Game2Learn Research Lab NC STATE UNIVERSITY

Investigating the Impact and Student Perceptions of Guided Parsons Problems for Learning Logic with Subgoals

go.ncsu.edu/csedm2025

Sutapa Dey Tithi

stithi@ncsu.edu

Dr. Xiaoyi Tian

xtian9@ncsu.edu

Dr. Min Chi

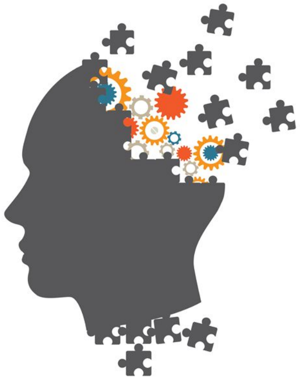
mchi@ncsu.edu

Dr. Tiffany Barnes

tmbarnes@ncsu.edu

Game2Learn Research Lab NC STATE UNIVERSITY

Instructional Design for Optimizing Learning Outcomes



- ❑ Three primary types of cognitive load (Sweller et al. 1998 [4])
  - ❑ Intrinsic load
    - ❑ Inherent difficulty of the material
    - ❑ May vary based on a student’s prior knowledge
  - ❑ Extraneous load
    - ❑ How information is presented and the ease with which a student comprehends it
  - ❑ Germane load
    - ❑ Integrating new information and how we process it into long-term memory

Instructional design for optimizing learning outcomes (Cont.)

- ❏ **Worked examples can save students time** without reducing their learning ([6])
- ❏ Nievelstein et al. found that worked examples may not be beneficial for students with high prior knowledge when problems are structured [7].
- ❏ Worked examples often lead to **passive engagement** by not clearly explaining the reasoning behind each step [3].

- ❏ In contrast, **unstructured problem solving** can place **high cognitive demands** on students as they try to construct multi-step proofs [4].

Parsons Problems

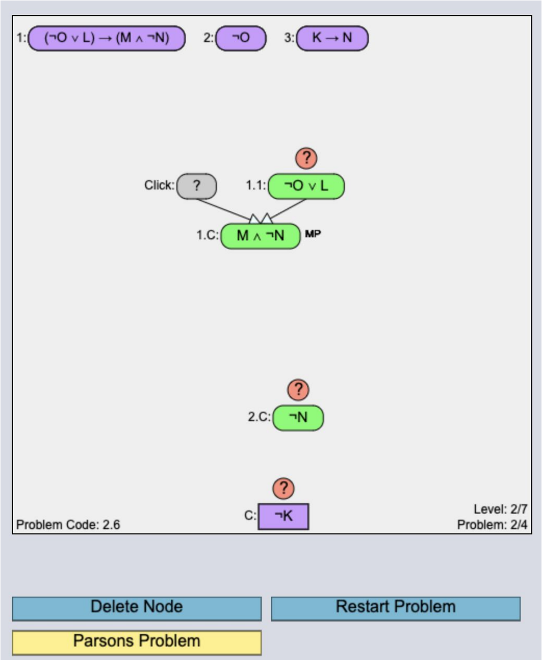
- Parsons problems have emerged as a promising scaffold for teaching structured problem solving.
  - Enable learners to **reconstruct jumbled proof steps** into valid solutions while reducing cognitive load [1].
- In programming education, Parsons problems have been extensively explored and found to improve students' code writing abilities [10, 11 , 9 , 12].

Parsons Problems in Logic

- Shabrina et al. demonstrated that data-driven, subgoal-oriented Parsons problems can **enhance students' subgoaling skills** in solving propositional logic proofs [2].
- They also found that students **struggle with Parsons problems**
  - when they **first encounter** this type of structured problem, or
  - when the **connections** among different parts of the problem are **complex**.

Parsons Problem

Frequent approaches were decomposed using data-driven subgoals and presented as chunks.



Experts Formalized the Structure of Chunk Explanations:

1. What the chunk derives.
2. How it is derived.
3. Why it is derived.

Improving Learning with Guided Parsons Problems

- Add **step-specific hints** to address the **“rationale gap”** of worked examples [16]
- Add **self-explanations** to understand the impact on students’ perception of problem subgoals
- Designed to maintain
  - **low intrinsic**
  - while facilitating **active problem solving**.

Research Questions (RQs)

RQ1: What is the **impact** of Guided Parsons problems (GPP) on student performance and learning outcome?

