Data Analysis Tutoring Effect Evaluation

Term: 2025Q1

Masae Wen
Mentor: Bryan Queme
8/14/2025

Agenda

❖ Executive Overview

➢ The Question

➢ The Answer

➢ The Evidences

➢ The Insights

❖ Comprehensive Statistical Overview (Appendix)

Executive Overview

Agenda

- ❖ Executive Overview
  - ➢ The Question
  - ➢ The Answer
  - ➢ The Evidences
  - ➢ The Insights

The Big Question: Does Tutoring Work?

- **Our Question:** Are students who use our tutoring sessions making more progress than those who don't?

- **Our Goal:** To understand the effect of tutoring on student progress
  - Key metric of learner success: competency completion (0 - 12)

- **Data Used:** Cleaned Gradebook and Cleaned Post Tutor Form
  - Data Range: 2025 Q1
    - Overall: 2558 observations
    - Tutored Group: 137 observations
    - Untutored Group: 2421 observations

Agenda

- ❖ Executive Overview
  - ➢ The Question
  - ➢ The Answer
  - ➢ The Evidences
  - ➢ The Insights

The Answer is a Resounding YES!

Finding 1: Tutored students complete more than DOUBLE the coursework.

- Students who receive tutoring complete on average **2.16 times*** as many competencies as students who don't.
  - *95% Confidence Interval (CI): [1.56, 3.01]
  - $p < 0.00001$ (highly statistically significant)

Finding 2: Tutoring Appears Critical for Student Engagement

- **77%** of untutored students completed zero competencies (untutored: 2421)
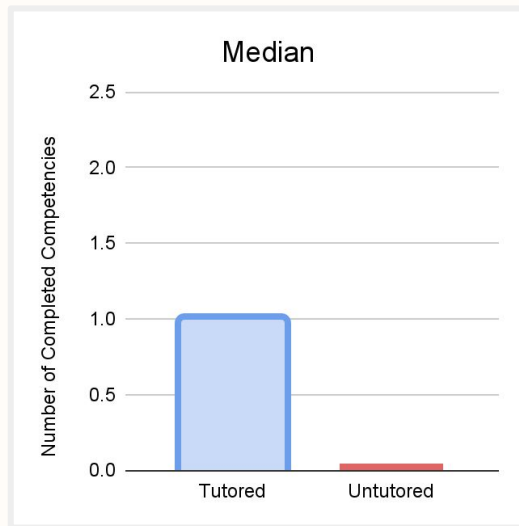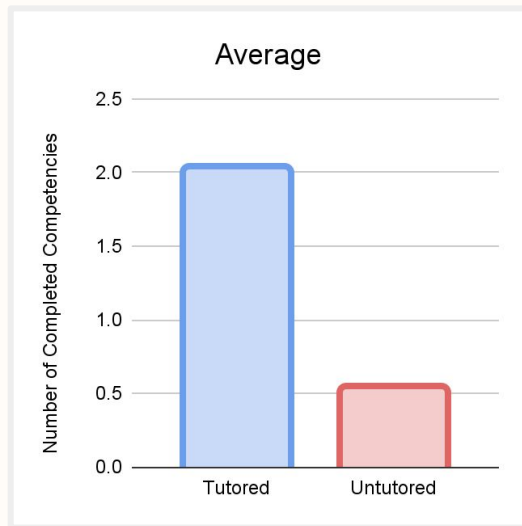- **33%** of tutored student completed zero competencies (tutored: 137)

