Exploring the utility of student think-alouds for providing insights into students' metacognitive and problem-solving processes during assessment development

Deni Basaraba, Ph.D., Yetunde Zannou, Ph.D., Dawn Woods, Paul Yovanoff, Ph.D., & Leanne Ketterlin-Geller, Ph.D.

SMU. ANNETTE CALDWELL SIMMONS
SCHOOL OF EDUCATION
& HUMAN DEVELOPMENT

CHANGING MINDS

Research Questions

1. What is the relation between students' predictive, planning, and procedural metacognition and their performance on multiple-choice mathematics items?

2. Do items written to require different levels of cognitive engagement with the mathematical content explain any additional variance in student performance above and beyond the target construct of algebra readiness?

   – Do these items differ significantly in their difficulty?

Objectives for this presentation

- Describe the assessment context for this study

- Provide theoretical rationale for verbal protocols (aka "think-alouds")

- Explore the relation between metacognition and performance on mathematics items

- Describe methods for coding students' responses collected during the verbal protocols

- Present results of analyses

- Explore directions for future research

Assessment Context:
Elementary Students in Texas Algebra Ready (ESTAR) Universal Screener

- Developed as one component of a comprehensive, response-to-intervention initiative designed to increase the preparedness of students to meet standards and pass assessments in algebra in Grades 2-8 (Texas Algebra Ready)

- Components of the initiative include:

  – Online professional development academies that focus on core and supplemental mathematics instruction

  – Assessments (Universal Screener and Diagnostic)

  – Sample intervention lessons

Assessment Context:
Development of increasingly sophisticated knowledge and skills

- Learning progressions represent hypotheses about the development of students' understanding about a target construct (e.g., algebra readiness) (Duschl, Schweingruber, & Shouse, 2007)

  – Include descriptions of successively more sophisticated ways of thinking about the target construct students engage in as they learn over time (Corcoran, Mosher, & Rogat, 2007)

Assessment Context:
Development of increasingly sophisticated knowledge and skills

- For the ESTAR Universal Screener, this development is represented via two components:
    - Knowledge representation (foundational, bridging, and target content knowledge as defined by state content standards)
    - Levels of cognitive engagement (levels of cognitive processing with which students are expected to engage with the content)
- Although the content and levels of cognitive engagement increase in their sophistication and complexity, items were also written to three different levels of relative difficulty (e.g., an item could assess foundational content knowledge, target strategic competence, and be considered a relatively "easy" item)

Assessment Context:
Development of increasingly sophisticated knowledge and skills

Foundational	Bridging	Target

Prerequisite knowledge and skills that support the Target content
 	
Are accumulated from previous learning
	Needed to connect or support students' learning from the Foundational to Target knowledge and skills
	
	Often:
	    Represents integration of knowledge & skills
	    Knowledge and skills student learns during mathematics instruction
		Grade level mathematics reasoning and knowledge
		
		Supports future success in mathematics
		
		Often abstract representations of formal mathematical knowledge

Assessment Context:
Increasingly sophisticated ways of cognitively engaging with the content

- According to the National Research Council (2001) and others, mathematical proficiency requires the following:

  – Conceptual understanding (CU): Understanding of mathematical concepts and operations, the relations *between* them, and/or *why* a procedure works

  – Procedural fluency (PF): Ability to follow a sequence of certain, defined actions flexibly, efficiently, and accurately

  – Strategic competence (SC): Ability to formulate, represent, and solve mathematical problems

  – Adaptive reasoning (AR): Ability to think logically about a problem, rationalize and justify strategies used to solve the problem , and/or appropriately explain a procedure or concept

Conceptual

Which line has a slope of -3?
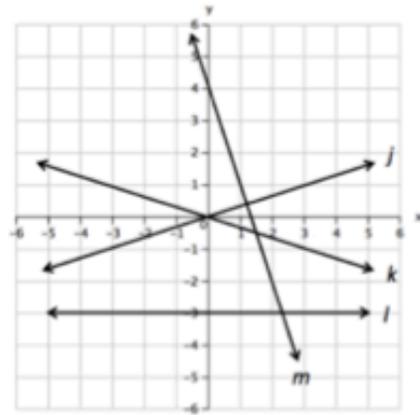
	Line *m*
	Line *j*
	Line *k*
	Line *m*

