

Learning from Mistakes: Word Problem-Solving Performance Patterns by Grade and in Response to Intervention

Maegan Colbert

April 26, 2025

Maegan Colbert¹, Isabel Valdivia¹, Jisun Kim², Tamika L. McElveen³, Sarah R. Powell⁴, Amanda S. Mayes⁵, Ma Bernadette Andres-Salgarino⁶, Sara A. Schmitt⁷, David J. Purpura², & Caroline Byrd Hornburg¹



Background: U.S. Students Struggle in Math

- Decline in U.S. student math proficiency from elementary to middle school (National Center for Education Statistics, 2025)
 - 39% of students “proficient” or better in 4th grade
 - 28% “proficient” or better in 8th grade
- Word problem solving (WPS) is especially challenging due to its complexity (Cummins et al., 1998; Fuchs et al., 2020):
 - Decode and understand text
 - Identify the relevant mathematical schema
 - Make a plan based on the identified schema
 - Execute the plan

Background: Conceptual and Procedural Knowledge

(Rittle-Johnson & Alibali, 1999)

- **Conceptual knowledge:** identifying “equal groups” schema
- **Procedural knowledge:** understanding the steps needed to execute a given process (e.g., how to multiply, how to regroup)

C. Mr. Lopez sold 5 boxes of peaches at the Farmer’s Market. Each box holds 12 peaches. How many peaches did Mr. Lopez sell?

$$\begin{array}{r} 12 \\ \times 5 \\ \hline \end{array}$$

Mr. Lopez sold 60 peaches

Mr. Lopez sold 60 peaches

C. Mr. Lopez sold 5 boxes of peaches at the Farmer’s Market. Each box holds 12 peaches. How many peaches did Mr. Lopez sell?

$$\begin{array}{r} 12 \\ \times 5 \\ \hline \end{array}$$
$$\begin{array}{r} 12 \\ 12 \\ 12 \\ 12 \\ \hline 60 \end{array}$$

She did 60 boxes

Background: Functional Use of Errors in Math Learning

- Gaps in conceptual and procedural knowledge are associated with different errors (Kingsdorf & Krawec, 2014)
- Understanding error patterns in response to intervention and across grades supports intervention development (Bailey et al., 2018)
- Little work has been done exploring how student-level characteristics relate to responsiveness to metacognitive interventions
- Analyzing student work is an important part of understanding the nature of their error patterns

The Present Study: Research Questions

1. How do patterns of problem-solving accuracy and error types change across grades?
2. How predictive are conceptually poor pre-test error patterns for students' response to intervention (i.e., increases in accuracy and/or decreases in proportion of conceptually poor errors)?