RQ2: To what extent does student **proficiency level moderate the relationship** between GPPs and student learning outcomes?

RQ3: What **common themes** emerge from students' self-explanations on their learning experiences with GPPs?

Context: DT, The Intelligent Logic Tutor

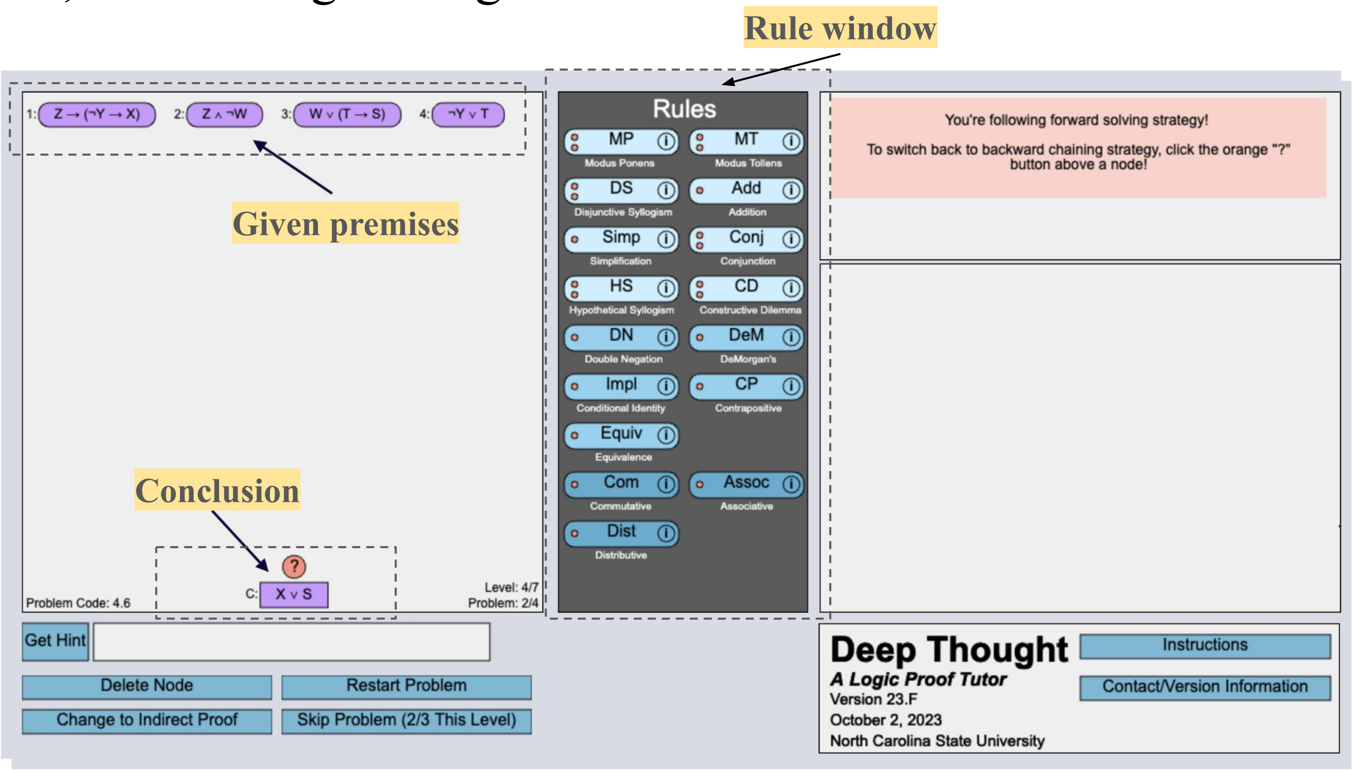


Figure 1: Full Interface of Deep Thought with Student Workspace (left), Rules (middle), Instructions (top-right)

Problem Organization

7 levels

- Level 1
  - 3 Intro problems | Pretest
- Level 2-6
  - 3 training problems
  - One level-end test problem
- Level 7
  - 6 posttest problems