Procedural

What is the slope of the line that goes through point A and point B? A (0, 2) B (3, 0)	$-\dfrac{2}{3}$
	$-\dfrac{3}{2}$
	$\dfrac{2}{3}$
	$\dfrac{3}{2}$

Adaptive

Why is the slope from point A to point B the same as the slope from point C to point D?
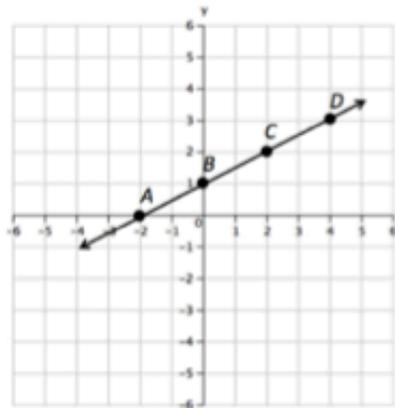
	$\overline{AB}$ and $\overline{CD}$ are on the same line.
	The distance between A, B, C, and D is the same.
	$\overline{AB}$ and $\overline{CD}$ do not overlap.
	The distance from A to B and C to D is the same.

Strategic

Which expression can be used to represent the slope of the line through point A and point B? A (3, -1) B (0, 4)	$\dfrac{4-(-1)}{0-3}$
	$\dfrac{4-1}{3+0}$
	$\dfrac{3-0}{-1-4}$
	$\dfrac{3-(-1)}{0-4}$

Limitations

Verbal Protocol Data

- Small sample, only one grade level
- Lower levels of inter-rater agreement for coding some of the questions
- Likely a relation between students' content knowledge and their ability to plan to solve a problem and follow procedures to solve that problem that weren't accounted for in our analyses
- Unable to fully examine relation between predictive metacognition and student performance

Bifactor Models

- Only one grade level
- Sparse data matrix as a result of assessment pilot design made it impossible to examine the contributions of CU, PF, SC, and AR without first considering knowledge representation
- Imbalanced number of items per each level of mathematical proficiency likely to influence the amount of variance that each can explain in the overall model

What are Verbal Protocols?

- Process of having students "think-aloud" while completing a task

- Students are asked to say what they are looking at, thinking, and doing (including strategies they are using) while completing a task

> I am going to ask you to solve some math problems and to talk about how you solved the problems, just like you do in class. We are interested in understanding the thinking you use while solving math problems. Today, I want you to say **all** of your thoughts about how you came to your answer, rather than thinking them in your head.

- Goal: To see first-hand the *process* of task completion, rather than just the final product

Theoretical Rationale for Verbal Protocols

- Can be useful during the test development process because having students "think-aloud" while solving problems can provide information about

  – Cognitive processes students were engaged in while solving the problem (Ericsson & Simon, 1993)

  – Students' understanding (or misunderstanding) of the content and constructs being assessed (Almond et al., 2009)

  – Whether items are of comparable difficulty or understood similarly for students from different demographic subgroups (Ercikan et al., 2010)

Verbal Protocols and Metacognition:
Insights into students thinking about their thinking

- **Researchers** (Deseote, 2009; Deseote, Roeyers, & Buysse, 2001; Jacobse & Harskamp, 2012) have used verbal protocols to explore the relationship between procedural, predictive, and planning metacognition and students' performance on mathematics problems

  - <u>Predictive:</u> Congruence between students' prediction of their selecting the correct response and whether they selected the correct response

  - <u>Planning:</u> Ability to articulate what you would need to do to solve the problem (identification and application of problem-solving steps)

  - <u>Procedural:</u> Accurate understanding of the procedures and strategies needed to solve the problem using the information given

- Findings include (a) moderate correlations between metacognition and mathematics performance and (b) that indicators of metacognition can explain some of the observed variability in student math performance

SMU. ANNETTE CALDWELL SIMMONS SCHOOL OF EDUCATION & HUMAN DEVELOPMENT

CHANGING MINDS

RQ 1 What is the relation between students' predictive, planning, and procedural metacognition, and their performance on multiple-choice mathematics items?

