The application of trajectory analysis for an early warning system in STEM courses

A Dissertation Presented

by

Un Jung Lee

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Applied Mathematics and Statistics

(Statistics)

Stony Brook University

May 2015

Stony Brook University

The Graduate School

Un Jung Lee

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

Stephen J. Finch, Professor
Department of Applied Mathematics and Statistics

Song Wu, Assistant Professor
Department of Applied Mathematics and Statistics

Wei Zhu, Professor
Department of Applied Mathematics and Statistics

Ross H. Nehm, Associate Professor
Center for Science and Math Education & Dept. of Ecology & Evolution

This dissertation is accepted by the Graduate School

Charles Taber
Dean of the Graduate School

Abstract of the Dissertation

The application of trajectory analysis for an early warning system in STEM courses

by

Un Jung Lee

Doctor of Philosophy

in

Applied Mathematics and Statistics

(Statistics)

Stony Brook University

2015

The retention of STEM (science, technology, engineering, and mathematics) majors has become a national concern. "Early warning systems" (EWS) are being developed to identify students who perform poorly early in the semester so that interventions can be implemented. The research reported here utilizes clicker scores and review quiz scores collected in every class session for the longitudinal analysis, as well as pre-course concept inventory scores and self-reported student characteristics.

Pre course concept inventory scores were significantly predictive of final course grade. Student demographic characteristics had a smaller fraction of final course grade explained. The cumulative average student clicker score was highly predictive of final course grade. The cumulative average student review quiz score was also highly predictive of final course grade in spring 2014 semester, but was less predictive and less correlated with final course grade in the fall 2014 semester. The trajectories of transformed clicker and review quiz scores identified student longitudinal patterns of scores. Students with scores that were high at the beginning of the semester had consistently higher scores through the semester. In addition, the Bayesian Posterior Probabilities (BPPs) of clicker score trajectory were significant predictors of final course grade. In a trajectory analysis of ACF and PACF, the number of zero clicker scores was associated with final course grade. In conclusion, pre-course concept inventory scores and clicker scores were effective predictive variables for an EWS.

Contents

List of Tables

List of Figures

List of Abbreviations

STEM: Science, Technology, Education and Mathematics

PCAST: President’s Council of Advisors on Science and Technology

GBTM: Group-Based Trajectory Model

ACF: Autocorrelation Function

PACF: Partial Autocorrelation Function

BPP: Bayesian Posterior Probability

Acknowledgments

I dedicate this thesis to
my parents, Young Ju Suh and Jong Keun Lee, and my friends for their constant support and unconditional love.

In addition, I really thank my advisor, Stephen J. Finch for his academic support.

I love you all dearly.

Chapter 1. Introduction

1.1. STEM retention as a national problem

Many college students who initially choose science majors switch to a non-science major (Daempfle 2003; PCAST 2012; Strenta, Elliot, Adair, Matier, & Scott 1994). Some of the reasons for leaving STEM majors are loss of interest, mismatch of their talent, poor quality of instruction, or the relationship between students and the instructors in a class (Seymour and Hewitt 1997). STEM dropout has become a national concern, and many studies have suggested instructional strategies to retain STEM majors. One finding was to use Peer Instruction (PI) during an introductory science course by promoting class interaction to engage students in formative assessment (Watkins and Mazur 2013). A second approach was to develop an innovative course that was offered to students with low mathematics placement scores. This course resulted in significantly higher STEM retention rates for students completing the course compared with students having similar math placement scores who did not enroll in the course. (Koenig et al. 2012). A third approach is to use an "Early Warning System" (EWS) that identifies students struggling in STEM courses. A variety of data sources may be used in an EWS. One objective of this research is to identify variables that can be used in an EWS, especially those derived from technological advances like clickers.

1.2. Early Warning System helpful in increasing retention

Early warning systems

A course EWS is a data-base tool to identify students early in the semester who are at greater risk of dropping out or performing poorly so that interventions can be implemented. An EWS may use academic or non-academic predictors (Dobson 2008; Eddy et al. 2014; Orr and Foster 2013; Rath et al. 2007; Richardson et al. 2012; Lee et al. 2008). Academic predictors include, for example, test scores, high school GPAs, and class rank. Non-academic predictors include gender, ethnic background and English language learner status. It is possible to gather non-academic predictors, such as the student's education history and demographic information, by using a survey given before the semester starts. Similarly, concept inventory data can also be collected early. The most powerful predictors of student academic performance included course performance and attendance. However, it is difficult for many instructors to generate accurate predictors of academic performance in real time during a semester. For example, most instructors give one or two midterms in a course. This data is often collected late in the semester, so that it is not possible to give students accurate predictors of their standing early in the semester. Clicker data collected from each class session is a potentially useful academic predictor of academic performance. There may be a timepoint early in the semester when an accurate prediction of class performance is possible.

