

# Predicting End of Chapter Grades: Exploring the Impact of Self-Index and Engagement in Online Learning

Jinyi Xue, Department of Statistics, College of Arts and Sciences, Ohio State University



## INTRODUCTION

This study investigates the effectiveness of CourseKata, an innovative e-learning platform for statistics and data science, by predicting End-of-Course (EOC) performance. The analysis is based on anonymized data from 1,625 students across 48 college-level courses at 11 U.S. institutions in 2023. Key variables include Engagement Time, On-Page and Off-Page Idle Time, and four self-evaluation measures (Intrinsic Value, Cost Value, Utility Value, and Expectancy), providing insights into student interaction with the platform and their perceived performance. By examining these factors, the study aims to identify elements influencing student success in online statistics and data science education.

## METHODS

Descriptive statistics were summarized by chapter groups using means and standard deviations for normally distributed variables (Intrinsic Value, Cost Value, Utility Value, Expectancy, and End-of-Chapter (EOC) grades); and medians and quantiles for skewed distributions (engagement time, on-page and off-page idle time). Comparisons between chapter groups were conducted using ANOVA for normally distributed variables and the Kruskal-Wallis test for non-normally distributed variables.

A Linear Mixed Model (LMM) was employed to analyze factors affecting EOC grades, accounting for repeated measures of student ID and chapter number as random effects.

The final LMM, after backward selection, after backward selection is specified as:  
**EOC ~ Total\_Level + idle\_brief + chapter\_group + (1 | student\_id) + (1 | chapter\_number).**

## RESULT

Table 1: Descriptive Statistics for Chapter Groups and Related Variables

Variables	Level	Chapter 1 to 4 n = 4728	Chapter 5 to 8 n = 4884	Chapter 9 to 13 n = 5949	Total n = 15561	P-Value
EOC	Mean (SD)	Missing = 874 0.67 (0.20)	Missing = 982 0.60 (0.19)	Missing = 2576 0.56 (0.20)	Missing = 4432 0.61 (0.20)	<0.001
	(min, max)	(0, 1.0)	(0, 1)	(0, 1)	(0, 1)	
Engagement Time (min)	Median	Missing = 865 149.77	Missing = 950 102.14	Missing = 2568 89.33	Missing = 4383 111.5	<0.001
	(p25, p75)	(83.75, 226.74)	(55.78, 162.85)	(49.23, 144.08)	(60.5, 183.9)	
On-Page Idle Time	Median	Missing = 865 43.59	Missing = 950 36.07	Missing = 2568 27.79	Missing = 4383 35.0	<0.001
	(p25, p75)	(14.07, 92.85)	(12.51, 79.16)	(8.06, 67.76)	(11.6, 79.0)	
Off-Page Idle Time	Median	Missing = 865 31.97	Missing = 950 30.6	Missing = 2568 22.78	Missing = 4383 28.3	<0.001
	(p25, p75)	(10.81, 68.66)	(10.34, 69.48)	(6.41, 51.28)	(9.5, 63.8)	
Intrinsic Value	Mean (SD)	Missing = 2245 4.56 (0.98)	Missing = 1607 4.48 (1.01)	Missing = 3079 4.54 (1.02)	Missing = 6931 4.53 (1.00)	0.009
	(min, max)	(1, 6)	(1, 6)	(1, 6)	(1, 6)	
Cost Value	Mean (SD)	Missing = 2178 4.14 (1.17)	Missing = 1509 3.97 (1.23)	Missing = 3017 4.06 (1.23)	Missing = 6704 4.05 (1.21)	<0.001
	(min, max)	(1, 6)	(1, 6)	(1, 6)	(1, 6)	
Utility Value	Mean (SD)	Missing = 1426 4.53 (0.99)	Missing = 1536 4.49 (0.98)	Missing = 3783 4.62 (0.94)	Missing = 6745 4.54 (0.97)	<0.001
	(min, max)	(1, 6)	(1, 6)	(1, 6)	(1, 6)	
Expectancy	Mean (SD)	Missing = 1382 4.28 (1.04)	Missing = 1506 4.24 (1.06)	Missing = 3765 4.32 (1.08)	Missing = 6653 4.28 (1.06)	0.015
	(min, max)	(1, 6)	(1, 6)	(1, 6)	(1, 6)	

Based on ANOVA for normally distributed numerical values and the Kruskal-Wallis test for skewed numerical values, significant differences ( $p < 0.05$ ) between chapter groups are present across all variables.

