



Theory-Guided GPT Tutoring and Math Placement Preparation in Community College

A Quantitative Research Proposal · Andrew Egbert
Instructional Design & Development · Idaho State University
EDLP 7706 · Dr. Rick Wagoner · April 22, 2026

The Developmental Math Pipeline

Raise readiness before placement, not after. Incoming CWI students face a math placement test that gates their entire course pathway — yet the Math Solutions Center reports many students arrive chronically underprepared for that gatekeeping moment.

The Sequence Problem

Underprepared placements route students into developmental sequences that delay — and often derail — progress toward credit-bearing coursework (Bailey et al., 2010; Hodara & Jaggars, 2014).

Who Bears the Cost

The burden is not evenly distributed. Low-income and first-generation students are disproportionately placed into remediation — and every additional developmental course raises attrition risk for those who can least afford to drop out (Ngo & Melguizo, 2016).

① Each remedial course added to a student's pathway measurably reduces the probability of credential completion.

Test Whether Design Quality in an LLM Tutor Produces Measurable Gains

The central question: does a GPT tutor engineered around instructional theory outperform an unstructured GPT tutor on placement-aligned math outcomes? Three converging realities make this question both timely and consequential.

Already Here

LLMs are scalable, low-cost, and already in students' hands — the infrastructure barrier has collapsed. Access is effectively universal.

Evidence Base

ChatGPT-generated help can match human-authored help on mathematics problem-solving outcomes (Pardos & Bhandari, 2024). The floor is established.

Design Matters

Access is not instruction. Design quality — scaffolding, feedback loops, mastery pacing — is the decisive differentiator (VanLehn, 2011).

Three Streams, One Gap

The proposal synthesizes three intersecting bodies of research into an integrated theoretical framework that directly motivates both research questions.

Stream 1 · AI / GPT-Based ITS in Math

Pardos & Bhandari (2024); Wang et al. (2024) — Tutor CoPilot; Nickl et al. (2024); Létourneau et al. (2025). Establishes feasibility and benchmarks for AI-tutoring effectiveness in quantitative domains.

Stream 2 · Distributed Scaffolding

Puntambekar (2022); Puntambekar & Kolodner (2005); Tabak (2004). Frames how scaffolding distributed across tools and interactions supports learner agency and knowledge construction.

Stream 3 · Interaction & Engagement

Chi (2009) — ICAP framework; Chi, Roy & Hausmann (2008); VanLehn (2011) synthesis. Defines the gradient from passive to interactive engagement and its link to learning outcomes.

❏ **The Gap:** No community-college placement-prep RCT comparing theory-guided vs. baseline GPT tutors exists. No study models which interaction patterns predict gains in this population.

Two Questions, Four Hypotheses

RQ1 · Group Comparison

Do students using a theory-guided GPT tutor score higher on a placement-aligned posttest than students using a baseline GPT tutor, controlling for pretest performance?

H1

Theory-guided group mean posttest score > baseline group mean posttest score, controlling for pretest.
Directional hypothesis.

RQ2 · Within-Group Regression

Among theory-guided users, which interaction variables predict posttest score, controlling for pretest?

H2a

Hint-use frequency positively predicts higher posttest scores.

H2b

Mastery-check completion rate positively predicts posttest scores.

H2c

Attempts before solution reveal positively predicts posttest scores — operationalizing effortful retrieval practice.

Two Conditions, Run in Parallel

Participants are assigned to one of two lanes for the full prep window. The theory-guided condition receives a fully scaffolded experience; the baseline condition accesses an unstructured LLM with equivalent content but no engineered pedagogical features.

A · Theory-Guided GPT Tutor

Step-based scaffolding · Tiered hints · Mastery checks · Attempts-before-reveal gate. Automated event-level interaction logs captured throughout the prep window.

Pretest → Prep Window → Posttest

Both conditions complete a placement-aligned pretest (blueprint form) before tutor access and a parallel-form posttest after. RQ1 uses cross-condition comparison; RQ2 uses within-condition regression on interaction logs.

B · Baseline GPT Tutor

No scaffolding features · No tiered hints · No mastery checks · No interaction logging. Students receive the same content domain with free-form LLM assistance only.

POPULATION, SAMPLE & DATA PROVENANCE

N $\approx$ 60–100 · Three Study-Generated Data Stream

All data are generated by the study itself — not drawn from any secondary dataset. The population is incoming or continuing CWI students required to take or retake the math placement assessment.

01

Placement-Aligned Pretest

Built from the local CWI placement test blueprint. Administered on-site before any tutor access. Used as a covariate to control for prior knowledge in both analyses.

02

Placement-Aligned Posttest

Parallel form drawn from the same blueprint. Administered after the full prep window concludes. Serves as the primary outcome variable for both RQ1 and RQ2.

03

Automated Interaction Logs

Event-level records generated by the theory-guided tutor: every hint request, mastery-check item response, and solution attempt per student. Source data for all RQ2 predictors.

Variable Architecture Mapped to Research Questions

Role	Variable	Operationalization	RQ
Dependent	Posttest Score	Continuous 0–100%, placement-aligned parallel form	RQ1 & RQ2
Independent	Tutor Condition	0 = baseline, 1 = theory-guided	RQ1
Control	Pretest Score	Continuous covariate, prior knowledge control	RQ1 & RQ2
Predictor	Hint-Use Frequency	Count per student, optionally per-item	RQ2 · H2a
Predictor	Mastery-Check Rate	Proportion (0–1) of items completed	RQ2 · H2b
Predictor	Attempts Before Reveal	Continuous average per item	RQ2 · H2c
Optional Controls	Demographics	Age, first-gen status, prior math, test anxiety (if N permits)	RQ2

Full variable-to-RQ crosswalk including coding schemes and reliability considerations is provided in Table 1 of the accompanying paper.

