Proceedings for the 51st Annual Meeting of the Research Council on Mathematics Learning

Celebrating 51 years of Research on Mathematics Learning

February 29 – March 2
Columbia, South Carolina

RCML Officers & Conference Team

PRESIDENT, 2023-2025
Sarah Pratt
Baylor University
Waco, TX
Sarah_Pratt@baylor.edu

PAST PRESIDENT, 2023-2024
Travis Olson
University of Nevada at Las Vegas
Las Vegas, NV
travis.olson@unlv.edu

VICE PRESIDENT FOR CONFERENCES, 2022-2024
Jamaal Young
Texas A&M University
College Station, TX
jamaal.young@tamu.edu

VICE PRESIDENT FOR PUBLICATIONS, 2021-2024
Colleen Eddy
University of North Texas
Denton, TX
colleen.eddy@unt.edu

VICE PRESIDENT FOR PUBLICATIONS – ELECT, 2023-2024
Babette M. Benken
Cal State University Long Beach
Long Beach, CA
babette.benken@csulb.edu

TREASURER, 2022-2026
Tyrette Carter
North Carolina A&T University
Greensboro, NC
tscarte1@ncat.edu

SECRETARY, 2023-2025
Tina Mitchell
Delaware State University
Dover, DE
tmitchell@desu.edu

HISTORIAN
William Speer
University of Nevada at Las Vegas
Las Vegas, NV
william.speer@unlv.edu

INVESTIGATIONS EDITOR
Jonathan Bostic
Bowling Green State University
Bowling Green, OH
jbostic@bgsu.edu

INVESTIGATIONS ASSOCIATE EDITORS
Drew Polly
University of North Carolina at Charlotte
Charlotte, NC
abpolly@uncc.edu

Jennifer Cribbs
Oklahoma State University
Stillwater, OK
jennifer.cribbs@okstate.edu

NEWSLETTER CO-EDITORS
Tina Mitchell (2021-Present)
Delaware State University
Dover, DE
tmitchell@desu.edu

Byungeun Pak (2023-Present)
Utah Tech University
St. George, UT
b.pak@utahtech.edu

PROCEEDINGS EDITOR
Dennis Kombe
Cal State University, Monterey Bay
Seaside, CA
dkombe@csumb.edu

PROCEEDINGS CO-EDITOR
Ann Wheeler
Texas Woman's University
Denton, TX
AWheeler2@twu.edu

CONFERENCE COMMITTEE
Seanyelle Yagi (2021-2024)
University of Hawaii at Manoa
Honolulu, HI
slyagi@hawaii.edu

Barbara Allen-Lyall (2021-2024)
Manhattanville College
Purchase, NY
barbara.allenlyall@mville.edu

Nickolaus Ortiz (2022-2025)
Georgia State University
Atlanta, GA
nortiz1@gsu.edu

Karen Zwanch (2022-2025)
Oklahoma State University
Stillwater, OK
karen.zwanch@okstate.edu

Lucas Foster (2023-2026)
Northeastern State University
Broken Arrow, OK
fosterlb@nsuok.edu

Jennifer Gonzales (2023-2026)
Northside Independent School District
San Antonio TX
Jenny.Gonzalez@nisd.net

2024 Local Host Co-Chairs
Sean Yee
University of South Carolina
Columbia, SC
yee@math.sc.edu

Christie Martin
University of South Carolina
Columbia, SC
martinc1@mailbox.sc.edu

THANK YOU TO OUR REVIEWERS

Babette Benken	Tonya Garrett	Alesia Moldavan
Ken Butler	Nico Gòmez Marchant	Ayse Ozturk
Georgia Cobbs	Andrew Goodman	Byungeun Pak
Angie Crawford	Pat Jordan	Alyssa Sayavedra
Jen Crooks Monastra	Puskar Joshi	Dawn Teuscher
Colleen Eddy	Pavneet Kaur Bharaj	Carrie Toreky
Kristian Edosomwan	Karl Kosko	Sandra Trowell
Brian Evans	Karl Kruczek	Patty Wagner
Timothy Folger	Melinda Lanius	Ann Wheeler
Lucas Foster	Hope Marchionda	Nicholas Wong
Ryan Fox	Tina Mitchell	Sean Yee
		Alan Zollman

For citation of articles in the Proceedings:

Authors. (2024). Title of paper. In D. Kombe, & A. Wheeler (Eds.), *Proceedings of the 51st Annual Meeting of the Research Council on Mathematics Learning.* (pp.##-##). RCML.

