

Plagiarism increases during the semester, and when students write more they copy more.

J. Stephen Mansfield, Department of Psychology, State University of New York College at Plattsburgh

INTRODUCTION

This study explores the incidence of plagiarism in an Introductory Psychology course.

METHODS

- Participants: 176 undergraduate students enrolled in 3 sections of *PSY101: General Psychology*.
- A written homework was assigned for each chapter of the course textbook [Shaffer & Merrens (2001) *Research Stories for Introductory Psychology*, Allyn & Bacon]. The students answered questions about the methods and findings from the research study described in the chapter.
- Students were instructed that their work must be written in their own words.
- 4835 assignments were received.

TEXT ANALYSIS

Each assignment is compared word-by-word with the textbook to create *dot-plots* in which matching sequences or two-or-more words are flagged (Figure 1).

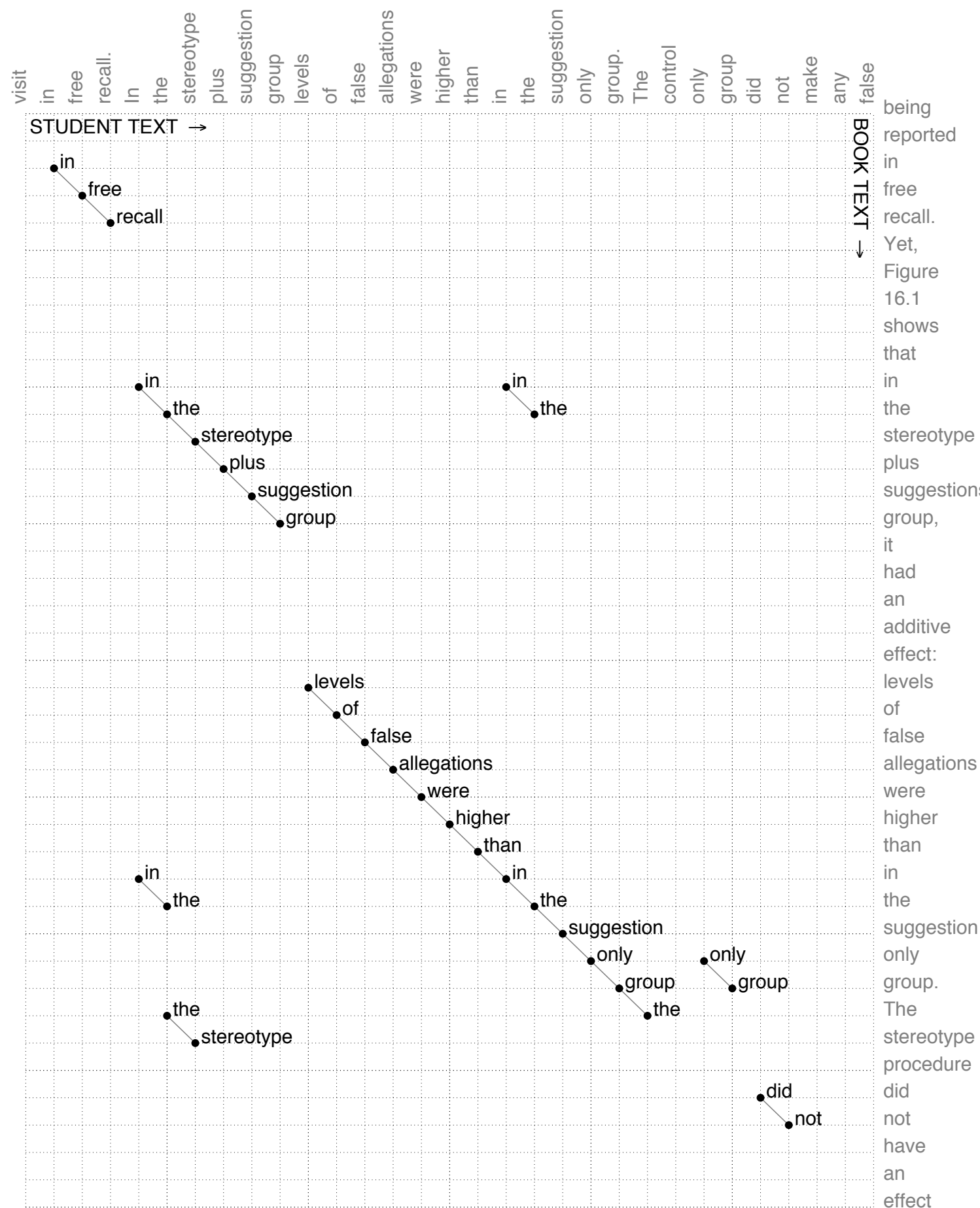


Figure 1. Part of a dot-plot showing matches between a student's work and the text book. The three matching sequences ('in free recall', 'in the stereotype plus suggestion group', and 'levels of false ... suggestion only group. The') are consecutive in the student's work and count as a single match 22-words long.

