10-21-2011

Validating Two Problem-Solving Instruments for Use with Sixth-Grade Students

Jonathan Bostic
Bowling Green State University, bosticj@bgsu.edu

Stephen J. Pape
University of Florida

Tim Jacobbe
University of Florida, jacobbe@coe.ufl.edu

VALIDATING TWO PROBLEM-SOLVING INSTRUMENTS FOR USE WITH SIXTH-GRADE STUDENTS

JONATHAN D. BOSTIC

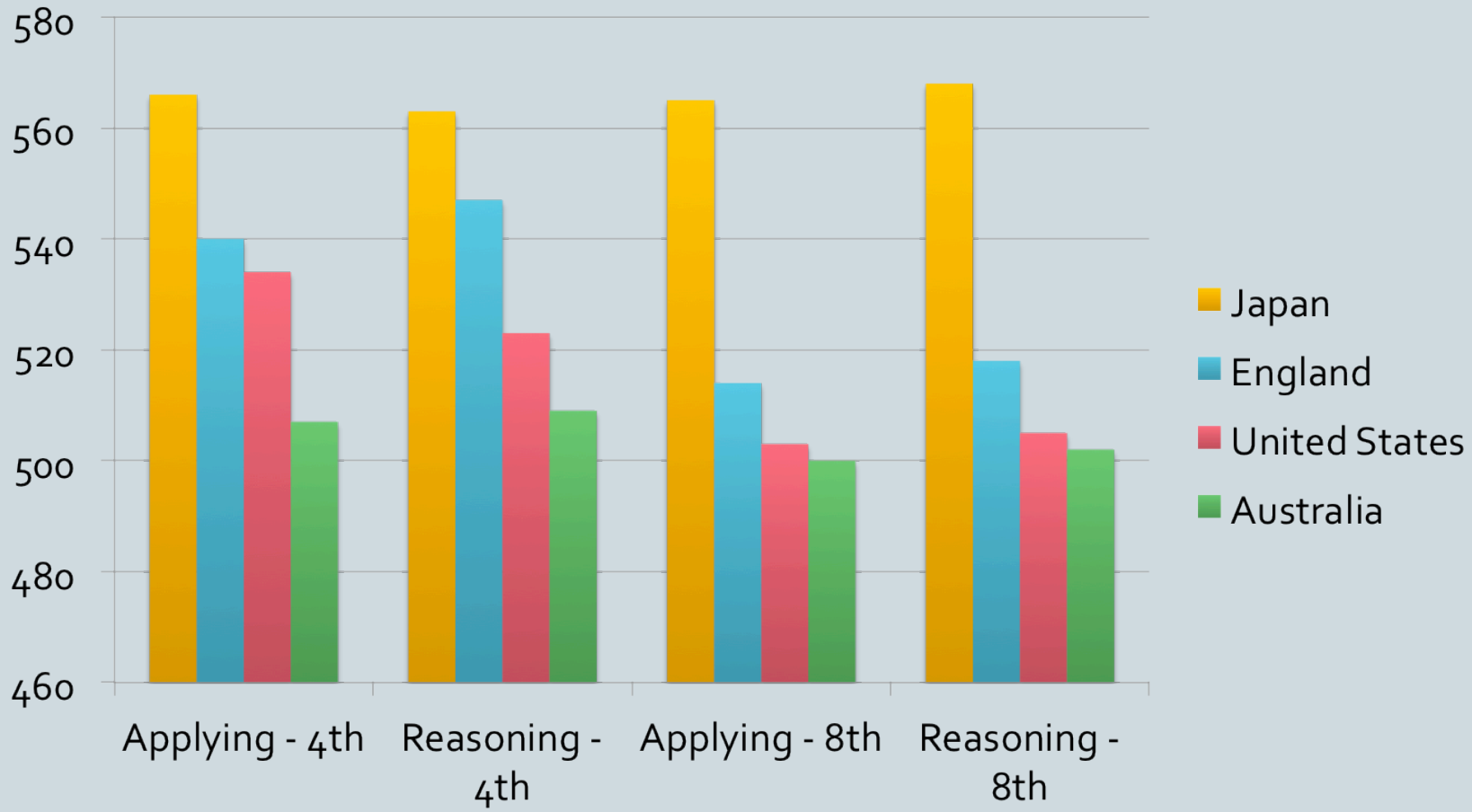
BOWLING GREEN STATE UNIVERSITY

STEPHEN J. PAPE
TIM JACOBBE

UNIVERSITY OF FLORIDA

OCTOBER 21, 2011

TIMSS 2007 results



Problem solving

- Necessary to reach a goal when an approach is not obvious (Mayer & Wittrock, 2006)
- Involves expressing, testing, and revising representations (Lesh & Zawojewski, 2007)
- Sorting, integrating, modifying, and revising/ refining mathematics from within and outside the classroom (Lesh & Zawojewski, 2007)