CALBRIGHT
COLLEGE

Agenda

❖ Executive Overview

    ➢ The Question

    ➢ The Answer

    ➢ The Evidence

    ➢ The Insights

A First Look: A Powerful Initial Effect

Competency Completion

Average

Number of Completed Competencies

Tutored — ~2.0
Untutored — ~0.5

Median

Number of Completed Competencies

Tutored — ~1.0
Untutored — ~0.0

Average Completions:
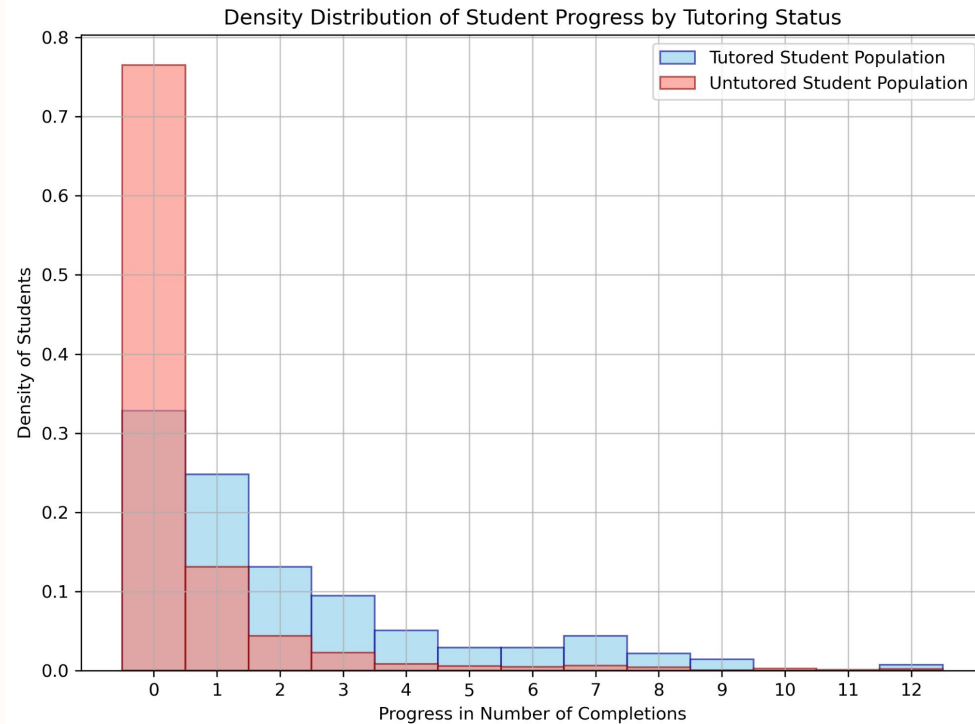2.0 (Tutored) vs. 0.5 (Untutored)

Median Completions:
1.0 (Tutored) vs. 0.0 (Untutored)

Initial Conclusion:
The raw data shows a strong positive effect, but its complexity requires a deeper analysis for an accurate measure.

How Tutoring Changes the Path to Completion

Density Distribution of Student Progress by Tutoring Status

Untutored group (in red)

- Majority at **Zero Completions**.

Tutored group (in blue)

- Many students showed **1 or more Completions**.

Measuring the Impact: A Rigorous Approach

Method 1: Confirming the Difference

- A **Wilcoxon Rank-Sum test** confirms the group difference is **highly statistically significant** (p < 0.00001).

A Rigorous Approach to Measuring the Impact

Method 2: Quantifying the Effect Size

- A **Zero-Inflated Negative Binomial (ZINB) model** was used to accurately measure the impact, accounting for the data's excess zeros and skewed distribution.

- ✅**Finding 1 Confirmation:** Tutored students complete **2.16 times*** as many competencies as untutored students.
  - *95% Confidence Interval (CI): [1.56, 3.01]
  - $p < 0.00001$

Tutored Students Show Drastically Less Disengagement

Tutored

- No Progress
- Progress

33.3%

66.7%

Untutored

- No Progress
- Progress

23.5%

76.5%
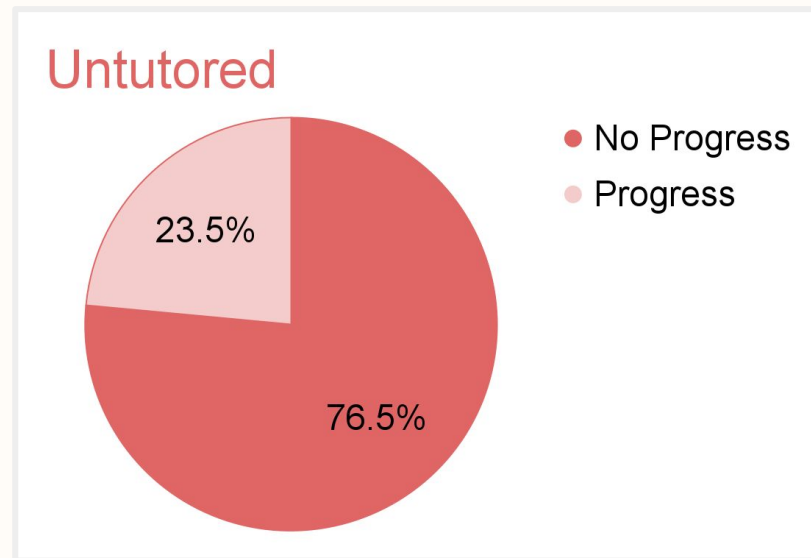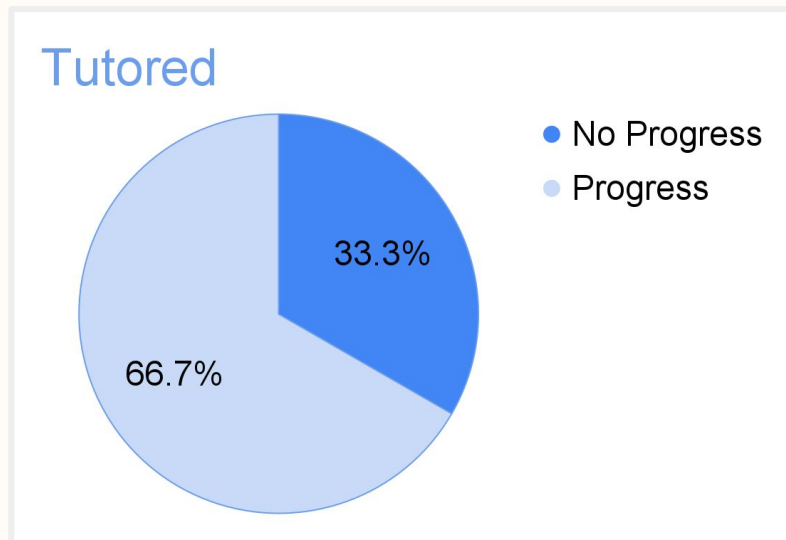
No Progress: Number of Completed Competencies = 0