SELECTING A PLAGIARISM CRITERION

How long should the matching sequence be in order to be flagged as plagiarized?

The goal is to choose a criterion that minimizes the likelihood of falsely flagging *incidental matches*.

Incidental Matches: Similarities in wording between the students' work and the textbook that occur merely due to the syntactic constraints in the language and the vocabulary constraints imposed by writing about psychology.

The frequency of incidental matches as a function of criterion length was estimated by comparing each assignment to the text chapters that the assignment was **not** about (e.g., assignments for chapter 9 were compared to chapters 1–8 and 10–36). This is shown by the green line in Figure 2.

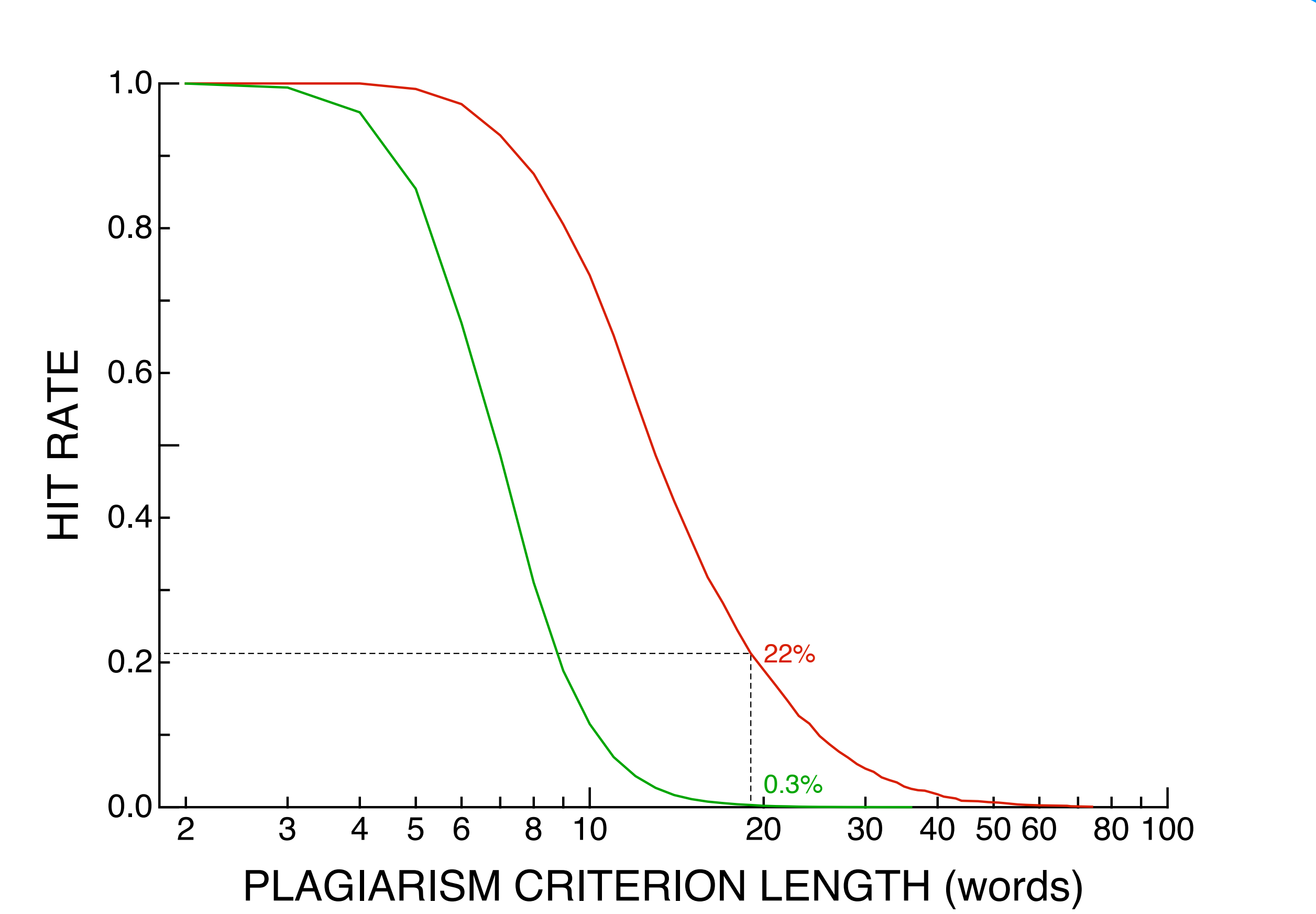


Figure 2. The proportion of assignments that contain a matching sequence that is at least as long as the criterion length, as a function of the criterion length. The green line shows data for incidental matches. The red line shows data for matches from the relevant chapter.