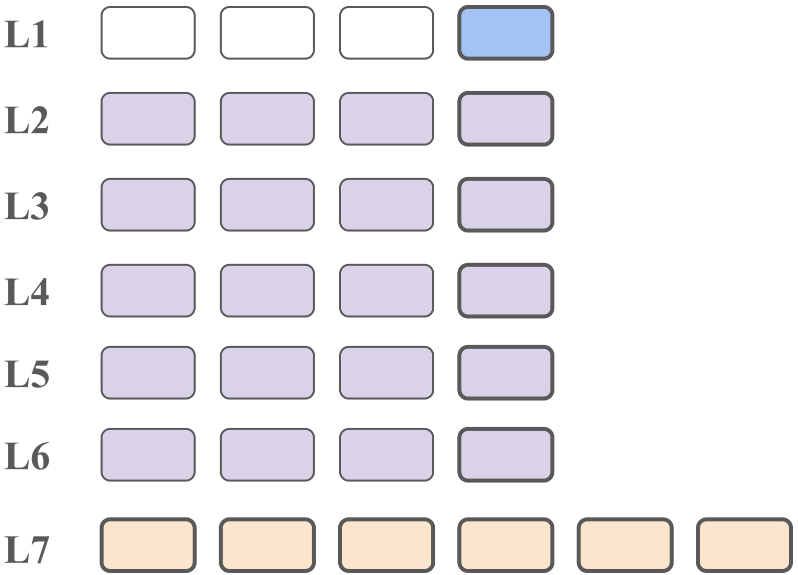


Figure 2: Problems in Different Levels

10

Problem Type: Problem-solving (PS)

➔ Clicking one or two existing statements or nodes, a rule button, and **entering the new derived statement**

Once a step is verified by the tutor, the new node appears.

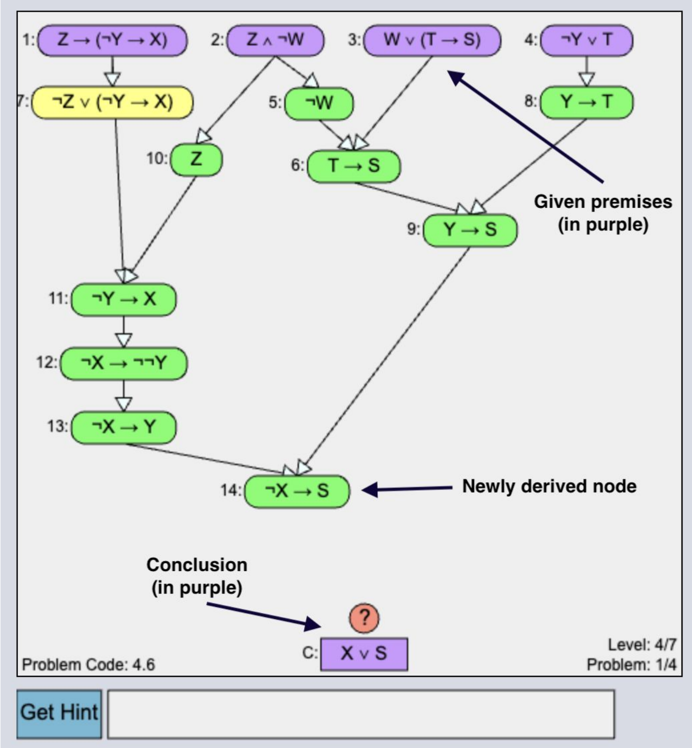


Figure 3: PS Interface

Problem Type: Worked Example (WE)

➔ The tutor shows one step at a time, consisting of adding a new node to the screen with its justification

Students press Next/Previous to progress between steps.