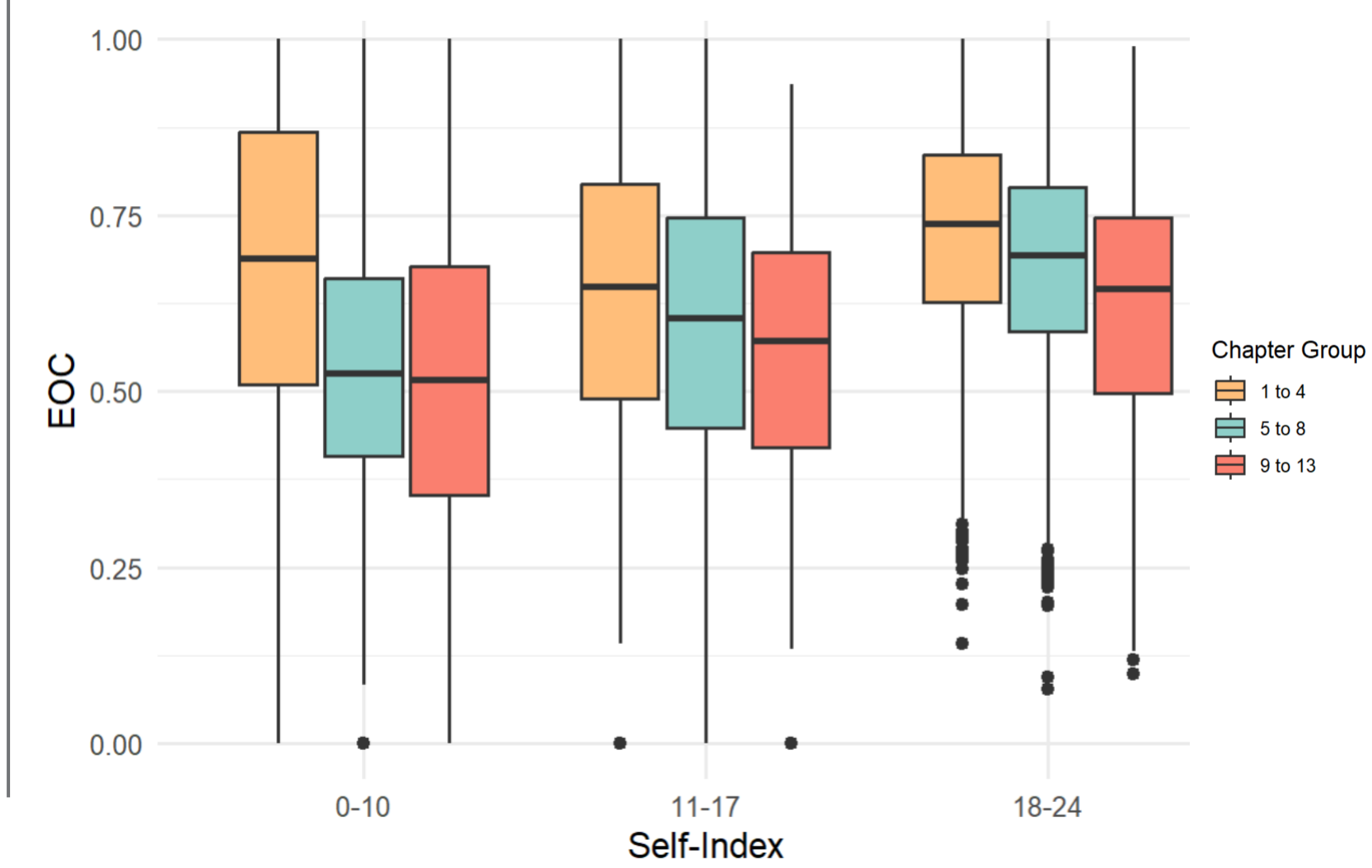


Figure 1: EOC by Self Index and Idle Time Category

This box plot displays the distribution of End of Chapter correction rates (EOC) by Self-Index categories (which based on quantile), with data segmented into three Chapter Groups. The EOC measures the average correction rate at the end of each chapter, ranging from 0 (no corrections) to 1 (all corrections made). Students with lower Self-Index tend to struggle with later chapters, while those with higher Self-Index show more consistent EOC rates across all chapters. The correction rates are generally higher for early chapters (1 to 4) regardless of the Self-Index.

Table 2: EOC With Self-Index, idle time and chapter group (Linear Mixed Model)

Predictors	EOC (End of Chapter Grades)			p-value
	Estimates	SE	95% CI (Upper, Lower)	
Fixed Effects (Intercept)	0.538	0.021	(0.500, 0.580)	< .0001
Total_Level (Self-Index)	0.007	0.0007	(0.005, 0.008)	< .0001
idle_brief (On-Page Idle Time)	-5.021*10 <sup>-5</sup>	2.129*10 <sup>-5</sup>	(-9.201*10 <sup>-5</sup> , -8.393*10 <sup>-6</sup> )	0.0184
Chapters	Ref			Type III P-value < .0001
Chapter 1 to 4	-0.03134	0.021	(-0.076, -0.013)	0.1598
Chapter 5 to 8	-0.111	0.021	(-0.155, 0.066)	< .0001
Chapter 9 to 13				

Number of Observations Used: 6997

The linear mixed model analysis of End-of-Chapter (EOC) scores reveals that student self-evaluation, measured by the Self-Index, significantly positively impacts grades ( $\beta = 0.007$ ,  $p < 0.0001$ ), indicating that students who rate their understanding more highly tend to perform better. Idle Time shows a slight negative effect on performance ( $\beta = -5.021 \times 10^{-5}$ ,  $p = 0.0184$ ). Notably, scores in later chapters (9-13) decline compared to earlier ones ( $\beta = -0.111$ ,  $p < .0001$ ).