Editorial Assistance:

Pavneet Kaur Bharaj, California State University, Bakersfield, CA.

Graduate Student Editorial Assistant:

Jennifer T. H. Hylemon, University of North Texas, Denton, TX

Publication Acceptance Rate:

Accepted 14 manuscripts out of 26 submissions. Acceptance Rate of 54% (last year 48%)

RCML History

The Research Council on Mathematics Learning, formerly The Research Council for Diagnostic and Prescriptive Mathematics, grew from a seed planted at a 1974 national conference held at Kent State University. A need for an informational sharing structure in diagnostic, prescriptive, and remedial mathematics was identified by James W. Heddens. A group of invited professional educators convened to explore, discuss, and exchange ideas especially in regard to pupils having difficulty in learning mathematics. It was noted that there was considerable fragmentation and repetition of effort in research on learning deficiencies at all levels of student mathematical development. The discussions centered on how individuals could pool their talents, resources, and research efforts to help develop a body of knowledge. The intent was for teams of researchers to work together in collaborative research focused on solving student difficulties encountered in learning mathematics.

Specific areas identified were:

1. Synthesize innovative approaches.
2. Create insightful diagnostic instruments.
3. Create diagnostic techniques.
4. Develop new and interesting materials.
5. Examine research reporting strategies.

As a professional organization, the **Research Council on Mathematics Learning (RCML)** may be thought of as a vehicle to be used by its membership to accomplish specific goals. There is an opportunity for everyone to actively participate in **RCML**. Indeed, such participation is mandatory if **RCML** is to continue to provide a forum for exploration, examination, and professional growth for mathematics educators at all levels.

The Founding Members of the Council are those individuals that presented papers at one of the first three National Remedial Mathematics Conferences held at Kent State University in 1974, 1975, and 1976.

Table of Contents

Framing Equitable Learning

COMPUTER ADAPTIVE MATHEMATICAL PROBLEM-SOLVING MEASURE: A BRIEF VALIDATION REPORT

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

Timothy Folger
Bowling Green State University
tdfolge@bgsu.edu

Kristin Koskey
Drexel University
kk3436@drexel.edu

Gabriel Matney
Bowling Green State University
gmatney@bgsu.edu

Toni A. May
Drexel University
tas365@drexel.edu

Gregory Stone
Clarity Assessments
gregory@clarityassessmentsystems.com

The purpose of this proceeding is to share a component to a validity argument for a new, computer adaptive mathematics Problem-Solving Measure that is designed for grades six through eight (PSM 6-8). The PSM is a single test, which uses computer adaptive features to measure students' performance using instructional standards. It is intended to measure students' problem-solving performance related to instructional standards.

Introduction

In prior studies (e.g., Bostic & Sondergeld, 2015; Bostic et al., 2017; Bostic et al., 2021), our research team described a series of vertically equated, paper-and-pencil measures of mathematical problem solving for grades 3-8. The measures draw upon guidance from the *Standards for Educational and Psychological Testing* (*Standards*; AERA et al., 2014), and align to the Common Core State Standards for Mathematics (CCSSI, 2011) for both content and practice. The tests filled a need for K-12 educators, researchers, and evaluators. K-12 educators recently requested, partially due to their experiences during COVID-19, a version that could be administered online, and could be adaptive to students' abilities. This presented an opportunity to develop a computer-adaptive version of the tests. This study's purpose is to present validity evidence based on test content and consequential (i.e., bias) considerations for a computer-adaptive test (CAT) of mathematical problem-solving for grades six through eight (CAT PSM 6-8).