1.3. Pre-diagnostic assessments (formative assessment)

Concept Inventory testinghas been a major research agenda in STEM education (Libarkin 2008; Haudek et al. 2011). These CIs have been used to help instructors in many ways. It is possible to identify common pre-instructional alternative conceptions, naïve ideas, and faulty reasoning models. In addition, CIs have been developed to help instructors diagnose student thinking about core ideas.

However, one of limitation of using CIs as predictive tools is that they cover a smaller range of content than an undergraduate course. Another limitation is that CIs are given to students at the beginning of the semester, so that instructors cannot consider the longitudinal performance patterns of student.

Data from two concept inventories and an attitude instrument to diagnose student ability and attitude were collected in the course studied prior to instruction: the constructed-response ACORNS (Assessing Contextual Reasoning about Natural Selection, Nehm et al. 2012), the multiple-choice CINS (Conceptual Inventory of Natural Selection, Anderson et al. 2002), and the Likert-scale MATE (Measure of Acceptance of the Theory of Evolution, Rutledge and Warden 1999). Data from three semesters of the course offering were studied here.

1.4. Usefulness of routinely collected data – clicker score and review quiz (formative assessment)

Formative assessment is a diagnostic testing tool of the details of content and performance employed by teachers during the learning process. There are many types of formative assessments that provide teachers and their students with the information they need to move learning forward. In the class studied here, clicker scores and review quiz scores were additional formative assessment tools used.

Clicker score

Wireless student response systems –"clickers"- are small keypad tools that can wirelessly transmit data from each student to a receiver at the front of the classroom. A clicker is one of the most popular educational technology tools, with the device relatively being low cost and easy for students and instructors to use in large lecture classes (Bruff 2011). Computer software programs summarize students’ answers, so that instructor can easily assess students’ understanding of the content of the class and adjust their teaching. The clicker also plays a role as communicator. For example, teachers have used the clicker structure to encourage students to engage in class discussions with their neighbors before answering the clicker question. Watkins and Mazur (2013) showed that using peer instruction (PI) using clicker quiz increased the retention rate in STEM majors.

There are common problems with using clickers and recommended solutions (Douglas 2006). All technologies have some errors. For example, clickers sometimes fail to transmit their information to the computer during class. It is not easy for instructors to identify who has a technical problem among the large number of students in the class. However, the failure rate for most clickers has decreased over time. A second problem is that clickers can be used dishonestly. For example, one student could simultaneously respond to clicker questions using his friend’s clicker as well as his own. One of solutions is that TAs monitor the class during clicker response entry. The other strategy is to keep the clicker on the student’s desk (Jones et al. 2009). Even though there are problems with clickers, researchers have demonstrated the effectiveness of clicker systems for student learning (see Caldwell 2007).

Review quiz

A review quiz is a type of formative assessment that can evaluate students’ academic performance. Review is a significant process in improving the extent of learning. McDaniel and Howard (2009) reported on an experiment with college students to compare the effectiveness of a study strategy that included the review process with various other methods. They found that the study strategy including the review process had more effectiveness in learning than rereading and note-taking. The review quiz process is an effective tool to improve student learning and examination performance. That is, the review process provides students with the opportunity to study the contents covered in the class for 24 hours before the quiz was given (McDaniel, et al. (2011)). Therefore, by giving a review quiz after every session, students can improve their academic performance and provide an indicator of class performance to the instructor.

1.5. Group-based trajectory modeling (GBTM)

GBTM was developed by Nagin (1990, 2005) is an application of finite mixture modeling. GBTM are designed to identify distinguishable groups of individuals with similar progressions of outcomes within the sample over time. Trajectory analysis has been widely used in the study of longitudinal patterns of substance abuse (e.g., Brook et al. 2014), criminology (e.g., Nagin and Land 1993) and clinical research (e.g., Nagin and Odgers 2010) but has only rarely been applied to educational research. Recently, trajectory analysis has been used to extract longitudinal patterns of subject matter mastery among students in the Chicago public school system (Torre et al. 2013).