Participants: 10 4th grade students with varying mathematics ability

Procedures:

– Asked students to solve 10 multiple-choice mathematics items

– Had students respond to 10 retrospective think-aloud questions after solving each problem (coded 5)

– Developed a rubric for each item outlining the expected components of responses of students demonstrating *Exemplary, Proficient, Developing,* and *Emerging* understanding of the assessed content

– Had 2 independent reviewers code students' responses targeting predictive, planning, and procedural metacognition for each item

Methods for Coding Student Responses

- Predictive Metacognition
  - After students read a problem but *before* they solved it we asked students to rate the likelihood that they would select the correct response: *Will you answer the problem correctly?* (Very Likely, Likely, Unlikely, Very Unlikely)
  - Coded for agreement between prediction and selection of correct response (Deseote et al., 2001)
    - (2): Complete congruence (Very Likely/Correct or Very Unlikely/Incorrect
    - (1): Partial Congruence (Likely/Correct or Unlikely/Incorrect)
    - (0): No congruence (Very Likely/Incorrect, Likely/Incorrect, Very Unlikely/Correct, Unlikely/Incorrect)

Methods for Coding Student Responses

- Planning Metacognition

  - Defined as the ability to analyze problems, retrieve relevant domain-specific knowledge, and then identify, sequence, and apply the problem-solving strategies needed to solve the problem (Deseote et al., 2001)

  - During the retrospective think-aloud, we asked students:

    - Q3: What is this problem asking you to do?

    - Q4: What information do you need to solve the problem?

Methods for Coding Student Responses

- Procedural Metacognition

    - Defined as one's knowledge of the methods or strategies needed to achieve one's goals (i.e., solving the problem), understanding *how* those strategies work, and how they can be applied to solve the problem (Deseote et al., 2001; Montague, 1992)

    - During the retrospective think-aloud, we asked students:

        - Q5: What strategies and steps did you take to solve the problem?

        - Q6: Does your answer for this problem make sense? Why?

Methods for Coding Student Responses

- Although not directly related to metacognition, there is reason to believe that students' understanding of the content might mediate his/her ability to identify or develop a plan to solve the problem

  - During the retrospective think aloud, we asked students:

    - Q2: What do you know about [content assessed by the problem]?

  - We identified the critical information needed to solve the problem and examined the extent to which students' responses to these questions included that information

VP Question	Exemplary	Proficient	Developing	Emerging
What do you know about _____?	S provides a complete explanation of the topic and/or an accurate example	S provides a partial explanation of the topic and/or a partial example	S describes the topic with related language and/or provides and example that does not reflect conceptual understanding	Student is not able to explain or describe the topic
What is this problem asking you to do?	S identifies and applies all mathematical concepts needed to solve	S identifies all mathematical concepts needed to solve	S identifies some of the mathematical concepts needed to solve	S is not able to identify any mathematical concepts needed to solve
What information do you need to solve the problem?	S identifies and interprets all mathematical information needed to solve	S identifies all mathematical information needed to solve	S identifies some of the mathematical information needed to solve	S is not able to identify any mathematical information needed to solve
What strategies and steps did you take to solve the problem?	S identifies all of the mathematical steps needed to solve	S partially identifies the mathematical steps needed to solve	S identifies a non-mathematical strategy to solve	S does not identify a strategy to solve
Does your answer for this problem make sense? Why?	S clearly justifies response by using mathematical reasoning	S partially justifies response by using mathematical reasoning	S does not provide a mathematical justification	S is not able to describe or justify his/her response