Relevant Literature

Problems and Problem Solving

Our research team drew upon two related frameworks for mathematical problems. Schoenfeld (2011) frames a mathematical problem as a task presenting to an individual such that (a) it is unclear whether there is a solution, (b) it is unknown how many solutions exist, and (c) the pathway to the solution is unclear. Verschaffel and colleagues (1999) frame mathematical word problems as tasks that are (a) open, (b) complex, and (c) realistic. Open tasks can be solved using multiple developmentally-appropriate strategies. Complex tasks are not readily solvable by a respondent and require productive thinking. Notions of open and complex are clearly related to Schoenfeld's framing of problems. The use of realistic adds a necessary element to effectively frame word problems for our assessment. Realistic word problems draw upon real-life experiences, experiential knowledge, and/or believable events. Schoenfeld (2011) and Verschaffel et al.'s (1999) frameworks provided our team with sufficient grounding to develop mathematical word problems for the CAT PSM 6-8.

Given our selection of two synergistic frameworks for creating CAT items, we retained Lesh and Zawojeski's (2007) problem-solving framework from past test development. That is, problem solving is a process of "several iterative cycles of expressing, testing and revising mathematical interpretations – and of sorting out, integrating, modifying, revising, or refining clusters of mathematical concepts from various topics within and beyond mathematics" (Lesh & Zawojewski, 2007, p. 782). Through these frames of mathematical word problems and problem solving, our team sought to develop the CAT PSM 6-8.

Validity and Validity Arguments

Validation studies are intended to provide a reader with information about how evidence supports an intended interpretation and use of test results (AERA et al., 2014; Carney et al., 2022; Kane, 2006, 2012). The *Standards* defines validity as "the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests" (AERA et al., 2014, p. 11). The *Standards* describe five sources of validity evidence: test content, response process, internal structure, relations to other variables, and consequences of testing. More information about these five sources is discussed in Folger et al. (2023). This proceeding focuses on test content and consequential evidence to explore three validity claims. The first claim is that the CAT PSM 6-8 items address mathematics content described in the CCSSM and have a known mathematical solution space. Our second claim is that CAT PSM 6-8 items adhere to the open, complex, and realistic framework. A third claim is that CAT PSM 6-8 items possess limited bias. Our broad research question is: What test content and consequential evidence supports our claims regarding CAT PSM development?

Method

A design science approach (Lesh & Sriraman, 2005; Middleton et al., 2008) was used to develop the CAT PSM 6-8. Design science research is valuable for creating products like tests that can be evaluated, refined, and re-evaluated (Middleton et al., 2008). Our test item development and subsequent validation process had multiple checks and balances across a diverse research team including mathematics educators, assessment scholars, psychometricians, and graduate students across all three areas. Recent scholarship indicates a deductive, a priori, approach for developing validity claims and then collecting validity evidence leverages the theoretical foundation from which the test is developed (Folger et al., 2023); hence, we use that approach in this study. Note, this study focuses on validity evidence based on test content and consequential considerations (i.e., bias) because (a) of similar data collection and analysis method and (b) these claims can be described within the proceedings page limitations.