The software program, PROC TRAJ in SAS, is used here for GBTM. The specific form of GBTM is determined by the type of the data being analyzed. PROC TRAJ provides options for three different distributions: the zero-inflated Poisson (ZIP) model for counts, the censored normal model (CNORM) for continuous data, and the logistic model (LOGIT) for dichotomous data. Further, a polynomial is used to describe each trajectory. The polynomial order in PROC TRAJ ranges from constant to fourth degree. One of the keys in GBTM is to determine the number of groups that best fit the data. In PROC TRAJ, researchers set the number of trajectory groups and the highest polynomial order that best describes the path which each trajectory group takes over time. Then, different models with a variety of numbers of trajectory groups and shapes can be compared to find the optimal model that best describes the data being analyzed. The Bayesian Information Criterion (BIC) is used to determine the best model in a sample. That is, the model with the highest BIC is selected as the best model. Researchers, however, often add the requirement that there be a minimum number of participants in each group.

The Bayesian posterior probability (BPP) that each student belongs to each trajectory group is estimated (based upon estimates of the model parameters for each trajectory). Some researchers assign each student to the trajectory group with the largest BPP (this is known as "modal assignment"). Others use the BPPs directly as predictors in a regression model.

Group-based dual trajectory model

The group-based dual trajectory model, as an extension of the standard group-based model, was designed to analyze the developmental course of two distinct but related outcomes (Nagin and Tremblay 2001). The advantage of the group-based dual trajectory model is modeling the simultaneous interrelationship between two longitudinal outcomes. In this study, the data collected in spring 2014 had two longitudinal variables, clicker scores implemented for every session held and review quiz scores after every scheduled session. I hypothesize that the dual trajectory pattern will be more effective in the analysis of student academic performance than either the GBTM of clicker scores or the GBTM of review quizzes alone. This dual trajectory modeling produces joint BPPs.

Collections of time series (ACF, PACF)

Data collected consecutively in time often have important serial correlations (Granger 1981). These are summarized by the autocorrelation function (ACF) and partial autocorrelation function (PCAF). These functions are a set of correlation coefficients between past values with specified lag periods. Autocorrelation has been widely used in many engineering and basic science fields. It is not common to apply the autocorrelation function with educational data. Each autocorrelation coefficient varies between -1 and +1.

1.6. Research questions

I address the flowing research questions in chapter 3, 4, and 5:

In chapter 3, I consider clicker scores and predictors obtained early in the course from the concept inventories and student questionnaires.

(RQ1) To what extent does each data source (i.e. clicker scores or concept inventory scores) predict final course performance?

(RQ2) When in the semester can accurate final course performance predictions be made using clicker scores and other data?

(RQ3) How many distinct trajectory patterns characterize students' clicker performances?

(RQ4) To what extent do academic--including trajectory results—and non-academic variables predict final course performance?

In chapter 4, I add review quizzes to the other variables. My questions are:

(RQ1) To what extent do review quiz scores predict final course performance?

(RQ2) How many distinct trajectory patterns characterize students' review quiz scores?

(RQ3) How many distinct bivariate trajectory patterns characterize student’ performance on clicker and review quizzes?

(RQ4) To what extent does adding review quiz scores--including trajectory and bivariate trajectory results—improve the prediction of final course performance?

In chapter 5, I assess the stochastic properties of the sequences of clicker scores.

(RQ1) Does a clicker score in one session have associations with clicker scores in other sessions? That is, is each student’s sequence of clicker scores “white noise”? Similarly, for the sequence of review quizzes?

(RQ2) How many distinct trajectory patterns characterize collection of student ACF and PACF functions?

(RQ3) How many distinct bivariate trajectory patterns of ACF and PACF are there?

I present my methods in chapter 2. I summarize my findings and directions for future research in chapter 6.

Chapter 2. Methods

This chapter introduces the outcome measure (final course grade), clicker scores, review quiz scores, pre-diagnostic assessments, and student characteristics used in this research. Furthermore, I transformed the raw clicker and quiz scores to make their distribution closer to the normal distribution assuming the statistical analyses methods.

Course grades

The outcome measure in this research is final course grade. Data from three semesters (fall 2013, spring 2014, and fall 2014) were collected from BIO 201 (Fundamentals of Biology: Organisms to Ecosystems). This course was an introduction to the major groups of living organisms. Structure, functions, the ecological roles of organisms in communities and ecosystems, and their evolutionary history were covered. Informed consent to take the concept inventories and questionnaires was required. The participation rate was 64% in the fall 2013, 80% in the spring 2014 and 66% in the fall 2014.