Inter-rater Agreement

Question	N	Cohen's K	95% Confidence Interval	
			Lower	Upper
Q2. What do you know about _____?	97	0.42	0.29	0.55
Q3. What is the problem asking you to do?	98	0.63	0.49	0.76
Q4. What information do you need to solve the problem?	98	0.23	0.11	0.36
Q5. What strategies and steps did you take to solve the problem?	98	0.54	0.41	0.67
Q6. Does your answer for the problem make sense? Why?	98	0.39	0.25	0.52

RQ 1 Results: Correlations

	Total Correct	Predictive Congruence	Q2 ratings	Q3 ratings	Q4 ratings	Q5 ratings	Q6 ratings	Planning Metacognition	Procedural Metacognition
1	--	0.36	0.29	0.55	**0.78****	**0.62**	0.44	**0.68***	0.23
2		--	0.002	0.16	0.47	0.37	0.28	0.35	0.19
3			--	0.53	0.51	**0.83****	**0.87****	0.25	**0.68***
4				--	0.59	**0.70***	**0.68***	**0.64***	0.25
5					--	**0.67***	0.56	**0.73***	0.20
6						--	**0.95****	0.46	**0.73***
7							--	0.42	**0.78****
8								--	0.27
9									--

* *p* < 0.05, ** p < 0.01

RQ 1 Results: Regression

- Due to limited number of cases, 3 single, linear regression models were conducted with the total number of items correct as the outcome and a type of metacognition as a predictor

Predictor	F	p	b	t	p	Adjusted R^2
Predictive Congruence	1.17	0.31	0.10	1.08	0.31	0.19
Planning Metacognition	6.99	0.03	0.14	2.64	0.03	0.40
Procedural Metacognition	0.45	0.52	0.03	0.67	0.52	- 0.07

Interpreting RQ 1 Results

- Our indicators of planning metacognition (Q3 & Q4) had the strongest relation with students' performance on the multiple-choice mathematics items

- Contrary to previous research (Deseote, 2009; Jacobse & Harskamp, 2012) our indicator of predictive metacognition, although similar to that used in previous studies, was not significantly related to students' performance on the items

- Although not all of our metacognition indicators were significantly related to students' performance on the items, the moderate to strong correlations between two of the three types of metacognition ($r = 0.67\text{-}0.95$), are consistent with previous research (Deseote, 2009; Jacobse & Harskamp, 2012)

Limitations

Verbal Protocol Data

- Small sample, only one grade level
- Lower levels of inter-rater agreement for coding some of the questions
- Likely a relation between students' content knowledge and their ability to plan to solve a problem and follow procedures to solve that problem that weren't accounted for in our analyses
- Unable to fully examine relation between predictive metacognition and student performance
  - Explore relation between prediction, evaluation, and/or indicators of persistence

Next Steps

Verbal Protocol Data

- Examine rationales for G4 ratings to see if there are specific reasons why the inter-rater agreement was low for specific items; revise rubric if necessary

- Analyze G2 and G3 data

- Consider revising indicator of predictive metacognition to include evaluation (confidence in selection of correct response after solving the problem), which is consistent with prior research (Deseote, 2009; Jacobse & Harskamp, 2012) and students' rating of their perceived difficulty of the item

- Consider logistic regression as alternative analysis to examine the relation of these predictors to students' performance on each item

RQ 2 Do items written to require different levels of cognitive engagement with the mathematical content explain any additional variance in student performance above and beyond the target construct of algebra readiness?

- Participants: 2,548 students who responded to between 22-25 multiple-choice items written for the ESTAR US item bank (206 Grade 4 items)

- Procedures:
  - IRT item difficulties estimated with a bifactor model in Testfact
    - Bifactor model includes a general, underlying latent factor (algebra-readiness) and four domain-specific factors (CU, PF, SC, AR) – 1 model for each knowledge representation (F, B, T)
  - Results of the bifactor model were compared to a single-factor model that hypothesized a single, underlying latent factor (algebra readiness)
  - Conducted non-parametric tests using item difficulties to test for differences in mean and median ranks of four item types (CU, PF, SC, AR)