Incidental matches longer than 18 words are very rare (hit rate = 0.3%)

Therefore, for this study, an assignment is classified as containing plagiarism if it contains a matching sequence longer than 18 words that matches the relevant textbook chapter.

THE INCIDENCE OF PLAGIARISM

22% of assignments are classified as containing plagiarism using the 18-word criterion (see the red line in Figure 2).

LENGTH DIFFERENCES

The incidence of plagiarism is higher when students write longer answers.

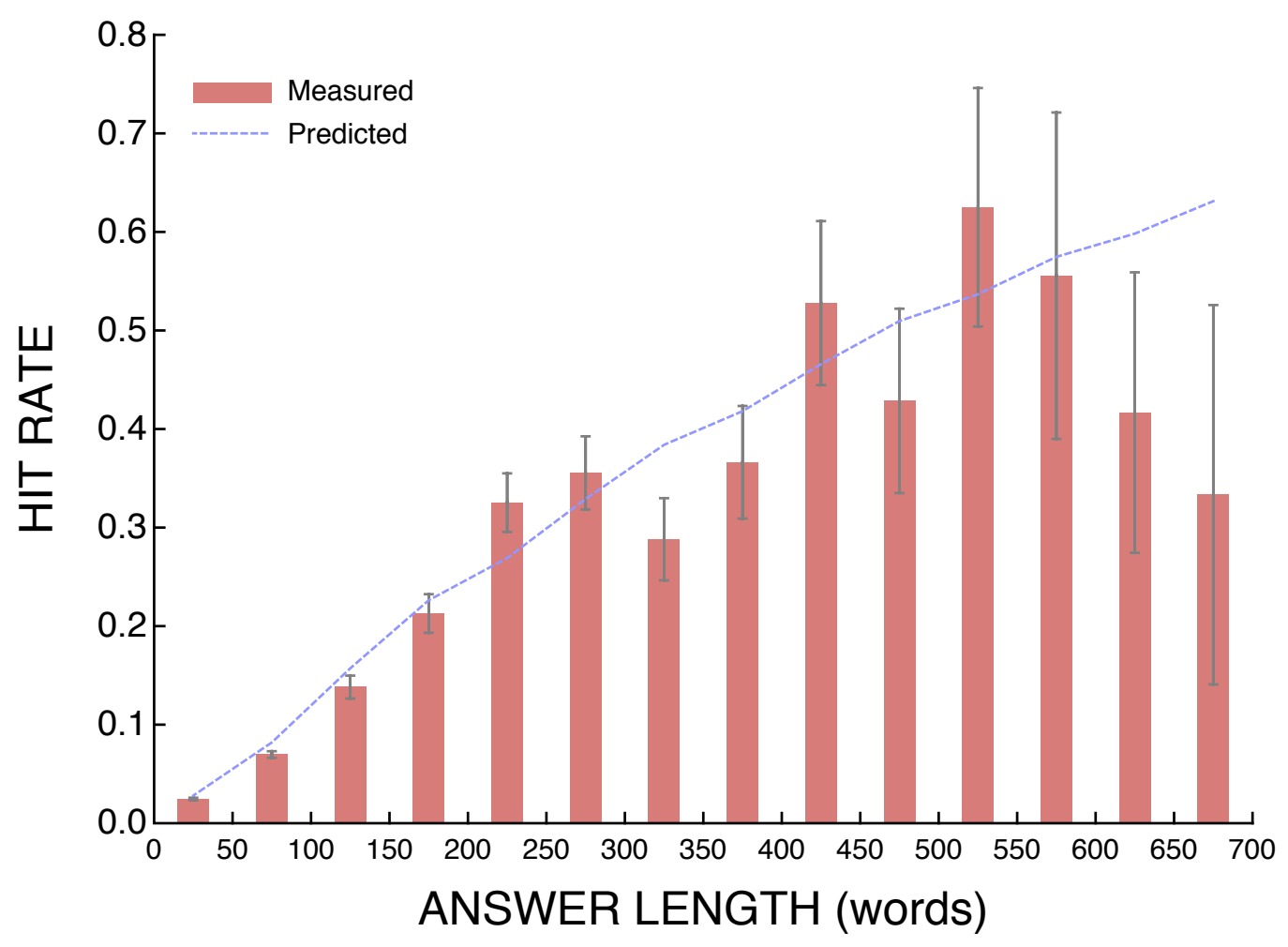


Figure 3. The proportion of answers containing plagiarism as a function of answer length. Error bars show standard deviations (the large SDs for long answers reflect the fact that the frequency of long answers was very low.) The dashed line shows the predicted incidence of plagiarism based on a simple *Poisson* model.