The final grade was calculated as follows: (1) 75% for examinations (two midterms and a final), 5% for homework, and 20% for clicker scores in fall 2013 (2) 80% for exams (three midterms and a final), 5% for review quizzes, and 15% for clicker scores in spring 2014 (3) 80% for exams (three midterms and a final), 16% for homework (that is, review quiz scores accounted for 20% of homework), and 12% for clicker scores. Students were not required to answer all of the clicker questions correctly to receive full credit for their clicker grade. This policy resulted in three quarters of the class earning 85% or higher of this portion of their grade. That is, clicker scores had a minimal impact on differentiating student’s final letter grades. In order to perform quantitative analyses, letter grades were transformed to numerical values: A to 4.0, A- to 3.67,

Table 2.1. The distribution of the final course grade of each semester and participation rate

Semester / Final grade	Fall 2013	Spring 2014	Fall 2014
Participant rate	64%	80%	66%
A/A-	20% (n=58)	44.5% (n=172)	12% (n=37)
B+/B/B-	39% (n=112)	27.2% (n=105)	26% (n=80)
C+/C/C-	37% (n=106)	20.5% (n=79)	35% (n=106)
D+/D/D-	4% (n=11)	7.3% (n=28)	24% (n=74)
F	0% (n=0)	0. (n=2)	3% (n=8)
Total	100% (n=287)	100% (n=386)	100% (n=305)

B+ to 3.33, B to 3.00, B- to 2.67, C+ to 2.33, C to 2.00, C- to 1.67, D+ to 1.33, D to 1.00, and F to 0.00.

Clicker scores

Students purchased hand-held clicker devices from Turning Technologies as part of the requirements of the course. These devices were used to capture student responses to 1-4 multiple choice clicker questions in each of the 37 class sessions in the fall 2013 semester, the 34 class sessions in the spring 2014 semester, and the 36 class sessions in the fall 2014 semester. Most clicker questions were designed to have one correct answer, but some questions had multiple correct answers. One point was given for each correct answer, and a minimal score above zero (i.e. 0.01) was given for incorrect answers (to encourage participation and distinguish class absence). A score of zero was given when no data were received by the Turning Point software.

This could be the result of many factors, such as a student's absence, misplacement of the clicker, or malfunction of the clicker. Clicker malfunctions occurred in only a small fraction of responses (< 5%).

Review quiz

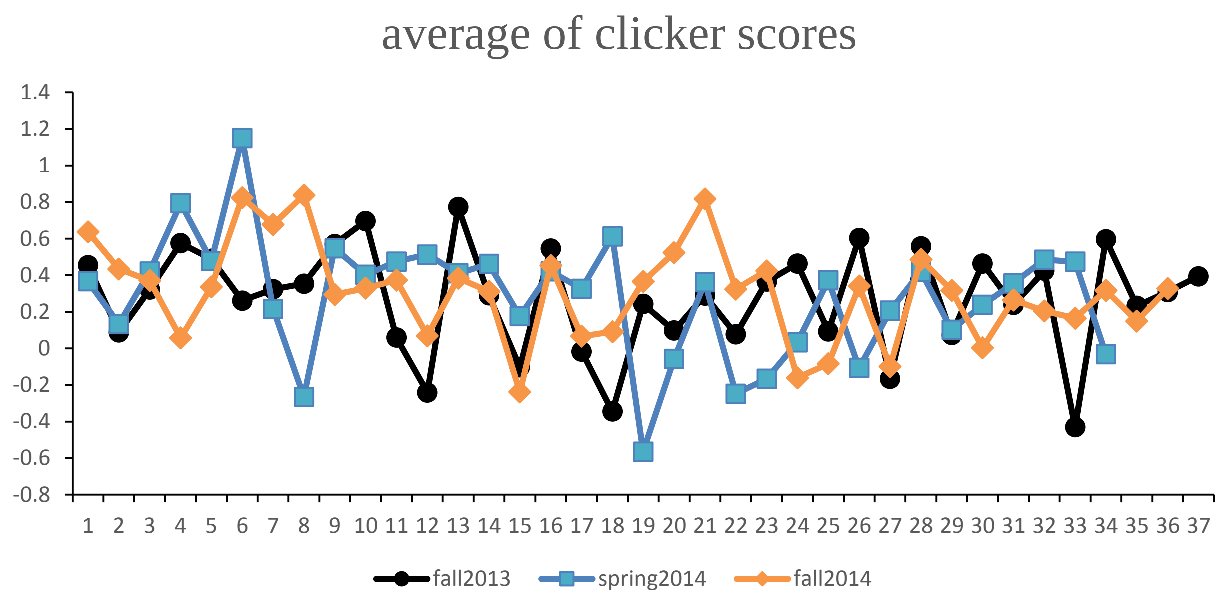
A review quiz was given to the class after each of the 39-scheduled class sessions in the spring 2014 semester and the 37-scheduled class sessions in the fall 2014 semester. These quizzes used multiple-choice questions. For example, one question for lecture 10 in the spring 2014 semester- what are proteins? A. They are sequences of nucleotides B. They are genes C. They are sequences of amino acids D. They are often enzymes.