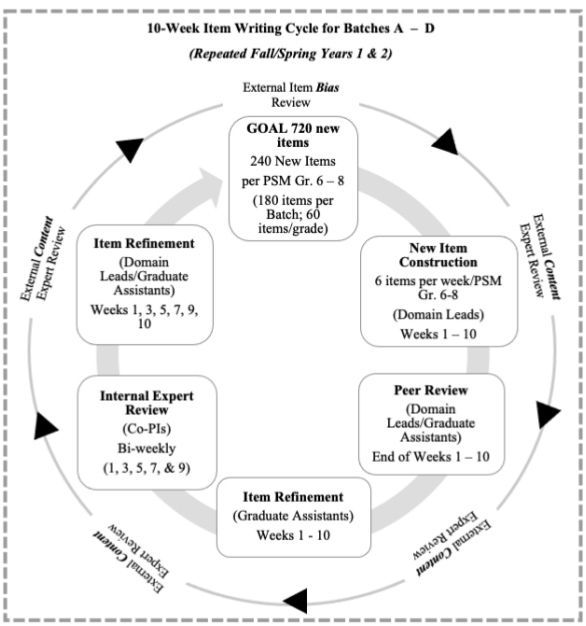
Participants and Instrumentation

CAT PSM 6-8 items were developed using an iterative design (see Figure 1). Item developers included mathematics teachers across the USA certified to teach grades 6-12. All teachers hold graduate degrees and valid teaching credentials in their state. Content panel experts included two terminally degreed (PhD) mathematics educators, mathematics education graduate students, as well as two terminally degreed (PhD) mathematicians who had expertise with instructional standards (e.g., CCSSI, 2011). Bias panel experts included two terminally degreed mathematics educators, an assessment scholar and doctoral student, as well as four purposefully recruited mathematically focused bias panel members representing diversity, equity, and inclusivity through different ethnicities, cultural backgrounds, and geographic regions of the USA. All bias panel members hold graduate degrees and work in education or education-adjacent fields (e.g., engineering). Also, students from participating districts were asked to review and respond to two questions following their work on three CAT items during 1-to-1 think alouds: (1) Do you believe this item is appropriate for other students in your grade level? (2) Do you feel there is any form of bias in the item that you completed? Participating school districts included over 500 diverse students from the Midwest, Southwest, and Mountain West regions, including rural, suburban, and urban school contexts. One group of students from the Mountain West region included a large number of multi-lingual learners.

A goal of CAT PSM 6-8 development was to develop 240 items for each grade level (grades six, seven, and eight). After the multilevel reviews during the item development phase (see Figure 1), the final item pool consisted of a total of 178 items, 178 items, and 182 items all aligned to their respective grade-level mathematics content standards. An example seventh-grade Expressions and Equations item is provided to contextualize the word problems created for the CAT PSM: “A water tower contains 16,880 gallons of water. Each day half of the water in the tank is used and not replaced. This process continues for multiple days. How many gallons of water are in the 4tower at the end of the fourth day?” One example of a reason that an item did not make the final pool of items was that some items could not be further shortened in text length without losing realism and complexity. Similar to past paper-and-pencil PSMs, CAT PSM 6-8 items are scored dichotomously as correct or incorrect.

Figure 1.

CAT PSM 6-8 item writing cycle process



Data Collection and Analysis

This study focuses on test content and consequential validity evidence; further validity claims and evidence will be presented in future scholarship. Gathering evidence based on test content involved multiple stages of data collection, which were analyzed and reviewed in light of each other to triangulate findings (Saldaña, 2013). Data sources are described in the order conducted. One source came from middle school teachers’ reviews. Teachers were instructed to review items to confirm that each item (a) addressed the mathematics content and practice standards indicated by the item writer (CCSSI, 2011), (b) adhered to our selected frames for mathematical word problems (i.e., open, complex, and realistic), and (c) were developmentally appropriate. A second data source came from the mathematicians who conducted reviews to explore the degree to which items (a) addressed the mathematics content and practice standards indicated by the item writer (CCSSI, 2011), (b) adhered to our selected frames for mathematical word problems (i.e., open, complex, and realistic), and (c) had a known mathematical solution space. A third data source was a review conducted by two mathematics educators and multiple mathematics education graduate students. That review paralleled prior reviews in that it combined elements from the mathematicians' and the mathematics teachers' reviews. A fourth data source came from students' feedback during think alouds regarding grade-level appropriateness of items.

Bias was reviewed in a similar fashion but with different individuals. Again, the data sources order coincides with the steps in the process. The first data source was a review conducted by practicing mathematics teachers. The goal was to explore the ways in which items might contribute negative bias towards students' outcomes. A second data source came from a review conducted by a bias panel led by a psychometrician and assessment-focused graduate student. This panel followed a protocol developed by the team, based upon past work on a similar project. A third data source came from a review conducted by two mathematics educators, two assessment scholars, and several mathematics education graduate students. This team reviewed items for potential bias, feedback was shared, and revisions were made. A fourth and final data source was students' responses to the bias question asked during 1-1 think alouds.