Dataset

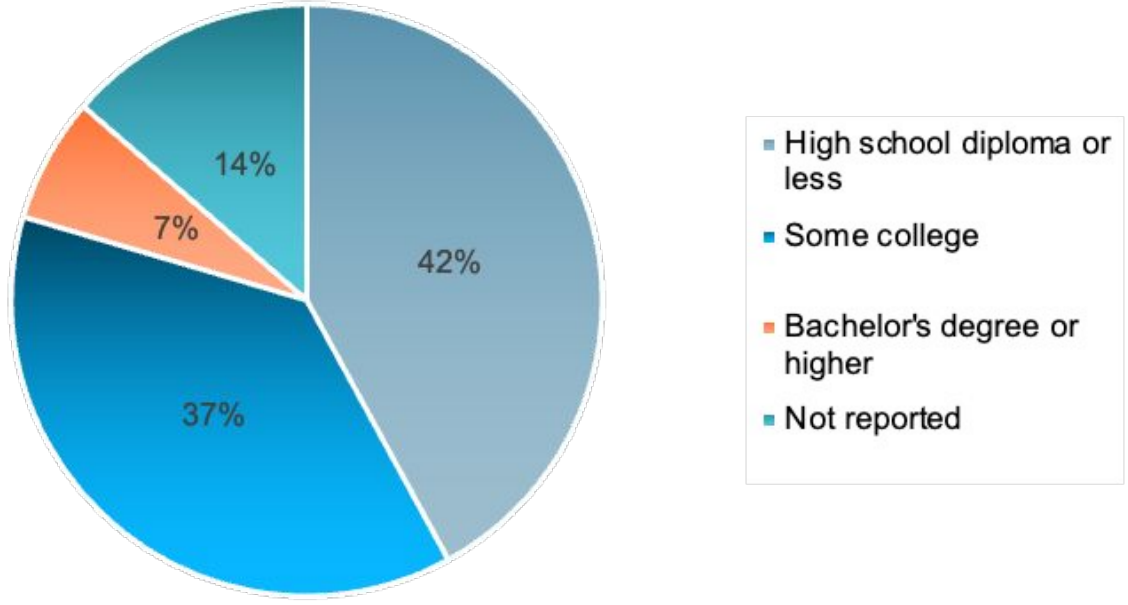
- Phase 3 of curriculum co-design project: “Our Mathematical World”
 - One California public school
 - Targeted 3rd-5th grades
 - One 7th grade teacher/class volunteer
 - Feasibility study - no control group
- Intervention targeted:
 - Metacognitive problem-solving approach (PULSE: Pause, Understand and Remember, Lay it Out, Solve, and Evaluate)
 - Executive function
 - Math identity
- Pre- and post-test data



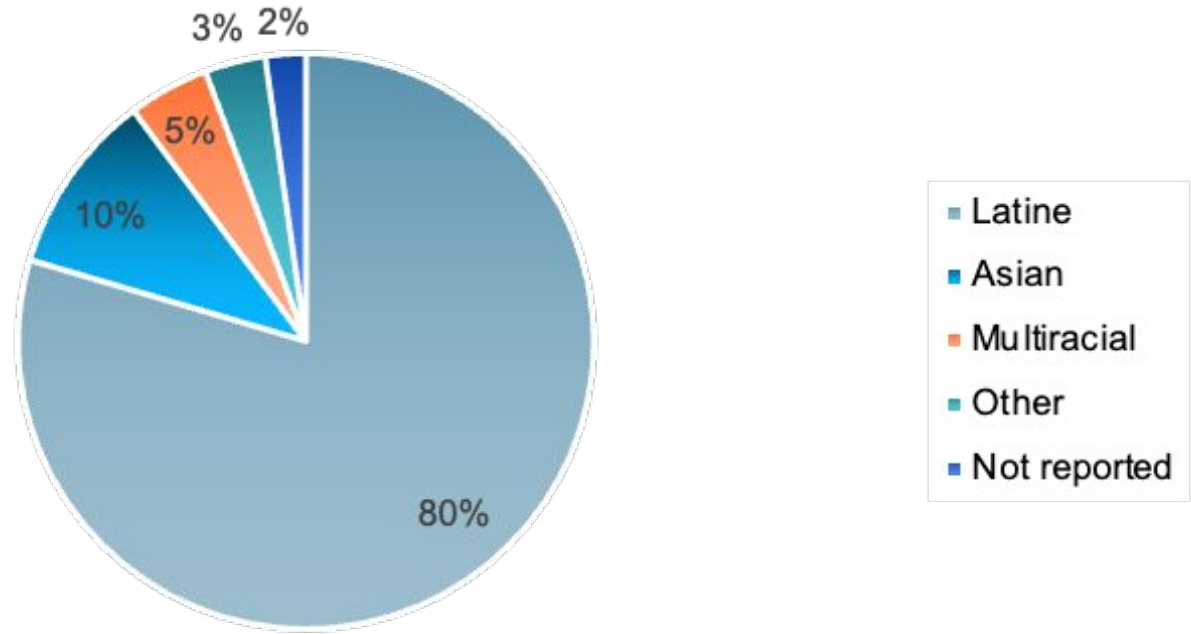
Population ($N = 88$; 49 girls)