In the spring 2014 semester, some of clicker scores did not exist because of the cancellation of classes due to weather conditions. Consequently, I dropped 5 review quiz scores for the cancelled classes in the spring 2014 semester in the bivariate trajectory analysis.

Transformation of clicker and review quiz scores

I normalized the daily clicker and review quiz scores so that each session contributed equally to the cumulative clicker score and review quiz scores. I transformed clicker and review quiz scores to make their distribution closer to the normal distribution assumed in a trajectory analysis. Specifically, the clicker score of student *i* at session *t*, clicker_{it}, was transformed to $s_{it} = \Phi^{-1}(\frac{clicker_{it}+0.5}{max\ (clicker_t)+1})$, where $max\ (clicker_t)$ was the highest score possible for the clicker on session *t* (fall 2013: *t* = 1, 2, 3, …37, spring 2014: *t* = 1, 2, 3, …34, and fall 2014: t= 1,2,3, …36) and Φ^{-1} is the inverse of the cumulative distribution function of the standard normal random variable (Draper and Smith 1998). For $review_{it}$, the review quiz score for student *i* on quiz t, $r_{it} = \Phi^{-1}(\frac{review_{it}+0.5}{max\,(review_t)+1})$ was the transformation of the review quiz score of student i at session t, where $max\,(review_t)$ was the highest score possible for the review quiz at session t.

Figures 2.1. Average of the transformed clicker scores and review quiz scores



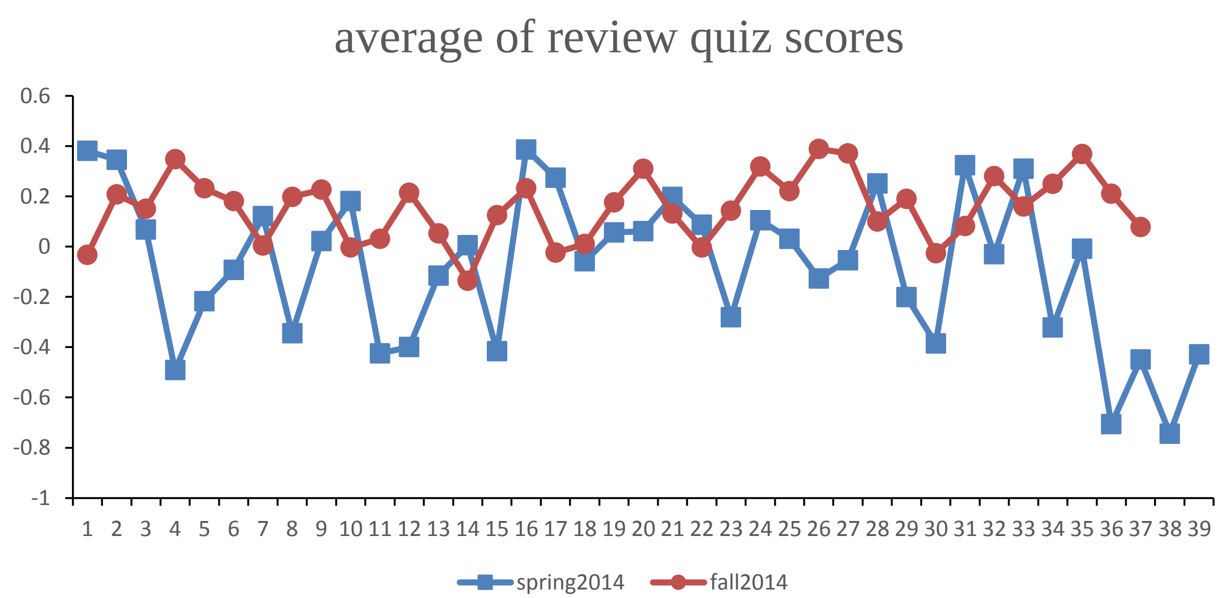


Figure 2.1 displays the patterns of the average of the transformed clicker and review quiz scores at each session. For each semester available, each of the three semesters had a similar patterns of the average of the transformed clicker scores. The average of the transformed review quiz scores in the spring 2014 decreased toward the end of the semester.

