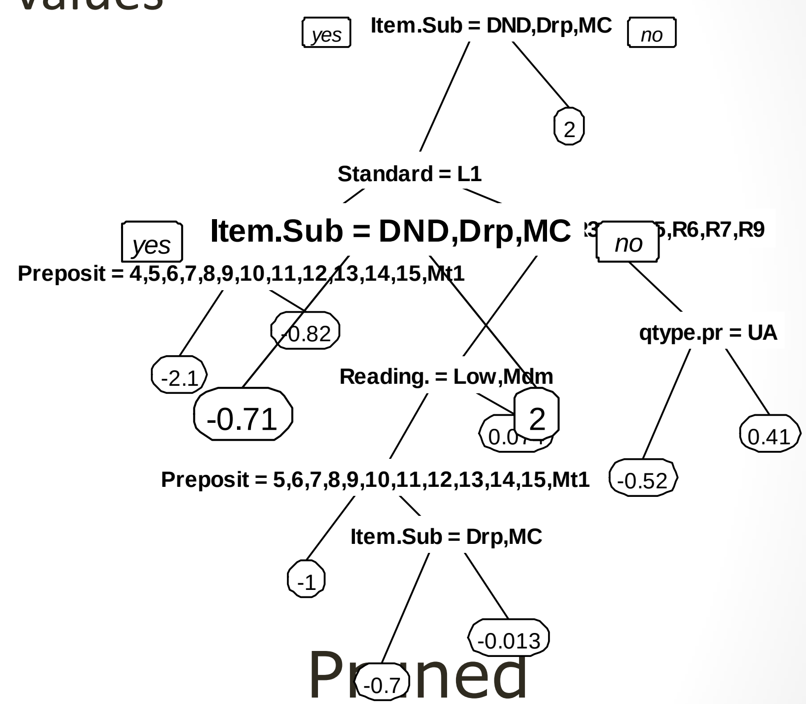
If so, relocate to ICU for constant monitoring.

Classification Tree for UC San Diego Medical Center Patients

Illustration: Reading, Program X

- Example analysis predicting IRT *b* values





Pruned Regression Tree, $R^2 = .25$

Cross-validation, R^2, pruning, and importance values

- Pruning based on cross-validation
- R^2 enables familiar, easily interpretable evaluation of the final, pruned tree
- Importance values indicate relative importance of the variables retained in the tree

Uses in large scale, summative assessments

- Item development
  - Use cognitive complexity to train item writers to hit difficulty range targets and to match item response demands with knowledge and skill requirements in Proficiency Level Descriptors
- Forms assembly
  - Assembling test forms: aligning items with KSA requirements PLDs—construct relevant item difficulty
  - Embedding new field test items in operational forms
- Item bank maintenance
  - Target cells in the bank with insufficient numbers of items
- Pre-equating
  - When item statistics are not available (e.g., in low volume programs, for embedding field test items in operational forms)

PARCC Partnership for Assessment of Readiness for College and Careers   ETS   PEARSON

Recommendations to item writers to achieve development targets

- Manipulating item response demands in construct relevant ways
- ELA
  - Creating items of high complexity is best supported by text, graphical and other stimuli with high text complexity
  - Manipulating Command of Textual Evidence levels
- Mathematics
  - Manipulating Mathematical Content and Practices demands of items is likely to enable creation of construct-relevant low, medium, and high complexity items
- These are testable hypotheses and candidates for efficacy research and validation

PARCC Partnership for Assessment of Readiness for College and Careers ETS PEARSON

Ideas for uses in classroom instruction and formative assessment

- Use text complexity measure(s) to match literary and informational texts with learners
  - PARCC: “Text” = print, digital, visuals, video, audio
- For lesson plans and quizzes, unit tests, etc.
  - E.g., explicitly target Command of Textual Evidence, Stimulus Material, and Mathematical Practices as well as Mathematical Content
- To check conceptual understanding in classroom discussions
  - E.g., use Command of Text Evidence or mathematical reasoning (a Practice) n combination with Understanding by Design (UbD) Essential Questions/understanding facets like Explanation, Interpretation, or Application

References

- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Belmont, CA: Wadsworth.
- Ferrara, S., Svetina, D., Skucha, S., & Murphy, A. (2011). Test design with performance standards and achievement growth in mind. *Educational Measurement: Issues and Practice, 30* (4), 3-15.
- Huff, K. (2003). *An item modeling approach to descriptive score reports.* (Doctoral dissertation, University of Massachusetts at Amherst, 2003). Available from Proquest, paper AAI3096285.

Partnership for Assessment of Readiness for College and Careers

Thanks!

http://parcconline.org/parcc-assessment

Enis Dogan: edogan@parcc.org

Nancy Glazer: nglazer@ets.org

Jeff Haberstroh: jhaberstroh@ets.org

Steve Ferrara: steve.ferrara@pearson.com