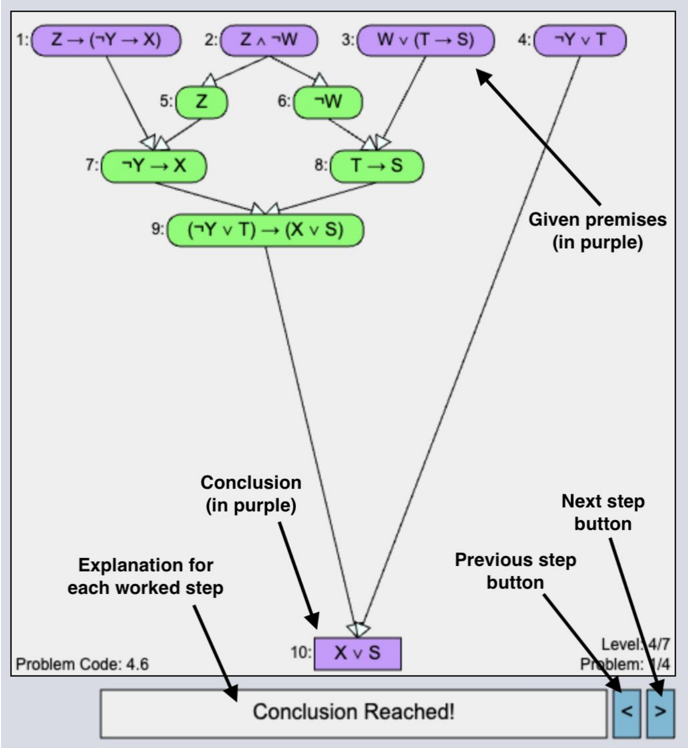


Figure 4: WE Interface

Problem Type: Guided Parsons Problem (GPP)

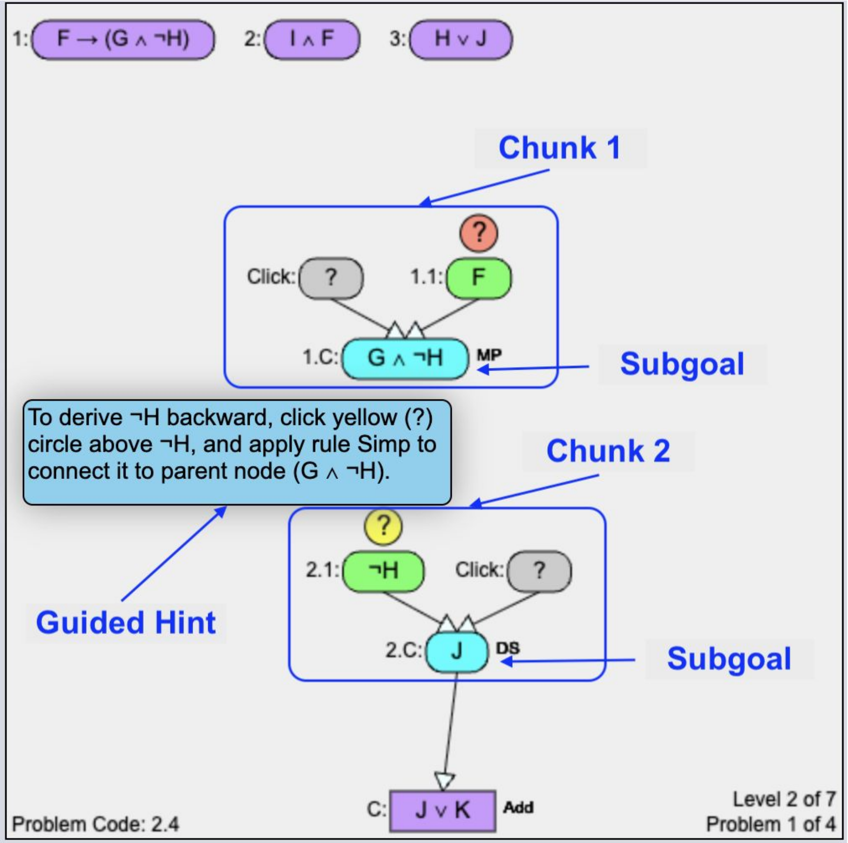
Game2Learn Research Lab

NC STATE UNIVERSITY

➔ Each GPP provides students with **all the statement nodes needed** to complete a proof.

Students must **add a few justifications to connect all the nodes** to one another with missing edges for rules.

GPPs **guide students** to justify each unjustified node by **specifying the rule** used to derive it.



13

Experimental Conditions

76 students in an undergraduate discrete mathematics course in Spring 2024

- **<u>Control:</u>** Random **PS or WE**:
  - Training problems were randomly Problem Solving or Worked Examples
- **<u>GPP:</u>** Random **PS or GPP:**
  - Training problems were Problem Solving or Guided Parsons Problems

Research Questions (RQs)

RQ1: What is the **impact** of Guided Parsons problems (GPP) on student performance and learning outcome?

RQ2: To what extent does student **proficiency level moderate the relationship** between GPPs and student learning outcomes?

RQ3: What **common themes** emerge from students' self-explanations on their learning experiences with GPPs?

