



# **TEACHING READING AND SUMMARIZING VIA ARGUMENT-FOCUSED TEXT ANNOTATION**

**By: Jessica Macaluso**



# Background: Summarization

- One reading skill students often struggle with is summarizing.
  - **Summarization** is the ability to convey the most crucial information from a text in a concise and clear form (i.e., the gist of the material).
- The skill of summarization is critical in formal education because comprehension of complex texts is essential to many academic disciplines like science, humanities, and law.

# Background: Annotation

- **Annotation** (i.e., highlighting key terms and ideas) has been identified as one of multiple cognitive literacy strategies that can help students when they are looking at structure, examining ideas, extracting meaning, and communicating understandings.
  - Annotation-based learning has proven to be beneficial: first-year college students in a randomized controlled trial did better on future comprehension exams when annotating narrative texts compared to those using a question-answer strategy.
- Thus, it seems feasible that annotation could be a useful activity for law students, such as when understanding legal opinions.

# Project Goal & Predictions

- **Project Goal:** I want to assess if annotation improves students' learning and, if so, to what extent.
  - Can the Gloss annotation tool help struggling students improve and close learning gaps?
- **Prediction:** I predict that reading *and annotating* legal opinions, rather than only reading them, will improve students' understanding of factor-based legal frameworks.

# Methodology

**Gloss** is an online annotation tool designed to help law students analyze and annotate legal cases, improving comprehension and legal reasoning skills.

gloss

Annotator

Data

Types **23** Annotations **11**

Auto **0**

1A Furtive Movement

1B Physical Appearance of Nervous...

1C Nervous Behavior

1D Suspicious or Inconsistent Answ...

2E Motorist License or Identification

2F Driver Status

2G Legal Indications of Drug Use

2G Refused Consent

2I Motorist's Appearance Related to ...

3J Possible Drug Route

Text

Suspicion Project 3.0 unitedstates\_v\_powell Done

Map

Sentence 8: Mr. Powell responded in the negative and then, without any prompting from the deputy, began volunteering information.  
Label: Nervous Behavior or Appearance

Sentence 9: Mr. Powell told Deputy Trammel that he rented the Yukon to move from Colorado to Lexington, Kentucky, where he planned to open a hip-hop clothing store.  
Label: Unusual Travel Plans, Unusual Vehicle Ownership

Sentence 10: He explained that he was transporting tires for his Chevy Suburban, which he owned but left at home in Colorado, because the tires were not made for long distance highway driving.  
Label: Unusual Travel Plans

Sentence 11: He added that he planned to unload the tires in Lexington before returning to Colorado to turn in his rental vehicle and pick up his Suburban.  
Label: Unusual Travel Plans

Sentence 12: Mr. Powell went on to say that he was traveling with his brother, who was "up ahead" because he did not wait for Mr. Powell when Mr. Powell stopped to use the restroom.  
Label: Unusual Travel Plans

Sentence 13: During Mr. Powell's narrative, Deputy Trammel observed that Mr. Powell appeared "extremely nervous" because, in addition to his talkativeness, he was breathing heavily and avoiding eye contact.