Item Response Theory Basics

- Ability is a unidimensional trait
- Items are locally independent
- As ability (i.e., θ) increases, then probability of correct response increases
- Item parameters are independent of respondents' abilities

Item Response Theory Modeling

- Odds of correctly answering an item = $\frac{\vartheta}{difficulty}$
- Item difficulty (b) characterizes necessary ability such that $P(\theta) = 0.5$
  - May be positive or negative
- Item discrimination (a) is the degree to which respondents with differing abilities can be distinguished
  - Good items located between $0.5 \leq a \leq 2.5$ (de Ayala, 2009)

Method

- Participants
  - N = 169
- Instrumentation
  - Translated items from Verschaffel et al. (1999)
  - Eight item pairs with updated contexts
  - Content review by mathematics educators and teacher for complex nature, realistic contexts, and opportunity to solve problems in multiple ways
- Data Collection
  - Measures completed one week apart
  - Approximately 65 and 45 minutes for pretest and posttest

Method

- Scoring using incorrect/correct categories (0/1) by two coders.
- Interrater Agreement (r_{wg}) greater than 0.9 (James, Demaree, & Wolf, 1984)
- Model fitting: Problem-solving ability
  - Chi-square, RMSEA, TLI, and CFI
- Reliability
  - Internal Consistency (Cronbach's alpha, α) and alternate-forms (Pearson's r)
- IRT modeling using Rasch constrained, Rasch unconstrained, and 2- PL
- IRT model Comparison (ANOVA)

Results

Structural Equation Modeling

- Pretest
  - Excellent Fit
    - (RMSEA = .005)
- Posttest
  - Good fit
    - (RMSEA = .021)

Reliability

- Pretest
  - α = .60
- Posttest
  - α = .62
- Alternate forms
  - r = .60

Model Comparison

Pretest	Posttest

• Rasch C vs. Rasch UC
  ○ $F(1) = 3.09$, $p = .08$
• Rasch C vs. 2-PL
  ○ $F(7) = 10.00$, $p = .19$
• Constrained Rasch model was selected.
	• Rasch C vs. Rasch UC
	  ○ $F(1) = 4.62$, $p = .03$
	• Rasch UC vs. 2-PL
	  ○ $F(7) = 15.92$, $p = .03$
	• 2-PL model was selected.

Pretest (Rasch Constrained) Results

Test Parameters	Item							
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
Difficulty	-0.49	2.82	2.47	1.74	0.99	1.65	1.69	1.53
Std. Error	0.17	0.39	0.34	0.25	0.19	0.24	0.24	0.23
Discrimination	1.24	1.24	1.24	1.24	1.24	1.24	1.24	1.24
Std. Error	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14

Posttest (2-PL) Results

Test Parameters	Item							
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
Difficulty	-0.38	4.07	3.32	0.99	0.75	1.01	1.31	0.68
Std. Error	0.38	2.13	1.38	0.17	0.19	0.27	0.34	0.13
Discrimination	0.71	1.03	0.79	2.31	1.44	1.10	1.05	2.80
Std. Error	0.25	0.67	0.38	0.71	0.38	0.31	0.31	1.02

Conclusions

- Validated measures for use with sixth-grade English-speaking students
  - Improving items two and three on both measures
- Students tended **not** to perform well on these problem-solving tasks

Future Directions

- Analyses using improved measures of internal consistency (e.g., Raykov, 2001)

- Measuring students' problem-solving ability using open, complex, and realistic tasks and aligning with Common Core State Standards

Thank you. Do you have any questions or comments?

JONATHAN D. BOSTIC
BOSTICJ@BGSU.EDU

STEPHEN J. PAPE
SJPAPE@COE.UFL.EDU

TIM JACOBBE
JACOBBE@COE.UFL.EDU

Prior problem-solving measure

- Pretest, posttest, and follow-up test (Verschaffel et al., 1999)
  - Items constructed to be parallel in nature
  - Constructed-response problem-solving items
  - Open, complex, and realistic word problems
  - No available validity-related evidence