Performance Metrics

- Problem scores are **a weighted sum of three metrics** in the range [0, 1]
  a. Solution length
  b. Problem-solving time
  c. Accuracy of rule application
- Normalized learning gain (NLG) and learning efficiency (LE)

Performance Metrics: NLG and LE

(Shabrina et al. 2023 [24])

$$\text{NLG} = \frac{(\text{posttest score} - \text{pretest score})}{\sqrt{(1 - \text{pretest score})}}$$

$$\text{LE} = \frac{\text{NLG}}{\text{Tutor completion time}}$$

where, NLG is scaled between 0 and 1, and the tutor completion time includes the total time students spent on the tutor (pretest, training, and posttest problems)

RQ1: No differences in Normalized Learning Gains

Table 1: Problem Score and Normalized Learning Gain (NLG) across the Two Training Groups.

Group (N)	Pre	Post	NLG	% Students with (+) NLG
Control (30)	62.8 (18.7)	70.4 (14.4)	0.26 (0.45)	73%
GPP (46)	63.8 (18.1)	72.6 (8.2)	0.27 (0.44)	78%

No significant differences in NLG, but a higher percentage of students with positive NLG scores

Note: NLG is often negative in this tutor because of posttest difficulty

RQ1: Rule accuracy improved by Guided Parsons Problems

Table 2: Rule accuracy (Mean (SD)) across two conditions in training level-end test and posttest problems.
[Note: **Blue*** indicates a significant difference using Mann-Whitney U]

Test	Control (30)	GPP (46)	Test Results
Level-End Tests	59.4 (24.6)	**68.7 (20.2)***	$p = .002$
Posttests	72.7 (22.1)	**79.8 (17.2)***	$p = .003$

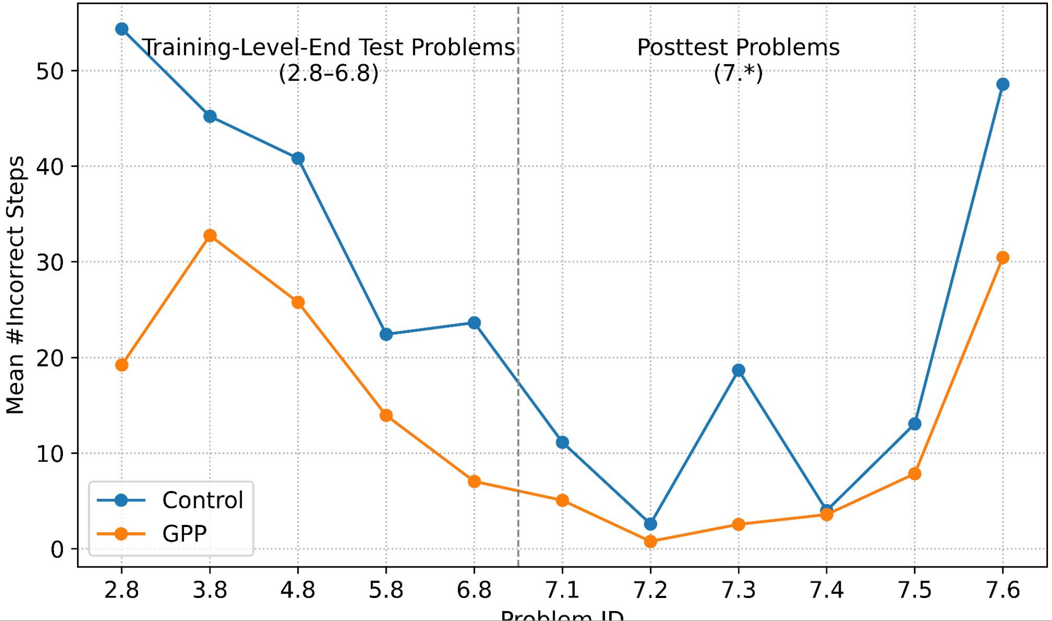
Results showed that students in the GPP condition had higher rule accuracy
Than students in the control condition

RQ1: Time in hours in each section of the tutor

Table 3: Comparison of Total Time to Complete the Tutor across Two Training Conditions.
[Note: **Blue*** indicates a significant difference.]

