

# Identification, Exploration, and Remediation: Can Teachers Predict Common Wrong Answers?<sup>1</sup>

Ashish Gurung

Final versions of this work is published in the Learning Analytics and Knowledge Conference(LAK), 2023

1

You failed to explain the title. You jump too what is a CWA with connecting to the "Can teacher predict"

## Common Wrong Answers(CWAs)

Defining Common Wrong Answers:

- Incorrect Attempts that commonly occur are considered CWAs
- Bugs (VanLehn[2, 4], Brown[1], and Sison[5])
- Diagnostic model (Brown et al. [1], Selent [8])
- Repair Theory (Brown and VanLehn[3])
- Feedbacks can be beneficial (Narciss[6, 7])

Important to note:

Proactive approach when predicting the CWAs

<sup>1</sup> McKendree, J. (1990). Effective feedback content for tutoring complex skills. *Human Computer Interaction*, 5, 381-413

Maybe add citations here in really small font

Advisrise kirk or aaron

The selent work is low quality.

I dont know Narciss.

YOu are still talking about Narciss but have not mentioned the title of the talk about can teacher predict this.

You are taking a proactive approach

(Simon Woodhouse's EEiD Edie or something like that has a think focused on CWA from 4 choice MC )

The screenshot shows a digital worksheet interface. On the left, there is a 'Problem' section with a math equation:  $\frac{\%v\{a\}}{\%v\{b\}} + \frac{\%v\{c\}}{\%v\{d\}} = \frac{\%v\{e\}}{\%v\{f\}}$ . Below it is an 'Answer' section with a green checkmark next to '%v\{f\}' and a red X next to '%v\{(d-c) \* b \* -1.0\}'. To the right, under 'Variables', is a table with the following entries:

a = {a,b,c,x,y}
Variable has string values.
e = {1:-1}
g = {1:-1}
c = rand(10)+2
bnum = rand(10)+2*e
d = rand(10)*2*g
b = if (bnum == 0) then (rand(9) + 1); else bnum
f = (d-c)*b

## CWAs

Example of CWAs

- How are CWAs established?  
analyze -> create diagnostic model ->  
generate templates -> generate  
feedback

Good slide!!!

Great

Problem: Solve for %v{a}

Variables:

```
a = {a;b;c;x;y}  
Variable has string values.  
e = {1:-1}  
g = {1:-1}  
c = rand(10)+2  
bnum = rand(10)+2*e  
d = rand(10)+2*g  
b = if (bnum == 0) then (rand(9) + 1); else bnum  
f = (d-c)*b
```

Answer:

✓ %v{f}  
✗ %v{((d-c) \* b \* -1.0)}  
Check your sign  
Positive x Positive = Positive  
Negative x Negative = Positive  
Positive x Negative = Negative  
Negative x Positive = Negative  
Step 1: Correct  
%v{a}%v{b} - %v{c} = %v{d}  
- %v{c} - %v{c}  
Step 2: Correct  
%v{b} \* %v{a} \* %v{b} = %v{d-c} \* %v{b}  
Sign: Incorrect  
%v{d-c} \* %v{b} IS NOT %v{((d-c) \* b \* -1)}

Problem ID: 711123 Comment on this problem  
Solve for a  
9a + 10 = 28

Check your sign  
Positive / Positive = Positive  
Negative / Negative = Positive  
Positive / Negative = Negative  
Negative / Positive = Negative  
Step 1: Correct  
9a + 10 = 28  
-10 -10  
9a = 18  
9 9  
Sign: Incorrect  
a = 18/9 IS NOT -2

Type your answer below as a number (example: 5, 3.1, 4 1/2, or 3/2): -2 75%  
Sorry, try again: "-2" is not correct  
Submit Answer Show hint 1 of 2

4

Good above vistionalie

Are we going to show the RCT (missing the thread from the title but- we are following details right now

You are on this

## Two types of CWAFs

Verbose and Detailed  
2 Step Equation

Problem ID: 493508 Comment on this problem  
What is the solution to the expression below?  
7 + 4 x 3

You should multiply before you add. PEMDAS

Type your answer below (mathematical expression): (3^4+7)  
Sorry, try again: "(3^4+7)" is not correct  
Submit Answer Show hint 1 of 2

Problem ID: 711123 Comment on this problem  
Solve for a  
9a + 10 = 28

Check your sign  
Positive / Positive = Positive  
Negative / Negative = Positive  
Positive / Negative = Negative  
Negative / Positive = Negative  
Step 1: Correct  
9a + 10 = 28  
-10 -10  
9a = 18  
9 9  
Sign: Incorrect  
a = 18/9 IS NOT -2

Type your answer below as a number (example: 5, 3.1, 4 1/2, or 3/2): -2 75%  
Sorry, try again: "-2" is not correct  
Submit Answer Show hint 1 of 2

Short and concise

If you want to show this show 5-10 of them?