Progress: Number of Completed Competencies > 0 (1 or more)

✅**Finding 2 Confirmation: 33%** No Progress (Tutored) vs. **77%** No Progress (Untutored)

Agenda

❖ Executive Overview

    ➢ The Question

    ➢ The Answer

    ➢ The Evidences

    ➢ The Insights

Our Findings Are Robust and Reliable

Robust Analysis: Advanced statistical methods ensure accurate and reliable conclusions.

Strong Evidence: Differences are not due to chance and are statistically significant.

Clear Impact: Tutoring is a major factor in observed improvements.

Bottom Line: Tutoring Works!

The data confirms tutoring more than doubles competency completions

CALBRIGHT
COLLEGE

Identifying Key Student Groups

Tutoring Impact on Student Progress	Tutored	Untutored
Made Progress (≥1)	67% Successful Engagement	23% Self-Starters
Made No Progress (0)	33% Opportunity Group **"Quick Win"**	77% Challenge

Tutoring Success:
67% of tutored students achieved progress.

Challenge:
77% of untutored students made no progress.

The "Opportunity Group":
33% of tutored students made no progress, showing a quick win chance.

CALBRIGHT
COLLEGE

What This Means for Us: Actionable Insights

Tutoring is a Proven Success: Effective tool for student progress.

Recommendation 1: Use Tutoring as a **Targeted** Intervention.

- Proactively support the **"Opportunity Group"** with personalized outreach.
- Implement an onboarding survey to identify **"at-risk" students** even earlier.

Recommendation 2: Promote and Expand Tutoring.

- Increase awareness and accessibility for all students via Canvas and Email.

.

Additional Tutoring Impact: Supports Our Instructors

- Effective tutoring reduces repetitive questions for instructors.

- Frees up instructor time for faster grading and feedback.

- Leads to higher overall student satisfaction.

.

Looking Ahead: Next Steps

- Continue this analysis for future terms to track ongoing impact.

    - Identify "Opportunity Group" students sooner than Quarterly analysis if possible.

- Explore applying this analytical framework to other Calbright programs.

.

SMALL STEPS ARE STILL PROGRESS

Acknowledgement

- **Byan Queme:** Data Analysis Program Mentor

- **Richard Garcia:** ExlI Intern Supervisor

Thank You!

Questions?

Appendix

Evaluating the Impact of Peer Tutoring on Competency Completion in the Calbright Data Analysis Program

A Statistical Analysis Using Zero-Inflated Negative Binomial Regression

CALBRIGHT COLLEGE

Agenda

❖ Comprehensive Statistical Overview

➢ Introduction

➢ Exploratory Data Analysis

➢ Modeling

➢ Conclusion

➢ Future Direction

CALBRIGHT
COLLEGE

Introduction & Research Question

- **Objective:** To evaluate the effect of a peer tutoring program on student competency completion within the Calbright Data Analysis Program.

- **Dataset:** The analysis utilizes cleaned gradebook data, separating learners into two groups: those who received tutoring (n=137) and those who did not (n=2,421).
    - Data Range: 2025 Q1
        - Overall: 2558 observations
            - Key Identifier: Calbright Student Email
    - Tools: Google Sheets, Python (Jupyter Notebook)

- **Response Variable:** The primary outcome is Completion Count Progress (Competency Completion), an integer count variable ranging from 0 to 12.