Data sources were analyzed qualitatively for evidence in support of conjectured claims. Our team used an iterative, inductive analysis with a goal of generating themes (Hatch, 2002; Saldaña, 2013). Step one was becoming familiar with the available data for analysis. Step two was to more closely examine data sources to clarify any ambiguity that arose during the first review of data. Step three was making notes about potential ideas that seemed relevant to the claims, based upon the data sources for each validity source. Step four aimed to categorize notes into general notions, which had potential to become a claim. Step five was discussions about categories that might be eliminated or revised based upon the findings occurred. Step six was to review each category and consider the amount and quality of evidence related to it, which made generate a validity claim. Those categories with two or more pieces of counterevidence or a paucity of evidence were removed. Step seven involved synthesizing those categories into support for the a priori validity claims.

Findings

Test content and consequential validity evidence are presented in relation to validity claims. Based on the study findings three validity claims were generated: (1) CAT PSM 6-8 items address mathematics content described in the CCSSM and have a known mathematical solution space; (2) CAT PSM 6-8 items adhere to the open, complex, and realistic framework; and (3) CAT PSM 6-8 items have evidence of limited bias.

Claim #1: Mathematics content

There was consistent and resounding support that the final drafts of items were aligned to CCSSM standards and had a known mathematical solution space. Closure was important because if there were multiple solution sets, then scoring dichotomously could be problematic. Regarding standards alignment, initial reviews during the item writing stages flagged certain items for further revisions. Flagged items were revised and demonstrated full content alignment with standards. As one example, one review on an Expressions and Equations item suggested that the item changed from "the most" to "the greatest number". The latter statement is mathematically accurate whereas 'most' does not necessarily imply larger numbers.

Claim #2: Open, complex, and realistic

Content panel members consistently indicated that items could be solved with two or more developmentally appropriate strategies. In some cases, teachers and mathematicians shared four unique strategies. As one example, a mathematician shared numerous strategies for an eighth-grade functions item that spanned different representations (i.e., symbols, graphs, and tables), as well as different procedures using those representations. Items were also complex enough such that teachers believed potential test takers would need time to reflect on a viable solution strategy pathway. In many cases, teachers emphasized the need for respondents to reflect on the problem's situational context, then connect ideas to mathematical strategies, and ultimately act on a strategy that had potential to arrive at a solution. Realism was often discussed among panel members. During later stages of data collection, items were deemed to need further revisions because some contexts may be real to one group of students (e.g., a cell phone bill) but not others. Some students shared that a cell phone plan may have limited minutes but that many students might not understand specific components of a cell phone bill (e.g., overage). For instance, Elias was a multilingual student from an urban Mountain West district. He shared during a think aloud that "I don't think most students understand that minutes cost money. Most students don't pay their own cell phone bill. Also, many people just have unlimited minutes like on the commercials." Items like this one were revised and resent to panel members for review to confirm that changes were adequate. In this case, the cell phone bill item was revised to focus more on the quantity of cell phone minutes used rather than connecting cell phone minutes used and cost of a cell phone.

Claim #3: Limited bias

Some initial items were flagged for bias for reasons involving topics like specialized knowledge of different sports (e.g., free throw in basketball) or phrasing in the item that suggested a student had cultural experience with a context (e.g., you went to the beach.) Feedback led to revisions such that the team handled it and resent items to the panels. As an example, the free throw item was revised to focus on throwing a basketball into a hoop. Item revisions helped to orient tasks like this one to general sports ideas that were also located in K-8 physical or health education standards. Items that used "you" were revised to include an individual's names. Care was taken to be culturally relevant; using names suggested by students from different geographic regions. Thus, it was more likely that students could perceive peers in their problems when they saw names relevant to their local culture. Revised items that were deemed limited in bias were shared with potential respondents during the think alouds. Students confirmed that they felt items contained no observed bias. Our team does not believe there is zero bias across the items but rather, it is limited in scope and not detected by the scholars, practitioners, and potential respondents involved in this study.

Discussion and Next Steps

Our goal was to explore the degree to which evidence supports validity claims of the CAT PSM 6-8. The a priori claims approach adheres to modern standards and best practices in assessment development (AERA et al., 2014; Folger et al., 2023). Based on the evidence and claims presented in this proceeding, test users may feel confident knowing the CAT PSM 6-8 does what it intends. Far too many tests provide insufficient information for test users (Bostic et al., 2022), which can lead to issues including but not limited to (a) spurious findings, (b) negative implications for test takers, and (c) less instructional time for K-12 students (AERA et al., 2014; Cronbach, 1988).