Further analysis comparing students who completed all chapters versus those who didn't.

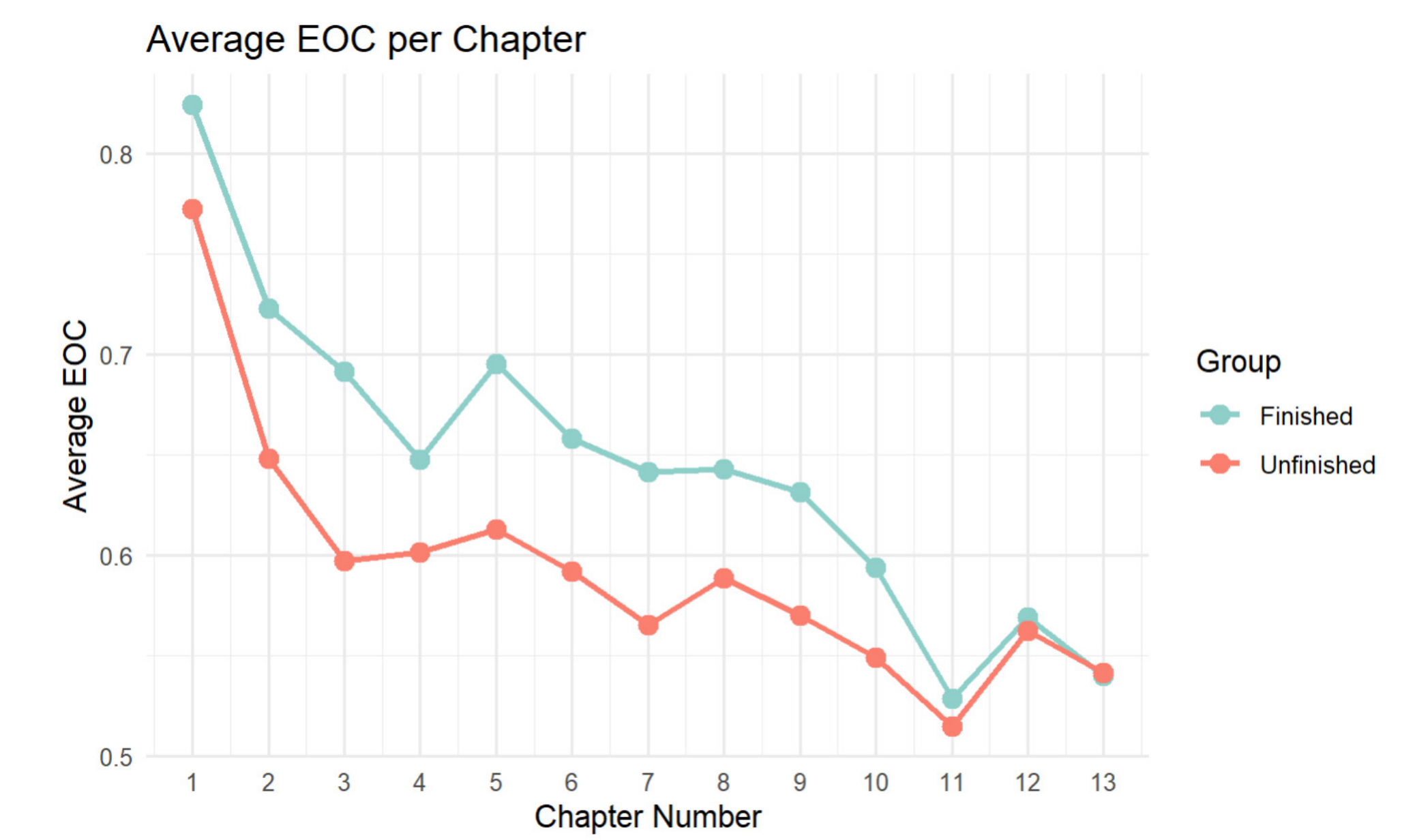


Figure 2: Average EOC per chapter by different student

The figure 2 shows that while both groups' EOC scores declined over time, students who completed all chapters consistently performed better (t-test p-value = .0805, significant at 0.1 level).

## DISCUSSION

Findings indicate that accurate self-assessment, focused study habits, and consistent engagement correlate with improved performance. Students who completed all chapters demonstrated better overall performance compared to those who did not. However, students faced increasing challenges as they progressed through the course material. These results align with the Practicing Connections Hypothesis, which posits that deliberate practice of core concepts across increasingly complex problems leads to more coherent and transferable knowledge. Future research should explore factors such as motivation and time management that may influence engagement and chapter completion. This additional inquiry could provide deeper insights into the relationship between self-assessment, course progression, and academic performance.