Exploratory Data Analysis & Modeling Rationale

Data Characteristics: The response variable exhibits properties that violate assumptions of standard linear models:

- **Count Data:** The outcome is a non-negative integer.

- **Excess Zeros:** A large proportion of students have zero completions, particularly in the untutored group where the median is 0 and the 75th percentile is also 0. For tutored students, the median is 1.

- **Overdispersion:** The variance in completion counts is substantially larger than the mean for both groups, as seen in the descriptive statistics.
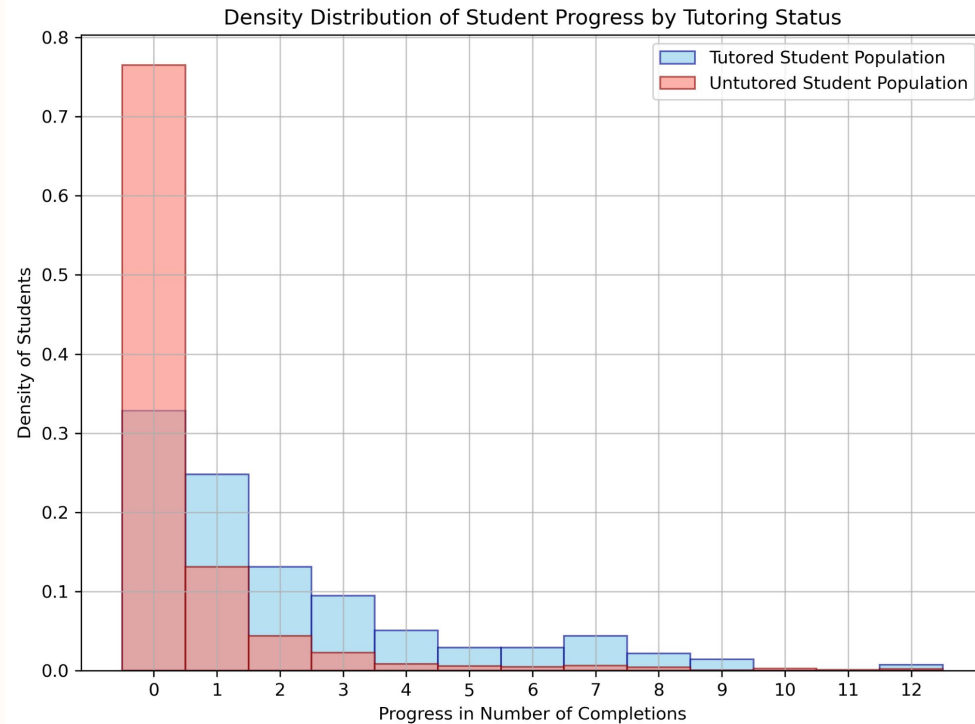
Exploratory Data Analysis & Modeling Rationale

Empirical Cumulative Distribution of Student Progress by Tutoring Status
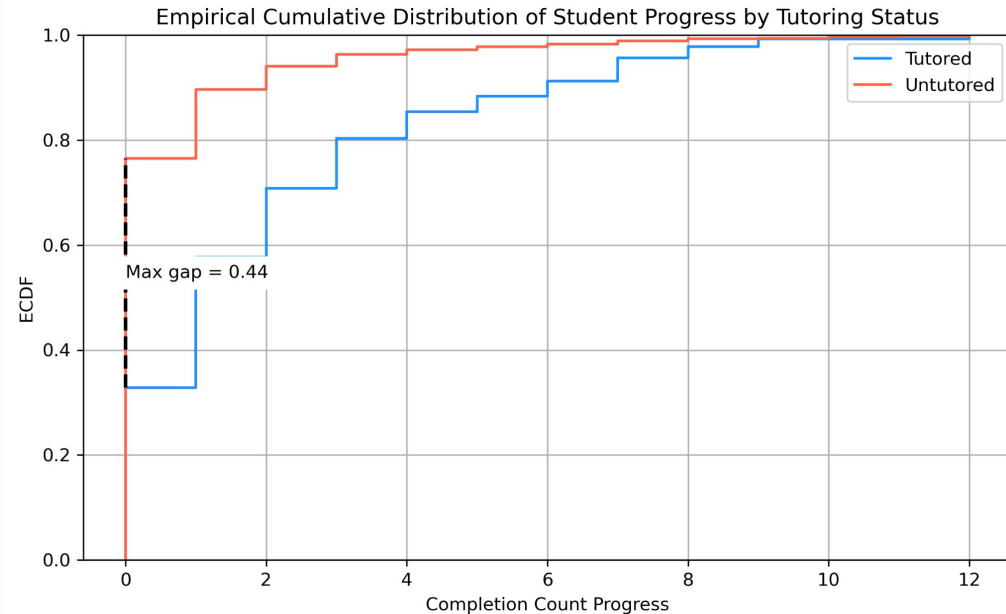
Visual Justification 1:

- The **ECDF plot** shows a maximum gap of 0.44 between the two groups at a completion count of 0, where 76.5% of untutored students have zero completions compared to only 32.9% of tutored students.