Proposed Bifactor Model
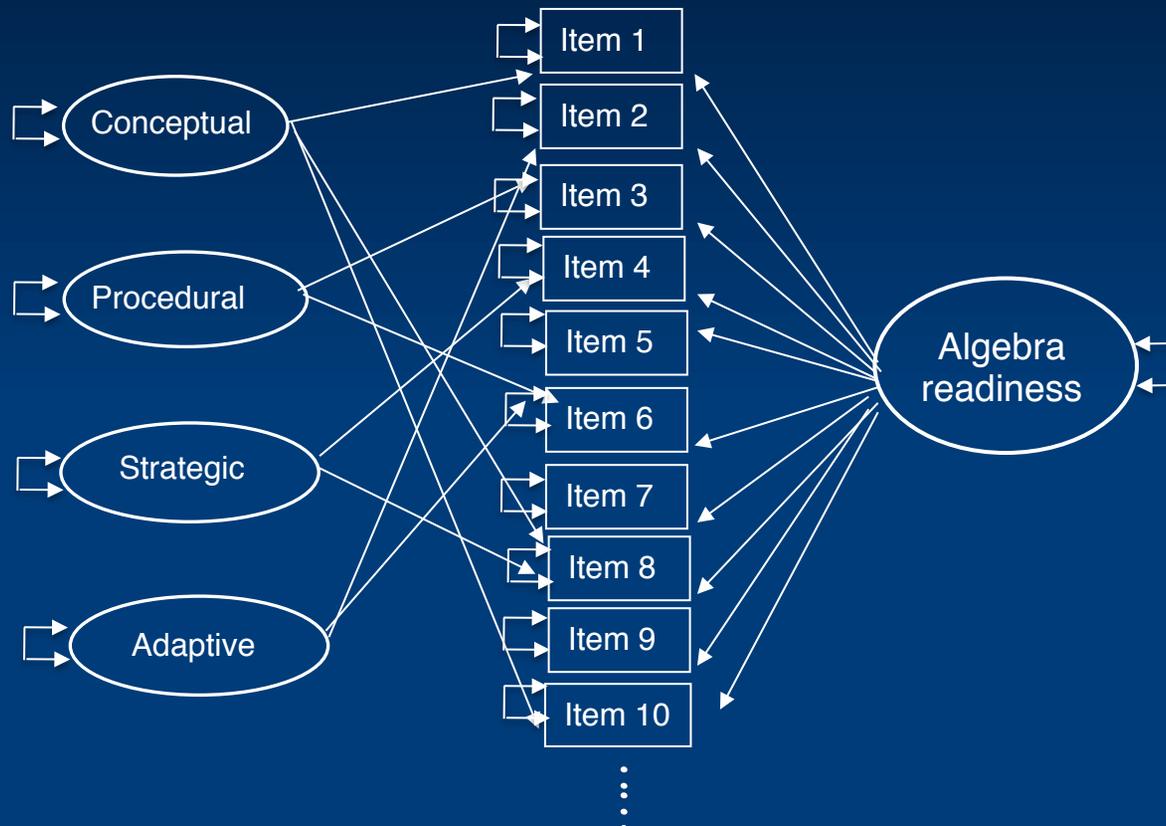
- Hypothesized general factor (algebra readiness) to account for commonality of items

- Multiple domain specific factors thought to account for unique variance over and above general factor

- Interest in domain specific factors in addition to general factor interest

Chen, West, & Sousa (2006)

RQ2: Model Comparisons

- Chi-square difference tests comparing the model fit of the unidimensional, single-factor and bi-factor model revealed that the bifactor model fit the data better for all three knowledge representations:

Knowledge Representation	Bifactor χ^2	Single Factor χ^2	χ^2 Difference	Difference df	p
Foundational	22727.3	17106.9	5620.37	69	< 0.001
Bridging	31660.6	26408.4	5252.22	71	< 0.001
Target	12479.6	8337.05	4142.53	58	< 0.001