Time	Control (30)	GPP (46)
Training	**0.81 (0.37)***	1.52 (0.77)
Level-End Test	1.87 (1.43)	1.50 (0.92)
Posttest	1.02 (0.96)	0.91 (0.59)
Total Tutor	4.71 (2.31)	4.75 (1.76)

RQ1: Avg. Incorrect Steps in Level-End Test & Posttest



Test	Control (30 students)	GPP (46 students)	p-value
Level-End Tests	37.4 steps	**19.6**	$p < .001$
Posttests	16.4	**8.4**	$p = .06$

21

RQ2: Impact of GPP - Moderation Analysis on Posttest cat. by Pretest Score

Metric	Test	Control (High)	GPP (High)	p	Control (Low)	GPP (Low)	p
Rule Accuracy	Pretest	67.8 (28.4)	60.1 (29.2)	0.22	39.9 (20.4)	42.3 (20.0)	0.23
	Level-End	67.2 (23.0)	**69.2 (20.6)**	0.56	51.6 (23.5)	**68.3 (19.9)***	< 0.001
	Posttest	78.5 (21.5)	**80.5 (16.9)**	0.68	66.7 (21.2)	**79.1 (17.4)***	< 0.001
Step Count	Pretest	5.2 (1.6)	5.0 (1.3)	0.70	6.9 (2.5)	7.3 (3.0)	0.85
	Level-End	11.1 (4.3)	**9.6 (3.6)***	0.02	**9.7 (3.6)***	11.6 (5.8)	0.03
	Posttest	**8.1 (3.2)**	8.2 (3.3)	0.81	**8.6 (3.8)**	9.6 (4.6)	0.14
Problem Time (minutes)	Pretest	16.2 (17.2)	12.4 (16.2)	0.74	38.3 (17.4)	31.1 (15.1)	0.60
	Level-End	19.8 (14.0)	**18.9 (12.5)**	0.38	24.3 (13.2)	**20.9 (13.5)**	0.20
	Posttest	9.3 (8.7)	**6.2 (9.2)***	0.01	13.6 (12.6)	**10.1 (16.7)**	0.58

22

Takeaways (RQ1 & RQ2)

- By maintaining **a balance between structured scaffolding and student autonomy**, GPPs address critical gaps in previously researched logic PPs.

- This approach proved particularly **beneficial for students with low prior knowledge**, who demonstrated significant improvements in rule application accuracy.

- **High prior knowledge** students benefited from the GPP to **improve their efficiency**, as evidenced by **the reduced number of steps** & **post-test time**.

RQ3: GPP Self-explanation Thematic Analysis methods

For students in the GPP group, we collected students' self-explanation responses after solving each GPP problem (e.g., "How did the subgoals ($G \wedge \neg H$), J help you derive the conclusion?").

We conducted a thematic analysis on 326 unique student explanations from 46 students in GPP group to determine whether and how students were learning about subgoals through GPPs (RQ3).

The themes were derived through an inductive coding process [20] following established thematic analysis methodology [21, 22].

RQ3: Five key GPP Self-explanation Themes

Theme	Description	Student Quote
Task Decomposition	GPPs helped break down logic problems into manageable steps and subgoals.	"They broke down the problem into more understandable smaller problems... like puzzle pieces."
Rule Understanding	Step-specific hints improved understanding of logic rules and when to apply them.	"The hints were useful... something I'll keep in mind in future."
Reduced Difficulty	GPPs reduced cognitive load by showing a solution skeleton and encouraging task planning.	"It allowed me to work on simpler goals and not get distracted on long mistakes."
Backward Reasoning	Though students typically use forward reasoning, GPPs encouraged effective backward chaining.	"They provided obvious stepping stones to move backward through the logic."
Difficulty	A minority found GPPs disrupted their natural problem-solving approach.	"It made the problem harder by disrupting my own way of working through the problem..."

Takeaways (RQ3)

- Three prominent themes, Task Decomposition, Rule Understanding, and Reduced Difficulty, emphasize how the subgoals and step-specific hints made the proofs **more manageable,** potentially **reducing cognitive load**.

- Conversely, several students perceived the structured nature of the proof as **disruptive to their own reasoning processes**. These results suggest that GPPs could be **further enhanced by making them adaptive** to individual student skill levels, which has been shown to be effective for programming [24].

Limitations and Future Work

- A confounding variable in this study could be the **lack of self-explanation prompts in the control group**. In the future, this can be addressed by asking **what aspects of the worked examples** students found helpful when solving problems.
- Future research should explore **a more adaptive implementation of GPPs** to dynamically adjust the amount of scaffolding according to learners' mastery levels and metacognitive needs.

Thank you!
Any questions/feedback?