All of this is skill builder or soe EIR stuff?

McKenderr I think did something on common wrong answer feedback

## RQ1 Proactively Identified CWAs

- Teachers and activity designers analyzed the problems
- Identified the mechanism that can result in the CWAs
- Generated the CWAFs to help address the underlying misconception, gap in knowledge or other potential learner behavior (think slip and guess)

	problems	Teacher Identified CWAs
Order of Operations	54	270
2-Step Equations	52	359

Table 1: Total CWAFs generated across two problem sets.

7

## Experimental Design

- Students were asked to work on mastery based assignments
- N-CCR design (N=3)
- Daily limit = 10
- Potential outcomes
  - Mastery
  - Wheel-Spinning [9]
  - Dropout
- Data was collected across 9 academic years 13-14 to summer 22

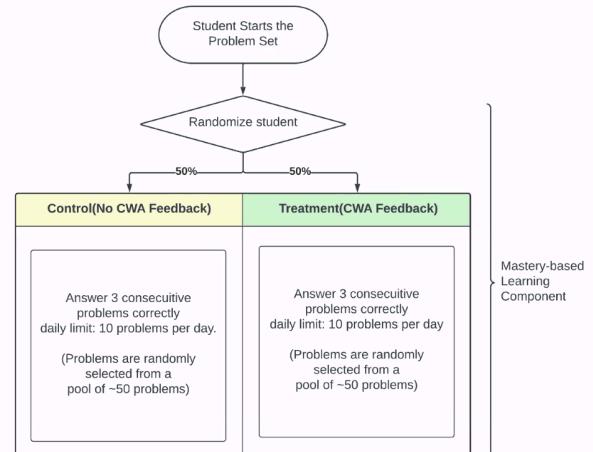


Fig: A/B test design

8

Like the slide but we did n

Get rid of the N-CCR acronym. Spell it out

How many problem sets do we have?

Surprised we don't have nice cases

Student Level randomize for the whole problem sets

## Experimental Descriptives

	Order of Operations	2-Step Equations	Combined
Teachers	202	458	587
Classes	386	954	1282
Assignments	497	954	1282
Students	6679	16976	23655

Table 4: Descriptives of the A/B test exploring the effectiveness of CWAs in two activities.

Good!

The n size is insanely high

## RQ1 Observed CWAs ( $N \geq 5$ )

- More unidentified CWAs in 2-Step Equation

	Teacher Identified CWAs	Observed CWAs
<b>Order of Operations</b>		
CWAs identified by teacher	270	88
CWAs not identified by teacher	-	15
<b>2-Step Equations</b>		
CWAs identified by teacher	359	228
CWAs not identified by teacher	-	192

Table 2: Observed CWAs when any incorrect attempt with  $n \geq 5$  is considered a CWA.

## RQ1 Observed CWAs ( $N \geq 5$ )

- More unidentified CWAs in 2-Step Equation

	Teacher Identified CWAs	Observed CWAs
<b>Order of Operations</b>		
CWAs identified by teacher	270	88
CWAs not identified by teacher	-	15
<b>2-Step Equations</b>		
CWAs identified by teacher	359	228
CWAs not identified by teacher	-	192

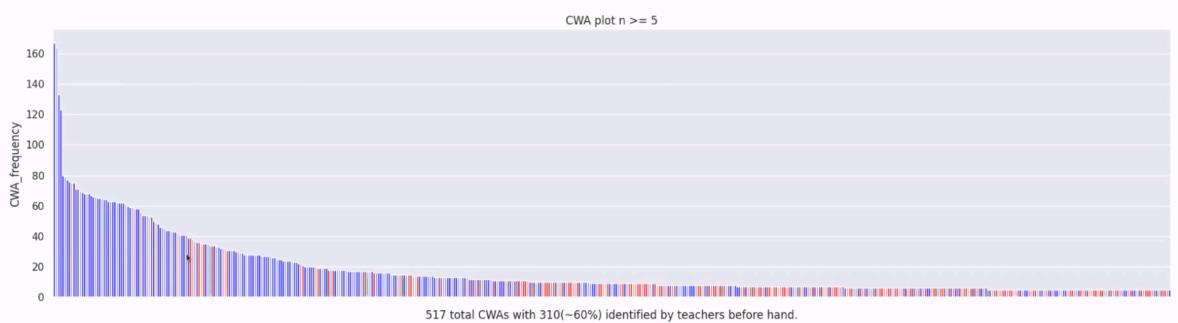
Table 2: Observed CWAs when any incorrect attempt with  $n \geq 5$  is considered a CWA.