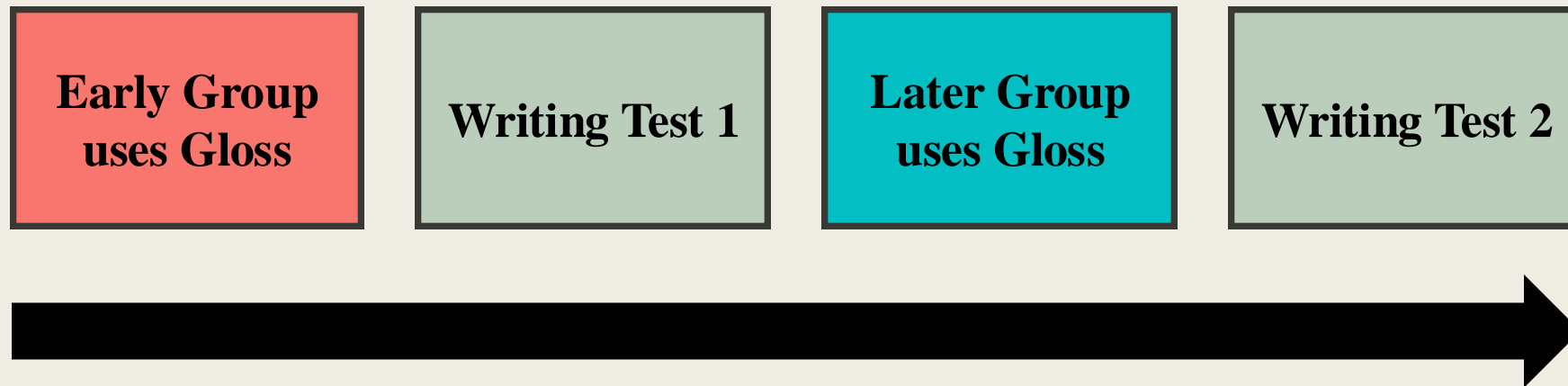
5

# Methodology

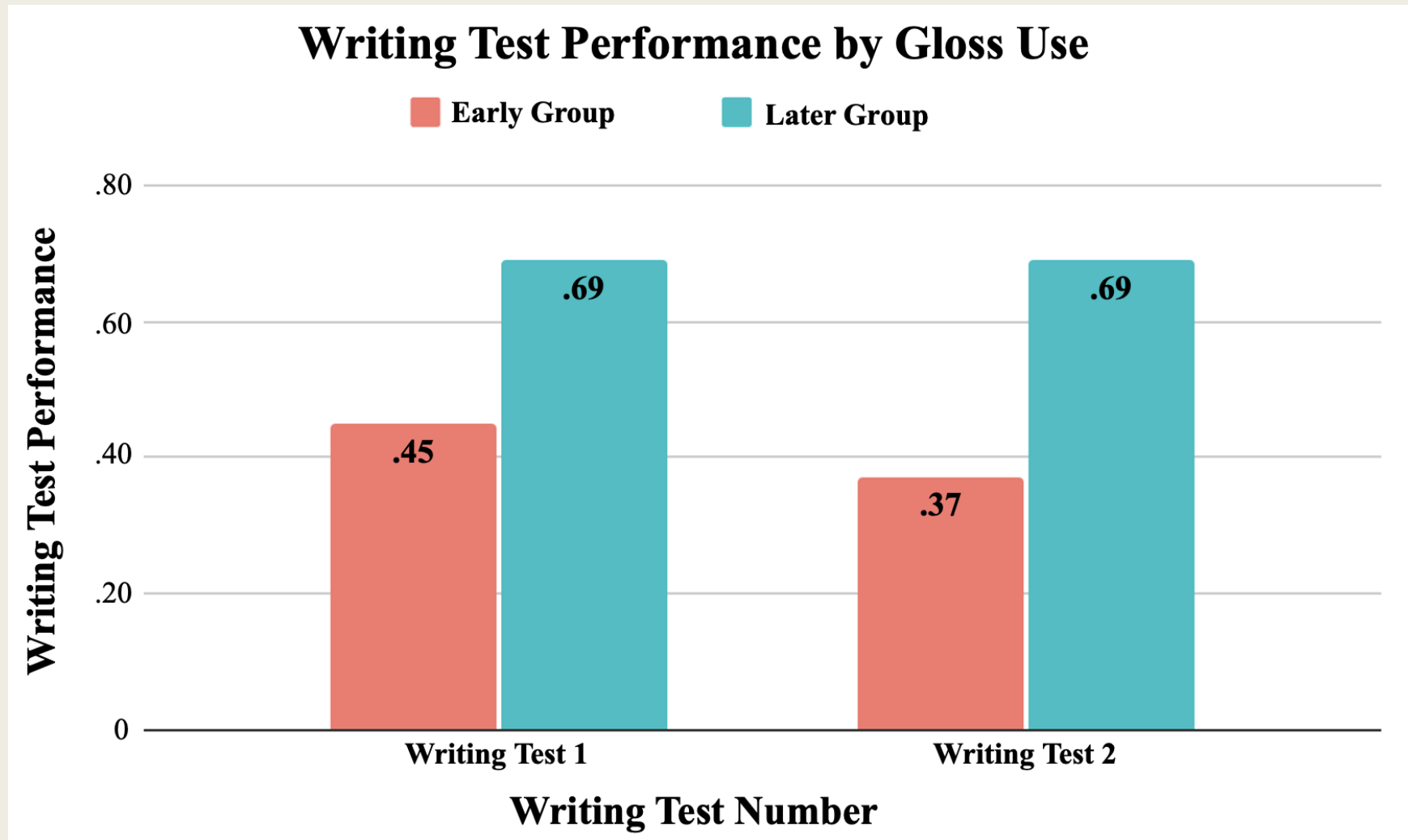
- To evaluate the impact of the Gloss annotation tool, I conducted a user study with first-year law students ( $N = 33$ ) in a legal writing class. Participants were randomly assigned to one of two groups in a waitlist-controlled design:
  - $\sim 1/2$  of the students ( $n = 18$ ) used Gloss early in the term
  - $\sim 1/2$  of the students ( $n = 15$ ) used Gloss later in the term
- This approach allowed us to compare how using Gloss at different times (i.e., “Early Group” and “Later Group”) influenced annotation behavior and learning outcomes.