RQ2: Bifactor Model Results

Knowledge Representations	Percent of Variance Explained					Additional variance explained by specific factors
	General Factor	Specific Factor 1 (CU)	Specific Factor 2 (PF)	Specific Factor 3 (SC)	Specific Factor 4 (AR)	
Foundational	33.55%	13.00%	7.94%	1.65%	0.20%	22.34%
Bridging	35.26%	9.51%	3.51%	2.96%	0.15%	16.13%
Target	34.00%	5.57%	7.78%	3.86%	0.35%	17.56%

RQ2: Nonparametric Test Results

- Kruskal-Wallis tests were conducted to examine the equality of population medians (and means) among groups of items

    – Items are rank-ordered with respect to item difficulties (low to high)

    – Item types (CU, PF, SC, and AR) are compared with respect to rank

    – If item types have different median or mean values, then item types are different with respect to item difficulties

- One-way ANOVAs were conducted to test for statistically significant differences across the means of item difficulty types

- Results indicated no significant differences across the item types

Interpreting RQ2 Results

- Comparisons of the single factor and bifactor model revealed that the bifactor fit the data better and that a not insignificant amount of variance in student performance could be explained by the domain-specific factors that represent 4 of the 5 strands of mathematical proficiency

    – Given that items were written to reflect specific differences in the strands of mathematical proficiency (e.g., PF items required students to accurately carry out computations or follow a sequence of steps while SC items required students to identify and use an appropriate strategy to solve a problem), the results indicate that differences in content knowledge alone may not explain variability observed in student performance

Limitations

Bifactor Models

- Only one grade level

- Sparse data matrix as a result of assessment pilot design made it impossible to examine the contributions of CU, PF, SC, and AR without first considering knowledge representation

- Imbalanced number of items per each level of mathematical proficiency likely to influence the amount of variance that each can explain in the overall model

Next Steps

Model comparisons

- Examine the percent of variance within each item type (CU, PF, SC, and AR) to see how much variance is explained by the general factor and that specific domain factor (decrease the influence of "weighting" due to different numbers of items for each item type)

- Conduct analyses with G2 and G3 data

- Conduct analyses with data from the three grades collected with the "live" test forms comprised of 24 items that are administered in Fall, Winter, and Spring starting the 2013-2014 school year

References

Almond, et al. (2009). *White paper: Cognitive interview methods in reading test design and development for alternate assessments based on modified academic achievement standards (AA-MAS).* Dover, NH: Measured Progress.

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research, 41,* 189-225.

Corcoran, T., Mosher, F., & Rogat, A. (2010). *Learning progressions in science: An evidence-based approach to reform.* (CPRE Research Report #RR-63). Consortium for Policy Research in Education

Deseote, A. (2009). Metacognitive prediction and evaluation skills and mathematical learning in third-grade students. *Educational Research and Evaluation: An International Journal on Theory and Practice, 15,* 435-446.

Deseote, A., Roeyers, H., & Buysse, A. (2001). Metacognition and mathematical problem solving in grade 3. *Journal of Learning Disabilities, 34,* 435-448.

SMU. ANNETTE CALDWELL SIMMONS SCHOOL OF EDUCATION & HUMAN DEVELOPMENT

CHANGING MINDS

References

Ercikan et al. (2010). Application of think aloud protocols for examining and confirming sources of differential item functioning identified by expert reviews. *Educational Measurement: Issues and Practice, 29*(2), 24-35.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as Data.* Cambridge, MA: MIT Press.

Jacobse, A. E., & Harskamp, E. G. (2012). Toward efficient measurement of metacognition in mathematical problem solving. *Metacognition Learning, 7,* 133-149.

Montague, M. (1992). The effects of cognitive and metacognitive strategy instruction on mathematical problem solving of middle school students with learning disabilities. *Journal of Learning Disabilities, 25,* 230-248.

Questions?

Deni Basaraba, Ph.D.
Assessment Coordinator
dbasaraba@smu.edu

Research in Mathematics Education
Annette Caldwell Simmons School of Education and Human Development
Southern Methodist University
PO Box 750114
Dallas, TX 75275-0114

http://www.smu.edu/Simmons/Research/RME