10

I are talking about them as teaches but I think its really drug select and Cristina heffernan failed to make buggy rules fo such instances

Does the buggy rules makkter across types

## RQ1 Observed CWAs ( $N \geq 5$ )



You could site Wang that shows that the non common wrong answer were associated with less knowledge

## RQ1 Observed CWAs ( $N \geq 10$ )

- Most of the CWAs were identified by the teachers if we consider 10 to be the cutoff point.

	Teacher Identified CWAs	Observed CWAs
<b>Order of Operations</b>		
CWAs identified by teacher	270	57
CWAs not identified by teacher	-	0
<b>2-Step Equations</b>		
CWAs identified by teacher	359	143
CWAs not identified by teacher	-	54

Table 3: Observed CWAs when any incorrect attempt with  $n \geq 10$  is considered a CWA.

Help them compare to the prior slide where you go from 192 to 54  
(that is hard for them to memorise )

## Experimental Design

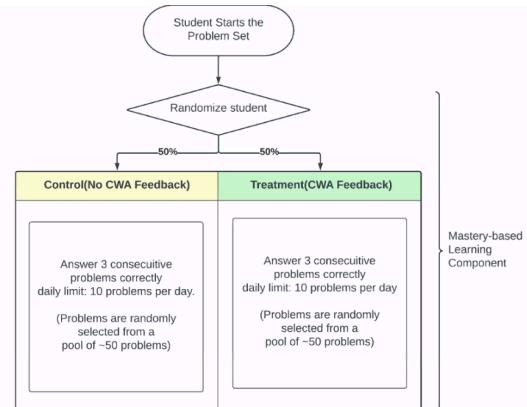


Fig: A/B test design

Why are we going over this again

## RQ2 Effects of CWAFs on Learning Outcomes

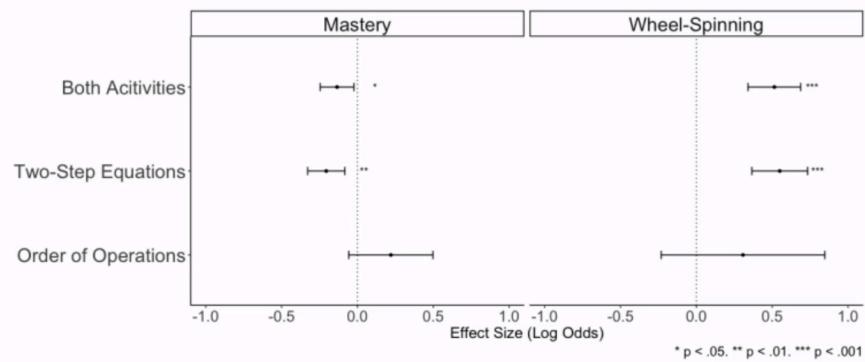


Fig: Comparison of student performance in the CWAF and the control condition

16

Why do we smoosh them together in "Both"

Dont b (did you tell them what Wheel spinning is) (first tell them the two means and and then tell them a effect size)

(explaine they are 95% confidence interval)

First show this

## RQ2 Effects of CWAFs on Learning Outcomes



Then say we want to break it out and then and only then show the split out

## RQ2 Effects of CWAFs on Learning Outcomes

Predictors	Both Activities		2-Step Equations		Order of Operations	
	Log-Odds	SE	Log-Odds	SE	Log-Odds	SE
Intercept	2.95***	0.08	2.68***	0.09	3.66***	0.17
CWAFs (Treatment)	-1.30**	0.06	-0.21**	0.06	0.22	0.14
<b>Random Effects</b>						
$\sigma^2$	3.29		3.29		3.29	
$\tau_{00}$	1.93 <sub>t</sub>		1.83 <sub>t</sub>		1.68 <sub>t</sub>	
ICC	0.37		0.38		0.36	
N	587 <sub>t</sub>		458 <sub>c</sub>		202 <sub>c</sub>	
	23604 <sub>i</sub>		16926 <sub>i</sub>		6678 <sub>i</sub>	

Table 5:  $\text{logit}(\text{mastery} \sim Z + (1|\text{teacher}))$  | Z is control or treatment assignment

## RQ3 Exploring Potential HTE of CWAFs

What is HTE? Get rid of these acronym (maybe even CWA)