# Methodology

- The outcome measure was a writing test embedded in a real class assignment, evaluated by the course instructor using their standard grading rubric.



# General Findings



Students who used Gloss early performed better on the first writing test than on the second, while students who used Gloss later performed consistently across both writing tests.

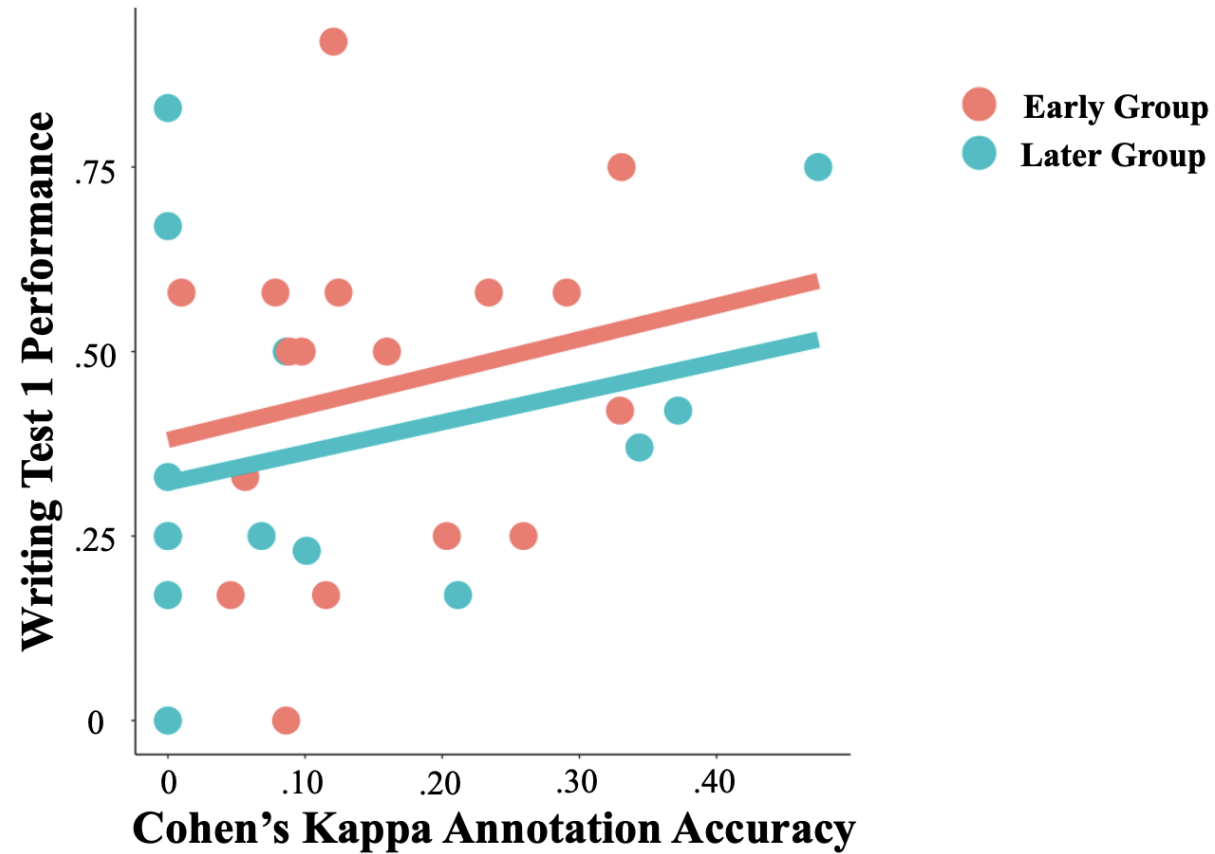


# Additional Analyses

- I also used Cohen's Kappa for to evaluate students' annotation accuracy.
  - **Cohen's Kappa** is a statistical measure used to assess the agreement between two or more raters when evaluating annotations, beyond what would be expected by chance.
- For example, if two raters independently assess whether student annotations correctly identify key legal arguments, Cohen's Kappa helps determine how reliably they agree.
  - Values range from -1 to +1: a value of +1 indicates perfect agreement, 0 suggests agreement is at chance level, and negative values mean agreement is lower than expected by chance.