Pre-diagnostic assessments

In the fall 2013 semester, there were six pre-course diagnostic assessments for each student: two ACORNS KC (key concept) scores, two ACORNS NI (naïve idea) scores, one CINS score, and one MATE score. In the spring 2014 semester and the fall 2014 semester, four parts of the pre-course diagnostic assessments were given to students: CINS, ISEA, Biology, and Human Life. The pre-course diagnostic assessments excluding CINS had 3 missing, respectively in the spring 2014 semester.

Table 2.2 displays the descriptive statistics of diagnostic assessments in each semester. Each semester has missing values from 11 students in the fall 2013 and 3 students in the spring 2014.

Table 2.2. Summary of diagnostic assessments

		Mean (SD)	Max	Min	Number of missing values
Diagnostic test of Fall 2013 (n=287)	Snail	1.48 (1.23)	5	0	11
	Mis-snail	0.48 (0.74)	3	0	11
	Rose	0.99 (0.96)	5	0	11
	Mis-rose	0.40 (0.61)	2	0	11

	CINS	11.45 (4.18)	20	2	11
	MATE	79 (12.07)	100	45	11
Diagnostic test of Spring 2014 (n=386)	CINS	12.17(4.13)	20	0	0
	ISEA	50.42 (6.73)	60	21	3
	Biology	19.43 (3.65)	25	6	3
	Human Life	20.09 (3.11)	25	9	3
Diagnostic test of Fall 2014 (n=305)	CINS	11.4 (4.2)	20	3	0
	ISEA	50.31 (7.21)	60	28	0
	Biology	19.59 (3.56)	25	7	0
	Human Life	20.21 (3.22)	25	11	0

Note: fall 2013 missing values from 11 students and spring 2014 missing value from three students

Missing data

A small number of pre-course diagnostic assessments values were missing (~ 5%). Multiple imputation (employing SAS PROC MI) was used to estimate missing values. This approach is widely used in statistical research (Peugh et al. 2004). I used a Markov chain Monte Carlo method to impute values for data set with an arbitrary missing pattern, assuming a multivariate normal distribution for the data. The number of imputations was set 20. I used the average of the imputed values in the regression analysis.

Student characteristics

Table 2.3 contains the summary statistics of student characteristics for each semester. Some participants did not respond to some (or all questions) in the self-survey. In the fall 2013, gender had 23 missing values, ethnicity had 23 missing values, English reading ability had 25 missing and English writing ability had 27 missing of 287 participants. In the spring 2014, English reading ability had 1 missing, and English writing ability had 6 missing of 386 participants.

There were 23 students who did not respond to any question in the self-survey in the fall 2013 semester; 2 students who did not answer only reading ability, and 4 students who did not respond to the writing questions. In the spring 2014 semester, 1 student did not answer the reading ability questions, and 6 students did not answer the writing ability question. In the fall 2014 semester, 11 students did not report their demographic information.

Table 2.3. Summary of students' characteristics

Lists \ Semester		Fall 2013	Spring 2014	Fall 2014
The number of participants		287	386	305
Gender	Male	46.2%	51.0%	49.0%
	Female	53.8%	49.0%	47.0%
Ethnicity	Non-Hispanic Whites	39.8%	44.0%	38.0%
	Asian	37.5%	36.0%	41.0%
	Minority	22.7%	20.0%	21.0%
Age		19.9 years (SD = 2.1, Max = 32, Min = 17)	19.7 years (SD = 1.9, Max = 41, Min = 18)	19.2 years (SD = 4.2, Max = 30, Min = 17)

English native speaker	73.1% out of 256 answering	77.7% out of 381 answering	70.5% out of 305 answering
English reading ability	5.1 (High ability = 6; SD = 1.0)	5.3 (High ability = 6; SD = 0.9)	4.7 (high = 6, sd = 1.6)
English writing ability	4.9 (High ability = 6; SD = 1.0)	5.1 (High ability = 6; SD = 0.9)	4.5 (high = 6, sd = 1.6)

Correlation

Let $\overline{S_{\iota T}} = \frac{\sum_{t=1}^{T} s_{it}}{T}$ be the average of the transformed clicker scores of student i from the start of the semester up to session T. Let $\overline{R_{\iota T}} = \frac{\sum_{t=1}^{T} r_{it}}{T}$ be the average of the transformed review quiz scores of student i for sessions up to session T. These cumulative averages were calculated at the end of each week (i.e., every three class sessions) of the semester. I calculated the correlation of F_i (the final grade) with $\overline{S_{\iota T}}$, the correlation of F_i with $\overline{R_{\iota T}}$, and the multiple correlation of F_i with $\overline{S_{\iota T}}$ and $\overline{R_{\iota T}}$. I then identified the point in the semester when this correlation was high enough to be actionable.