## RQ3 Exploring Potential HTE of CWAFs

Predictors	Both Activities		2-Step Equations		Order of Operations	
	Log-Odds	SE	Log-Odds	SE	Log-Odds	SE
Intercept	3.34***	0.09	3.06***	0.09	4.29***	0.17
CWAFs (Treatment)	-0.17*	0.07	-0.18*	0.08	-0.06	0.18
Prior Problem Correct (Z-Score)	0.77***	0.04	0.80***	0.05	0.79***	0.10
Treatment X Prior Problem Correct	-0.11*	0.05	-0.80	0.06	-0.36*	0.14
<b>Random Effects</b>						
$\sigma^2$	3.29		3.29		3.29	
$\tau_{00}$	1.62 <sub>t</sub>		1.45 <sub>t</sub>		2.03 <sub>t</sub>	
ICC	0.33		0.31		0.38	
N	564 <sub>t</sub>		443 <sub>c</sub>		191 <sub>c</sub>	
	21793 <sub>i</sub>		15835 <sub>i</sub>		5958 <sub>i</sub>	

Table 7:  $\text{logit}(\text{mastery} \sim Z * \text{prior performance} + (1|\text{teacher}))$   
 $Z$  is control or treatment assignment

Why not forest plots. I like that better.

## RQ3 Exploring Potential HTE of CWAFs

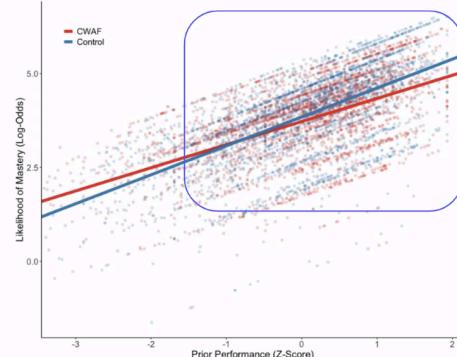


Figure: Exploring HTE of receiving CWAFs on students with different prior performance when working on the Order of Operations activity

Is this combined or just one of these?

Help them see the Y access is if they finish they homework of three write in a row (you call it mastery later but by change to mastery)

## Takeaways

- Proactively predicting CWAs and generating CWAFs is not the optimal approach
- Possible alternative: user historical data to establish the more common CWAs
  - Data Driven
  - Reduces total CWAs
  - Ranks CWAs for prioritization
- CWAFs can hurt students
  - Long and Verbose CWAFs hurt everyone
  - Short and Concise hurt higher performing students more than lower performing students
- We really need to be careful when designing CWAFs
- Additional work is required to establish the various factors that influenced the negative effect of CWAFs on student learning outcomes.

2

"When the topic is the same, student benefit" What the heck does "When the topic is the same" mean. I dont see how to square this negative results and your later EIR results can squared.

## Limitations & Future Work

- Automated nature of CWAFs
  - Learner Autonomy
  - Prior performance or Knowledge Tracing based heuristic
- Estimation of Student Effort on the Feedback using response time decomposition[]
- Causal Forest for Heterogeneous Treatment Effects(HTE)[]
  - Further expand on the treatment effects of CWAFs
  - Gain insight into various factors that influenced the effectiveness of CWAFs
- Analysis of CWAs and generation of CWAFs using Large Language Models.

Students can be more sensitive to Common Wrong Answer Feedback unlike other feedback(Hints and Explanations) lead to unforeseen learning outcomes.

What would it mean to account for autonomy mean?  
You talked about

- Automated nature of CWA  
Fs
  - Learner Autonomy
  - Prior performance or Knowledge Tracing based heuristic

Why suggest this thing when you can not document it yet

You are talking about

- Estimation of Student Effort on the Feedback using response time decomposition[]  
Causal Forest for Heterogeneous Treatment Effects(HTE)[]

Put back in the citations

Are a tiny.cc name url into the Slide Number thing so folks can get to the file via a OSF page  
(you have said nothing so far about OSF)

Open data

Open materials

I know it was not pre-registered

Why should we believe that you did not try this analysis 5 times?

Too long on

## Limitations & Future Work

- Automated nature of CWA  
Fs
  - Learner Autonomy
  - Prior performance or Knowledge Tracing based heuristic
- Estimation of Student Effort on the Feedback using response time decomposition[]
- Causal Forest for Heterogeneous Treatment Effects(HTE)[]
  - Further expand on the treatment effects of CWA  
Fs
  - Gain insight into various factors that influenced the effectiveness of CWA  
Fs
- Analysis of CWAs and generation of CWA  
Fs using Large Language Models.

Students can be more sensitive to Common Wrong Answer Feedback unlike other feedback(Hints and Explanations) lead to unforeseen learning outcomes.

Add tiny url to this page



# Thank You



Project GitHub



Webpage  
agurung@wpi.edu