- 3rd grade: 15 students (one teacher/class)
- 4th grade: 30 students (two teachers/classes)
- 5th grade: 27 students (two teachers/classes)
- 7th grade: 16 students (one teacher/class)

Highest Parent Education



Participant Race/Ethnicity



Procedure

Pretest	Intervention	Posttest	Error coding
<ul style="list-style-type: none">● 6 single step word problems (adapted from Powell et al., 2020)● Additional measures (not included in current analyses)	<ul style="list-style-type: none">● 9 week overlay● 3 sessions per week, 30 minutes per session● Targeted:<ul style="list-style-type: none">- PULSE approach to problem-solving- executive function- math identity	<ul style="list-style-type: none">● 6 single step word problems (same as pretest problems)● Additional measures (not included in current analyses)	<ul style="list-style-type: none">● Student work coded for error type (adapted from Kingsdorf & Krawec, 2014)● Errors were categorized as conceptually poor or conceptually sound (Rittle-Johnson & Alibali, 1999)

Notes on Word Problem Procedures

- 3rd & 4th grade level problems
- Same problems given to all grades
- Read aloud by teacher and printed on student worksheet
- Problems completed one at a time; 2 minutes per problem

Error Classifications: Conceptually Poor

- Omission
- Selecting an incorrect operation
- Using a number that was not provided in the problem
- Using numbers from the problem incorrectly
- Providing a final, incorrect answer without sufficient work to be able to discern their process

C. Mr. Lopez sold 5 boxes of peaches at the Farmer's Market. Each box holds 12 peaches. How many peaches did Mr. Lopez sell?

$$\begin{array}{r|l} & 5 \\ + & 12 \\ \hline & 17 \end{array}$$

$$\underline{5} + \underline{12} = \underline{17}$$

$$5 + 12 = 17$$

C. Mr. Lopez sold 5 boxes of peaches at the Farmer's Market. Each box holds 12 peaches. How many peaches did Mr. Lopez sell?

5

C. Mr. Lopez sold 5 boxes of peaches at the Farmer's Market. Each box holds 12 peaches. How many peaches did Mr. Lopez sell?

3

Error Classifications: Conceptually Sound

- Procedural errors (most common)

- Computation error
- Setting up an operation incorrectly (e.g., inverting subtrahend & minuend)

- Conceptually correct errors (rare)

- Running out of time
- A setup error unrelated to the concept
- Transcription error

C. Mr. Lopez sold 5 boxes of peaches at the Farmer's Market. Each box holds 12 peaches. How many peaches did Mr. Lopez sell?

$$12 \times 5 = 70$$

$$12 \times 5 = 70$$

G. Kavon rode the rollercoaster 44 times. His brother rode the rollercoaster 25 times. How many fewer times did Kavon's brother ride the rollercoaster?