Group Comparison: Independent-Samples t-Test with Pre-Registered Fallback

1

Preliminary

Descriptive statistics and baseline equivalence check on pretest scores across both conditions.

2

Assumption Checks

Normality assessed via histograms, skewness/kurtosis. Variance homogeneity via Levene's test.

3

Primary Test

Independent-samples t-test on posttest scores. Directional alternative consistent with H1.

4

Pre-Registered Fallback

If pretest scores differ significantly between conditions, ANCOVA is substituted with pretest as covariate — decision rule registered in advance.

5

Effect Size Reporting

Cohen's d or partial η^2 reported with 95% confidence intervals for all primary comparisons.

Hierarchical OLS Regression · Theory-Guided Condition Only

1

Block 1

Pretest · Prior-knowledge control entered first to isolate interaction effects.

2

Block 2

Hints + Mastery + Attempts · Three focal interaction predictors — core test of H2a-H2c.

3

Block 3

Interaction Term · Exploratory Hint × Mastery-check product term.

Assumption Checks

- Linearity · scatterplots of predictors vs. residuals
- Residual independence · conceptual review
- Homoscedasticity · residual plot inspection
- Residual normality · Q-Q plot
- Multicollinearity · VIF > 10 flagged

Reporting Standards

- B , SE , β per predictor
- R^2 and ΔR^2 for Block 2 entry
- Robustness: pooled model with condition dummy

$$\hat{Y} = \beta_0 + \beta_1(\text{Pretest}) + \beta_2(\text{Hints}) + \beta_3(\text{MasteryRate}) + \beta_4(\text{Attempts}) + \varepsilon$$

What This Study Adds — and What It Cannot Claim

Core Contribution

A direct, interaction-level test of whether instructional theory — when actually operationalized inside a GPT tutor — produces placement-aligned learning gains in a community-college setting. This context is largely absent from current AI-tutoring research, which skews toward K-12 and four-year institutions. The study also models which specific interaction behaviors within a theory-guided tutor predict outcomes, advancing both design practice and theoretical understanding of effortful engagement in LLM-mediated learning.

Acknowledged Limitations

- Single-site convenience sample limits generalizability
- Voluntary self-selection introduces participation bias

Mitigations

- Pre-registered decision rules for RQ1 analyses
- Transparent assumption-check reporting throughout
- Effect sizes reported with 95% confidence intervals

Thank you.

Andrew Egbert · aedsgn@gmail.com Instructional Design &
Development · Idaho State University
EDLP 7706 · Dr. Rick Wagoner · April 22, 2026



References

- Bailey, T., Jeong, D. W., & Cho, S.-W. (2010). Referral, enrollment, and completion in developmental education sequences in community colleges. *Economics of Education Review*, 29(1), 1–12. <https://doi.org/10.1016/j.econedurev.2009.12.001>
- Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, 1(1), 73–105. <https://doi.org/10.1111/j.1756-8765.2008.01005.x>
- Chi, M. T. H., Roy, M., & Hausmann, R. G. M. (2008). Observing tutorial dialogues collaboratively: Insights about human tutoring effectiveness from vicarious learning. *Cognitive Science*, 32(2), 301–341.
- Hodara, M., & Jaggars, S. S. (2014). An examination of the impact of accelerating community college students' progression through developmental education. *The Journal of Higher Education*, 85(2), 246–276.
- Létourneau, A., Martineau, M. D., Charland, P., Karran, J. A., Boasen, J., & Léger, P. M. (2025). A systematic review of AI-driven intelligent tutoring systems in K-12 education. *npj Science of Learning*, 10, 29. <https://doi.org/10.1038/s41539-025-00320-7>
- Ngo, F., & Melguizo, T. (2016). How can placement policy improve math remediation outcomes? *Educational Evaluation and Policy Analysis*, 38(1), 171–196.
- Nickl, M., Sommerhoff, D., Radkowsch, A., Huber, S. A., Bauer, E., Ufer, S., Plass, J. L., & Seidel, T. (2024). Effects of real-time adaptivity of scaffolding. *Learning and Instruction*, 94, 101994. <https://doi.org/10.1016/j.learninstruc.2024.101994>
- Pardos, Z. A., & Bhandari, S. (2024). ChatGPT-generated help produces learning gains equivalent to human tutor-authored help on mathematics skills. *PLOS ONE*, 19(5), 1–18. <https://doi.org/10.1371/journal.pone.0304013>
- Puntambekar, S. (2022). Distributed scaffolding: Scaffolding students in classroom environments. *Educational Psychology Review*, 34(1), 451–472. <https://doi.org/10.1007/s10648-021-10000-0>
- Puntambekar, S., & Kolodner, J. L. (2005). Toward implementing distributed scaffolding. *Journal of Research in Science Teaching*, 42(2), 185–217.
- Tabak, I. (2004). Synergy: A complement to emerging patterns of distributed scaffolding. *The Journal of the Learning Sciences*, 13(3), 305–335.
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197–221. <https://doi.org/10.1080/00461520.2011.611369>
- Wang, R. E., Ribeiro, A. T., Robinson, C. D., Loeb, S., & Demszky, D. (2024). Tutor CoPilot: A human-AI approach for scaling real-time expertise. *arXiv*. <http://arxiv.org/abs/2410.03017>