This relationship is explained by a simple Poisson model: If plagiarism occurs at a certain base rate then longer answers are more likely to contain plagiarism.

This finding indicates that answer length must be taken into account when we assess plagiarism.

QUESTION DIFFERENCES

Plagiarism was approximately 1.5 times more frequent when answering fact-based questions (e.g., "How did the researchers go about answering their research question?") than it was for opinion-based questions (e.g., "Describe and explain a strength or weakness of this study.")

However when the answers are compared using a length-independent measure of plagiarism, this difference is less marked.

STUDENT DIFFERENCES

Comparisons across the students revealed that only a small number of students (about 8%) frequently submitted plagiarized assignments, whereas the majority of students rarely submitted plagiarized assignments.

CHANGES DURING THE SEMESTER

The incidence of plagiarism increased during the semester from about 10% of assignments to over 30% (Figure 4a. χ^2 test for trend in proportions, $p < 0.001$).

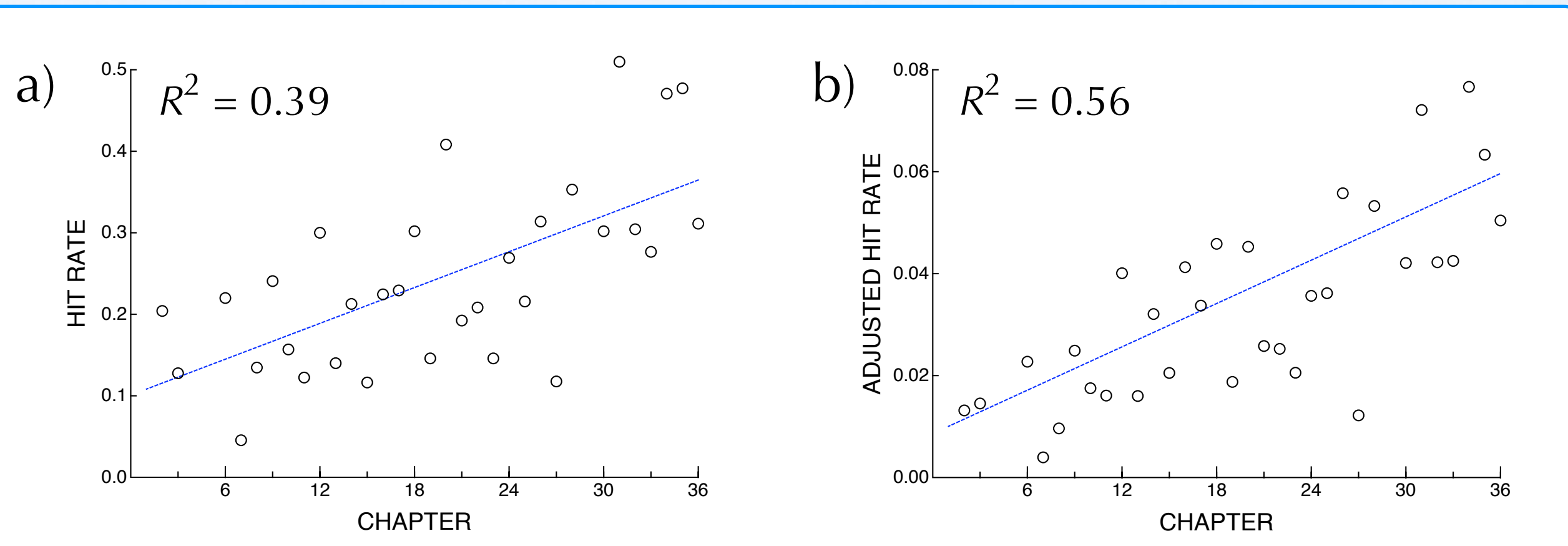


Figure 4. a) The proportion of assignments containing plagiarism as a function of the textbook chapter (chapters were read in numerical order during the 14 weeks of the semester). b) Plagiarism rate adjusted for assignment length.

A length-independent measure of plagiarism (Figure 4b) shows that the time during the semester accounts for 56% of the variance in plagiarism.

This trend was **not** found in sections of the course where the students were given feedback to prevent plagiarism.

CONCLUSIONS

This study used an objective criterion for plagiarism: a matching sequence longer than 18 words is unlikely to be an incidental match, and can be flagged as 'plagiarized.'

Using this criterion, 22% of assignments contained plagiarism. Plagiarism was more likely when students wrote long answers to fact-based questions.

This type of plagiarism may be considered to be *petty* — after all, the students are not copying whole paragraphs into their work!

However, the increase in plagiarism during the semester underlines the importance of detecting and dealing with plagiarism when it occurs.

Plattsburgh
STATE UNIVERSITY OF NEW YORK

Supported by the *Institute for Ethics in Public Life*, State University of New York College at Plattsburgh.