$$\begin{array}{r} 44 \\ - 25 \\ \hline \end{array}$$

Results

Note on Student Characteristics

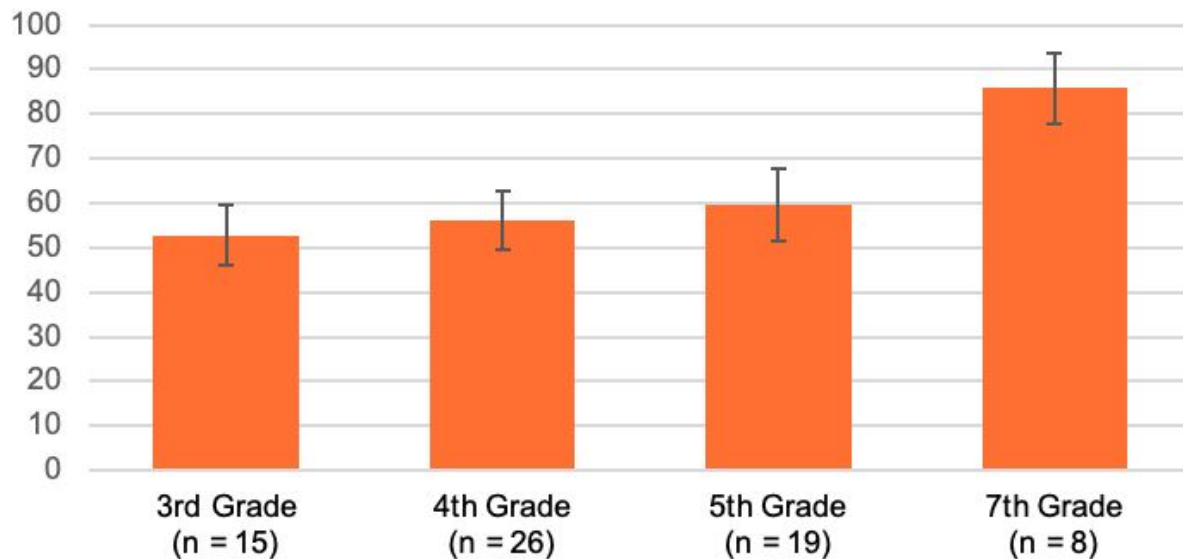
- Not significantly related to accuracy or error type at pre or post:
 - Gender
 - Error climate rating at pretest
 - Home language
 - Parent education

Research Question 1: How do patterns of problem-solving accuracy and error types change across grades?

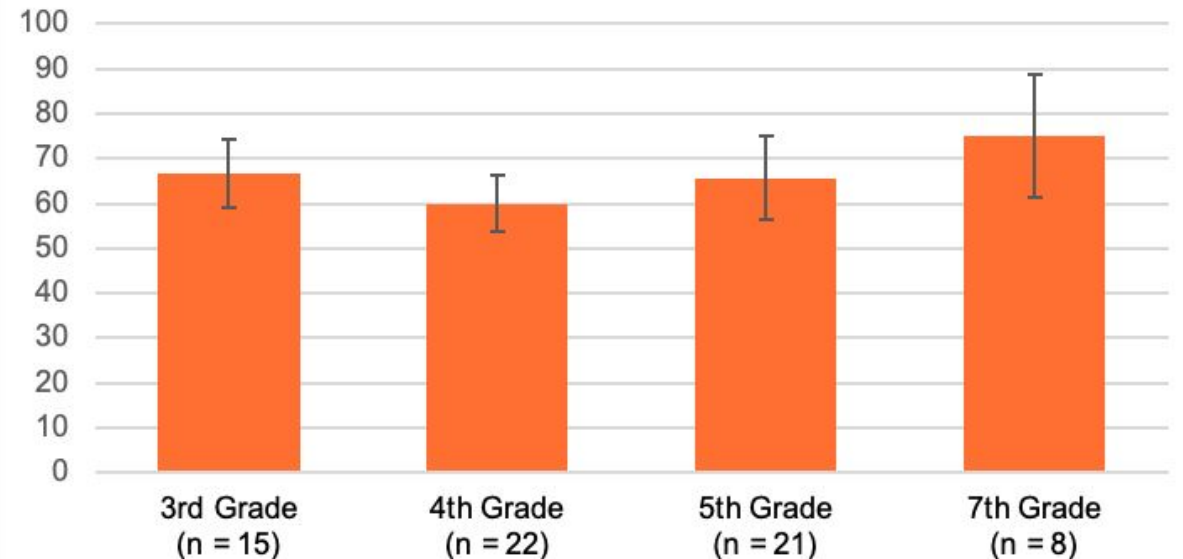


Research Question 1: How do patterns of problem-solving accuracy and error types change across grades?