# Additional Findings

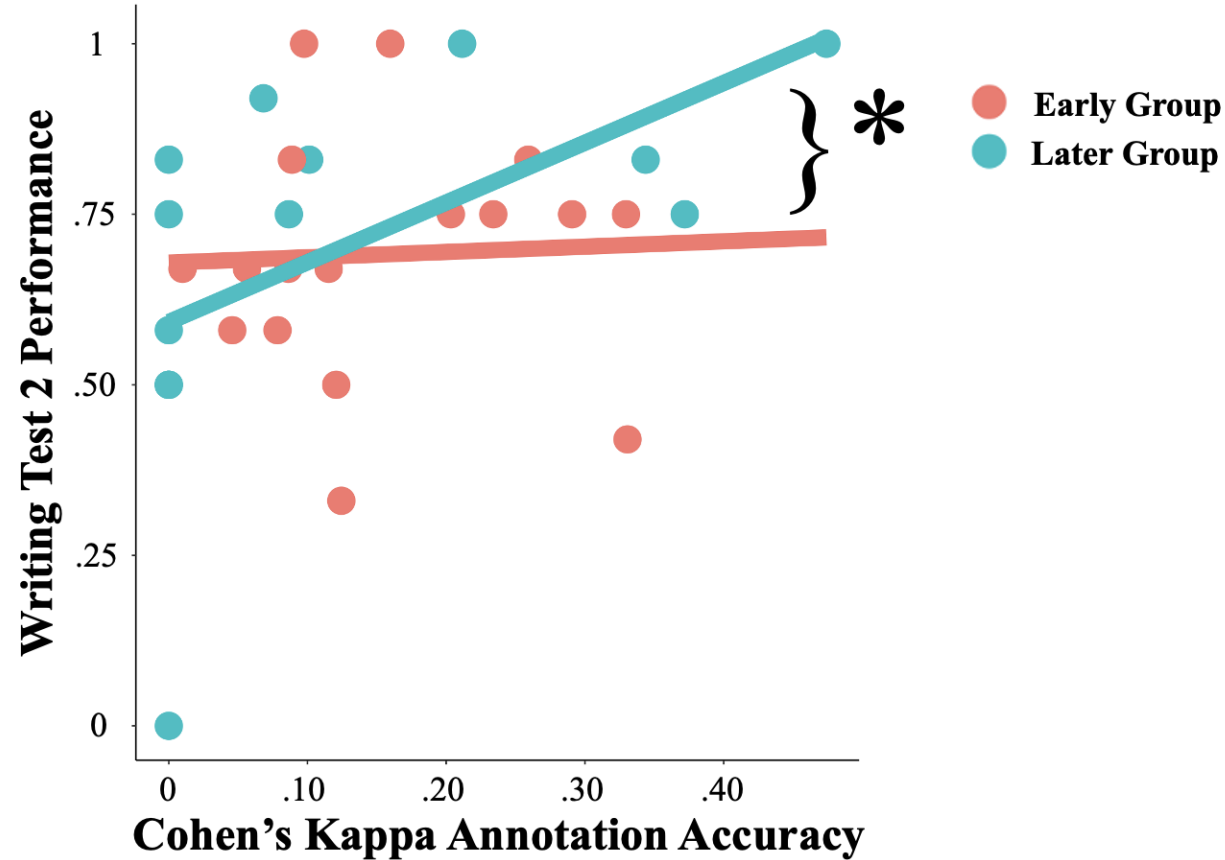
**Positive Relationship: as Writing Test 1 Performance Increases, Annotation Accuracy Increases for Both Gloss Groups**



For Writing Test 1, all students, regardless of condition, show a positive relationship between annotation accuracy and test score.

# Additional Findings

## Gloss Helps Weaker Students Catch Up: Softens Relationship Between Early Accuracy & Writing Test 2 Performance



For Writing Test 2, having access to Gloss helped ensure that early poor performance did not predict low exam scores. Gloss acted as a compensatory tool, enabling previously struggling students to achieve scores comparable to higher-performing students.

# Key Findings

- With Writing Test 1, annotation accuracy was positively correlated with test performance across all conditions, highlighting the role of effective engagement with learning tools.
- With Writing Test 2, access to Gloss reduced the impact of early struggles, demonstrating how well-designed digital interventions can support equitable learning outcomes.
- These findings emphasize the importance of designing tools that enhance user engagement and provide adaptive support, ensuring all learners can succeed regardless of initial performance differences.

# Next Steps

- In Summer 2024, I started a study with incoming first-year law students ( $N = 116$ ) to assess how Gloss impacts their comprehension and synthesis of legal opinions.
- The goal of this study is to explore its potential for predicting academic success and identifying at-risk students.
- Early findings suggest that annotation quality, rather than quantity, is a stronger predictor of first-semester law school GPA.
  - Analyses are ongoing.
- This insight highlights Gloss's potential as a tool for law schools to proactively identify and support students who may face academic challenges.