Bias is something that cannot be eliminated (AERA et al, 2014) and instead, is intended to be limited and balanced across an item set. In the case of these items, teachers, bias panel members, other scholars, and students did not perceive negative bias in the final versions of the items; however, that does not necessarily mean there is no bias. Instead, it provides strong qualitative evidence for a lack of observed bias. Further quantitative analyses will be performed using differential item functioning (DIF) to explore whether items grouped onto the CAT PSM have unbalanced bias, which will take place after test administration in May 2023.

Acknowledgement

Ideas in this manuscript stem from grant-funded research by the National Science Foundation (NSF #2100988, 2101026; #1720646, 1720661). Any opinions, findings, conclusions, or recommendations expressed by the authors do not necessarily reflect the views of the National Science Foundation.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Bostic, J., & Sondergeld, T. (2015). Measuring sixth-grade students' problem solving: Validating an instrument addressing the mathematics Common Core. *School Science and Mathematics Journal, 115, 281-291.*

Bostic, J., Sondergeld, T., Folger, T. & Kruse, L. (2017). PSM7 and PSM8: Validating two problem-solving measures. *Journal of Applied Measurement, 18*(2), 151-162.

Bostic, J., Sondergeld, T, Matney, G., Stone, G., & Hicks, T. (2021). Gathering response process data for a problem-solving measure through whole-class think alouds. *Applied Measurement in Education, 34*(1), 46-60.

Bostic, J., Krupa, E., Folger, T., Bentley, B., & Stokes, D. (2022, October). *Gathering validity evidence to support mathematics education scholarship*. In A. Lischka, E. Dyer, E., R. Jones, J. Lovett, J. Strayer, & S. Drown (Eds.), Proceedings of the forty-fourth annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education (pp. 100-104). Nashville, TN.

Carney, M., Bostic, J., Krupa, E., & Shih, J. (2022). Interpretation and use statements for instruments in mathematics education. *Journal for Research in Mathematics Education*, 53(4), 334-340.

Common Core State Standards Initiative. (2011). *Common Core State Standards for Mathematics*. http://www.corestandards.org/wp-content/uploads/Math_Standards.pdf

Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 3–17). Erlbaum.

Folger, T., Bostic, J., & Krupa, E. (2023). *Defining test-score interpretation, use, and claims: Delphi study for the validity argument. Educational Measurement: Issues & Practice*.

Hatch. J. A. (2002). *Doing qualitative research in educational settings.* State University of New York Press.

Kane, M. T. (2006). Validation. In R. L. Brennan, National Council on Measurement in Education, & American Council on Education (Eds.), *Educational measurement*. Praeger Publishers.

Kane, M. T. (2012). Validating score interpretations and uses. *Language Testing, 29*(1), 3-17.

Lesh, R., & Sriraman, B. (2009). Re-conceptualizing mathematics education as a design science. In *Theories of mathematics education: Seeking new frontiers* (pp. 123-146). Springer.

Lesh, R., & Zawojewski, J. (2007). Problem solving and modeling. In F.K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics* (pp. 763-803). Information Age.

Middleton, J., Gorard, S., Taylor, C., & Bannan-Ritland, B. (2008). The "compleat" design experiment. In A. Kelly, R., Lesh, & J. Baek (Eds.), *Handbook of design research methods in education: Innovations in science, technology, engineering, and mathematics teaching and learning* (pp. 21-46). Routledge.

Polya, G. (1945/2004). *How to solve it.* Princeton University Press.

Saldaña, J. (2013). *The coding manual for qualitative researchers* (2nd ed). London, UK: Sage.

Schoenfeld, A. H. (2011). *How we think: A theory of goal-oriented decision making and its educational applications.* Routledge.

Verschaffel, L., De Corte, E., Lasure, S., Van Vaerenbergh, G., Bogaerts, H., & Ratinckx, E. (1999). Learning to solve mathematical application problems: A design experiment with fifth graders. *Mathematical Thinking and Learning, 1,* 195-229.