Exploratory Data Analysis & Modeling Rationale

Density Distribution of Student Progress by Tutoring Status

Visual Justification 2:

- A **density histogram** visually confirms the high concentration of zeros in the untutored population (~77%) versus the tutored population (~33%).

Exploratory Data Analysis & Modeling Rationale

Chosen Analytical Framework:

- A **Wilcoxon Rank-Sum test** confirmed a highly significant difference between the groups ($p < 0.00001$), supporting further investigation.

- Based on these characteristics, Negative Binomial (NB) and Zero-Inflated Negative Binomial (ZINB) regression models were selected as appropriate tools for the analysis.

Model Selection & Diagnostics

Systematic Comparison: A systematic model selection process was undertaken to identify the best-fitting and most stable model. Configurations tested included:

- **Model Type:** Standard Negative Binomial vs. ZINB.

- **ZINB Inflation Structure:** A full model (with 'Tutored' as a predictor in both count and zero-inflation components) vs. an intercept-only inflation model.

- **Optimizers:** lbfgs, bfgs, and newton were tested to ensure convergence.

CALBRIGHT
COLLEGE

Model Selection & Diagnostics

Selection Criteria: Models were evaluated based on Log-Likelihood, AIC, BIC, Pseudo R^2, and successful convergence without warnings (e.g., Hessian inversion failures).

Winning Model: The ZINB model with a full inflation structure, using the lbfgs optimizer and 100 iterations, was selected. It demonstrated superior fit over the standard NB model:

Model	Log-Likelihood	AIC	BIC
NB	-2470.26	4946.51	4964.06
ZINB	**-2456.79**	**4923.57**	**4952.81**

Final Model Interpretation: ZINB Results

The selected ZINB model (Full_Infl_LBFGS_Iter100) provides the following insights:

Count Model Component (Negative Binomial):

- The coefficient for **Tutored** is **0.7722 (p < 0.00001)**. The Incidence Rate Ratio (IRR) is e0.7722≈**2.16***.
  - *95% Confidence Interval (CI) of [1.56, 3.01]

- **Interpretation:** For learners not in the "always-zero" latent class, receiving tutoring is associated with a **116% increase** in the expected number of completed competencies, holding other factors constant.

Final Model Interpretation: ZINB Results

Zero-Inflation Component (Logit):

- The coefficient for **inflate_Tutored** is **-12.0146 (p = 0.879)**.

- **Interpretation:** The effect of tutoring on the log-odds of being in the "always-zero" group is not statistically significant. There is no evidence that tutoring affects whether a student belongs to a latent class of structural non-completers.

CALBRIGHT COLLEGE

Final Model Interpretation: ZINB Results

Dispersion Parameter:

- **alpha** is **1.9521 (p < 0.00001)**.

- **Interpretation:** The significant and positive alpha confirms the presence of overdispersion, validating the use of a Negative Binomial distribution over a Poisson.

CALBRIGHT
COLLEGE

Conclusion & Summary of Findings

- **Primary Finding:** Tutoring is strongly associated with higher competency completion. The model estimates that tutoring more than doubles the expected completion count for students who are actively participating in the course.

- **Secondary Finding:** The mechanism of this effect appears to be through increasing progress among active learners, not by changing the status of structurally disengaged learners.

- **Model Validation:** The data strongly supports the use of a ZINB model over simpler alternatives, as evidenced by the significant overdispersion parameter and the superior AIC/BIC fit statistics compared to a standard Negative Binomial model.

CALBRIGHT
COLLEGE

Limitations & Future Directions

- **Observational Data:** This analysis is observational, not experimental. Causality cannot be definitively established due to potential self-selection bias (i.e., more motivated students may be more likely to seek tutoring).

- **Future Work:** A more robust analysis could include controlling for covariates such as program start date, demographics, or prior academic history. A randomized controlled trial would be the ideal design to isolate the causal effect of tutoring.

CALBR★GHT
COLLEGE

Thank You!

Questions?

CALBRIGHT COLLEGE

END