Regression analysis

Multiple regressions analyses were run to calculate the contribution to the $R^2$of the final grade from student characteristics and pre-course diagnostic measurements. Specifically, I used the final grade for student i, F_i, as the dependent variable and the pre-course diagnostic assessments and student characteristics as independent variables. I used PROC REG (in SAS) to analyze the predictive ability of the pre-course diagnostic assessments and used PROC GLM (in SAS) to incorporate self-reported ethnicity, gender, English reading skill, English writing skill, and native English speaker status.

I performed a multiple regression analysis of final course grade that combined all data sources (i.e., student characteristics, pre-course diagnostic assessments, and BPPs of trajectory groups) in order to compare the variation of the final grade explained by each of the predictor.

Prediction rate

I determined the value of the diagnostic tests, the $\overline{S_{\iota T}}$ (the average transformed clicker scores) and the $\overline{R_{\iota T}}$ (the average transformed review quiz scores) for predicting whether a student's final grade would be in the top or bottom half of the grade distribution. I performed a median split on the independent and dependent variables. The number of successful predictions of student final grades was calculated as the sum of the number of students who were below the population median for both the independent variable and the final grade plus the number of students who were above the median on both measures. I used SAS PROC FREQ to calculate the fraction of successful predictions, which was the ratio of the number of successful predictions to the total number of students. I plotted this ratio for each diagnostic measure for each week.

Univariate trajectory analysis

GBTM was performed to estimate the number of clicker score trajectory patterns in the sample as well as the patterns themselves. I used the PROC TRAJ plug-in for SAS developed by Jones and Nagin (2007) to calculate these results. I used the normal distribution specification (called CNORM) and the BIC to select the number of groups *G*. Raftery (1995) and Kass et al. (1995) recommended selecting the model with the maximum BIC. I estimated models with the number of trajectory groups *G* ranging from 1 to 7.

I compared the mean final course grades among the *G* trajectory groups, using ANOVA with modal assignment of students to trajectory group. I also used the regression approach described above to explore the relationship between the BPPs of the trajectory groups and final course grades. I did not use the BPP of the most prevalent trajectory group as an independent variable in this regression analysis to avoid collinearity.

I used three periods for the trajectory analysis: (1) clicker scores from all 37 sessions for fall 2013, 34 sessions for spring 2014 and 36 session for fall 2014; (2) clicker scores from the first three weeks (t = 1… 9); and (3) clicker scores from the first eight weeks. These analyses were used to determine whether trajectory analysis of the first three weeks of clicker scores was predictive of student final grades and whether trajectory analysis of the first three eight weeks of clicker scores was predictive of student final grades. The time just before eight weeks was chosen because the university deadline of withdrawing from the course with record is about nine weeks.

Bivariate trajectory analysis

Bivariate-trajectory groups were defined from two trajectories for $Y_t = (s_{it}, r_{it})$. One was for the review scores and the other was for the clicker quizzes. The basic form (Jones and Nagin (2007)) of the likelihood is

$$P(Y_1, Y_2, \dots, Y_k) = \sum_j \pi_j \prod_k f_k^j(Y_k),$$

where k is an index of the number of different outcome trajectories in each trajectory group j. $f(*)$ is the distribution for each such outcome by trajectory group, which can be different across the outcomes. As in the basic trajectory model, the probabilities of group membership determine the size of groups and are related to academic performance. In addition, posterior probabilities of group membership are computed, which assigns students to trajectory groups.

I used PROC TRAJ to compute posterior probabilities of group membership and related group membership assignments. In addition, the plot revealed the estimated number of clicker scores and review quiz score trajectory patterns in the sample as well as the patterns themselves. I used the normal distribution specification (called CNORM) and the BIC to select the number of groups G. I estimated models with the number of trajectory groups G ranging from 1 to 7.