Sutapa Dey Tithi - stithi@ncsu.edu

Dr. Xiaoyi Tian - xtian9@ncsu.edu

Dr. Min Chi - mchi@ncsu.edu

Dr. Tiffany Barnes - tmbarnes@ncsu.edu

References

1] J. Prather, J. Homer, P. Denny, B. A. Becker, J. Marsden, G. Powell, Scaffolding task planning using abstract parsons problems, in: IFIP World Conference on Computers in Education, Springer, 2022, pp. 591–602.

[2] P. Shabrina, B. Mostafavi, S. D. Tithi, M. Chi, T. Barnes, Learning problem decomposition recomposition with data-driven chunky parsons problems within an intelligent logic tutor., EDM 2023.

[3] A. Renkl, R. K. Atkinson, U. H. Maier, From studying examples to solving problems: Fading worked-out solution steps helps learning, in: CogSci conf. 2000, pp. 393–398.

[4] J. Sweller, Cognitive load during problem solving: Effects on learning. Cognitive Science 12 (1988) 257–285.

[5] D. Wood, J. S. Bruner, G. Ross, The role of tutoring in problem solving. Journal of child psychology and psychiatry, 17 (1976) 89–100.

[6] F. Paas, A. Renkl, J. Sweller, Cognitive load theory and instructional design: Recent developments, Educational psychologist 38 (2003) 1–4. F. Nievelstein, T. Van Gog, G. Van Dijck, H. P. Boshuizen, The worked example and expertise reversal effect in less structured tasks: Learning to reason about legal cases, Contemporary Educational Psychology 38 (2013) 118–125.

[8] A. Renkl, The worked-out examples principle in multimedia learning. (2005).

References

[9] P. Denny, A. Luxton-Reilly, B. Simon, Evaluating a new exam question: Parsons problems, in: ICER 2008, pp. 113–124.

[10] N. Weinman, A. Fox, M. A. Hearst, Improving instruction of programming patterns with faded parsons problems, in: CHI 2021, pp. 1–4.

[11] V. Karavirta, J. Helminen, P. Ihantola, A mobile learning application for parsons problems with automatic feedback, Koli Calling 2012 pp 11–18.

[12] B. J. Ericson, J. D. Foley, J. Rick, Evaluating the efficiency and effectiveness of adaptive parsons problems, in: ICER 2018, pp. 60–68.

[13] S. Poulsen, M. Viswanathan, G. L. Herman, M. West, Evaluating proof blocks problems as exam questions, ACM Inroads 13 (2022) 41–51.

[14] B. B. Morrison, L. E. Margulieux, B. Ericson, M. Guzdial, Subgoals help students solve parsons problems, in: SIGCSE TS 2016, pp. 42–47.

[15] L. Margulieux, R. Catrambone, Using learners' self-explanations of subgoals to guide initial problem solving in app inventor, ICER 2017, p. 21–29.

[16] A. Renkl, Worked-out examples: Instructional explanations support learning by self-explanations, Learning and instruction, 12 (2002) 529–556.

References

[17] K. R. Koedinger, A. T. Corbett, C. Perfetti, The knowledge-learning-instruction framework: Bridging the science-practice chasm to enhance robust student learning, Cognitive science 36 (2012) 757–798.

[18] K. VanLehn, R. M. Jones, M. T. Chi, A model of the self-explanation effect, The journal of the learning sciences 2 (1992) 1–59.

[19] K. Bisra, Q. Liu, J. C. Nesbit, F. Salimi, P. H. Winne, Inducing self-explanation: A meta-analysis, Educational Psychology Review 30 (2018) 703–725.

[20] D. R. Thomas, A general inductive approach for analyzing qualitative evaluation data, American journal of evaluation 27 (2006) 237–246.

[21] G. Guest, K. M. MacQueen, E. E. Namey, Applied thematic analysis, Sage publications, 2011.

[22] M. Muller, Curiosity, creativity, and surprise as analytic tools: Grounded theory method, in: Ways of Knowing in HCI, Springer, 2014, pp. 25–48.

[23] S. Kalyuga, The expertise reversal effect, in: Managing cognitive load in adaptive multimedia learning, IGI Global, 2009, pp. 58–80.

[24] C. Haynes-Magyar, B. Ericson, The impact of solving adaptive parsons problems with common and uncommon solutions, in: Koli Calling, 2022, pp. 1–14.