Average Percent of Pretest Errors that Were Conceptually Poor* (by Grade)



Average Percent of Posttest Errors that Were Conceptually Poor* (by Grade)



*Percent of mean **total** errors that were conceptually poor

Research Question 2: How predictive are conceptually poor pre-test error patterns for students' response to intervention (i.e., increases in accuracy and/or decreases in proportion of conceptually poor errors)?

Predicting Post-test Problem Solving Accuracy

- **Statistically significant predictors** (controlling for grade):
 - Pre-test accuracy
 - Pre-test conceptually poor errors (above and beyond pre-test accuracy)
- **Final model:**
 - 57.4% of post-test accuracy variance
 - Pre-test conceptually poor errors: 11.5% of post-test accuracy variance beyond pre-test accuracy

<i>Predicting Post-test Problem-Solving Accuracy (N = 68)</i>						
		β	t	p	r_p	$R^2\Delta$
Model 1						.165
	Grade	.406	3.610	<.001	.406	
Model 2						.354
	Grade	.225	2.502	.015	.296	
	Pre-test accuracy	.622	6.913	<.001	.651	
Model 3						.055
	Grade	.298	3.350	.001	.386	
	Pre-test accuracy	.606	7.093	<.001	.663	
	Pre-test conceptually poor errors	-.245	-2.880	.005	-.339	

Predicting Post-test Conceptually Poor Errors

- **Statistically significant predictors** (controlling for grade):
 - Pre-test accuracy
- **Final model:**
 - Pre-test accuracy predicted 7.6% of post-test conceptually poor variance when also accounting for pre-test conceptually poor errors

<i>Predicting Post-test Conceptually Poor Errors (N = 59)</i>						
		β	t	p	r_p	$R^2\Delta$
Model 1						.029
	Grade	.170	1.303	.198	.170	
Model 2						.072
	Grade	.244	1.858	.068	.241	
	Pre-test accuracy	-.278	-2.112	.039	-.272	
Model 3						.002
	Grade	.257	1.870	.067	.244	
	Pre-test accuracy	-.282	-2.119	.039	-.275	
	Pre-test conceptually poor errors	-.046	-0.347	.730	-.047	

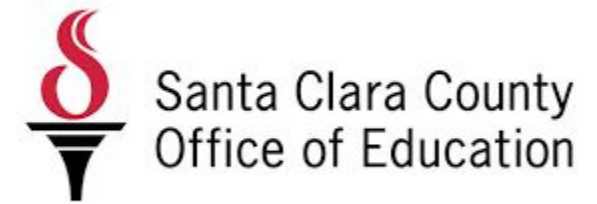
Takeaways...

- Error analysis can help us determine students' level of understanding and knowledge gaps to inform interventions
 - Conceptual difficulties
 - Procedural difficulties
 - Metacognitive and EF difficulties
- Consider non-significant relation between pre- and post-test conceptually poor errors:
 - Possible that some student pretest errors that we interpreted as conceptually poor were more related to difficulty with metacognition and EF
 - Those students may have benefited more from the intervention
- We need more information about baseline student characteristics and thought processes
 - Baseline metacognition
 - Student explanations of their own thought processes

Future Directions

- Larger sample sizes are needed to draw broader conclusions
- Experimental work (no control group here)
- Analysis of multi-step problems
- Problem-level and schema-level analyses
- Future work should consider baseline metacognition as a potential student characteristic relating to a student's response to metacognitive intervention
- Planned dissertation project: examining student characteristics related to showing work and response accuracy in WPS

Thank you to the students, parents,
teachers, and school staff who
contributed to this project!



This research was supported by the EF+Math Program of the Advanced Education Research and Development Fund (AERDF) through funds provided to Purdue University.

The opinions expressed are those of the authors and do not represent views of the EF+Math Program or AERDF.


EF+MATH

supported by  aerd



Thank you!

Questions?

- Email Maegan Colbert: maegan@vt.edu
- Virginia Tech Learning and Development Lab
 - www.vtlearnlab.edu
 -  VTLearnLab