Autocorrelation and partial autocorrelation applied into trajectory analysis

One of assumptions of the traditional regression analysis is that errors are independent. The ACF and PACF functions would have constant 0 value. I calculated ACF and PACF of clicker scores using software from the R language in chapter 5. The formula of autocorrelation function is $ACF_{i,\tau} = \frac{E[(s_{i,t}-\mu_t)(s_{i,t+\tau}-\mu_s)]}{\sigma_t\sigma_{t+\tau}}$, which is the value of lag τ ACF of student *i* and partial autocorrelation function is $PACF_{i,\tau} = \frac{Cov(s_{i,t}, s_{i,t-\tau}|s_{i,t-1} \dots s_{i,t-(\tau-1)})}{\sqrt{var(s_{i,t}|s_{i,t-1} \dots s_{i,t-(\tau-1)})var(s_{i,t-\tau}|s_{i,t-1} \dots s_{i,t-(\tau-1)})}}$, which is the value of lag τ PACF of student *i*. Here $s_{i,t}$ is the transformed clicker score of student *i* at session *t*, μ_t is the mean at session *t*, σ_t is the standard deviation at session *t*, and τ is the time lag (fall 2013: τ=1, 2, 3…36, spring 2014: τ= 1, 2, 3 … 33). Then, I applied those values into the univariate trajectory analysis and bivariate trajectories analysis upon these sequences.

Chapter 3. Clicker Score Trajectories and Concept Inventory Scores as Predictors

The purpose of this chapter was to identify the relation between session clicker scores and course final grade for the three semesters studied. I found that a student's performance on clicker questions was highly related to the student's final grade. I then sought to identify the time point of effective prediction in order to help develop an early warning system. I performed prediction and correlation analyses with the cumulative clicker scores. I then examined the usefulness of GBTM. I sought the optimal number of trajectory groups and trajectory patterns in clicker scores. Univariate trajectory analysis was used to show how many distinct trajectory groups were in each class and how the clicker scores of each trajectory group changed on average. In addition, fractions of variation explained were reported in regression analysis in order to assess the explanatory values of pre-course diagnostics and demographic variables for the final course grade.

Results

The median final grades were 3.0 (B) in fall 2013, 3.33 (B+) in spring 2014, and 2.33 (C+) in the fall 2014. As shown in figure 3.1, the prediction rate of cumulative transformed clicker scores by week for each semester gradually increased as the semester progressed. Table 3.1 displayed the prediction rate of cumulative transformed clicker scores at three time points.

Table 3.1 Prediction rate of cumulative transformed clicker scores

	Fall 2013	Spring 2014	Fall 2014
3 week	68.0%	68.0%	66.0%
8 week	71.0%	77.0%	68.0%
Compete semesters	71.0%	79.0%	70.5%

Figure 3.1. Fraction of correct predictions by week prediction made by clicker scores

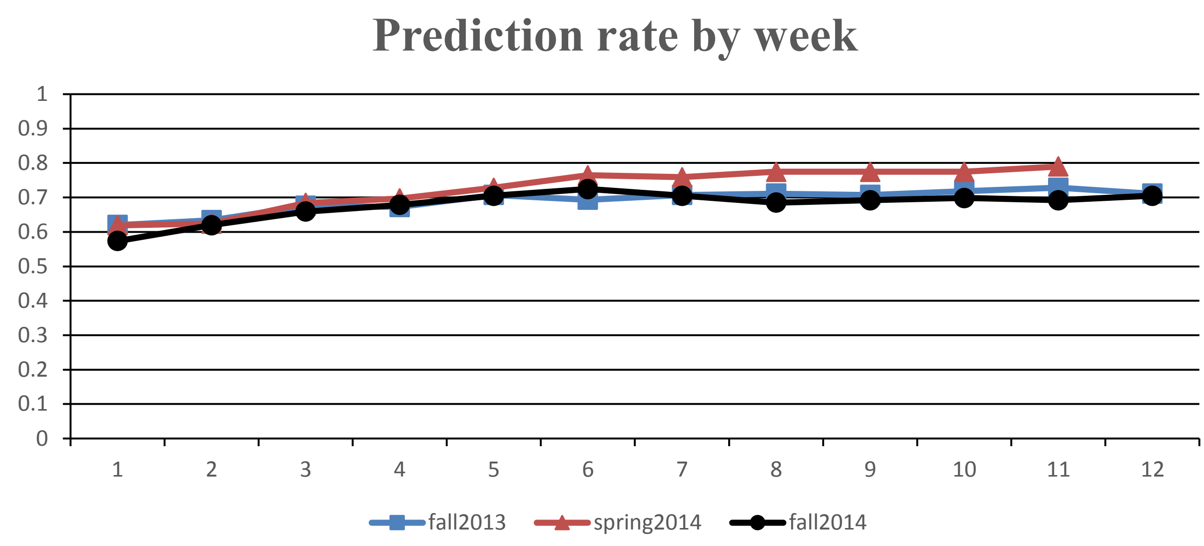
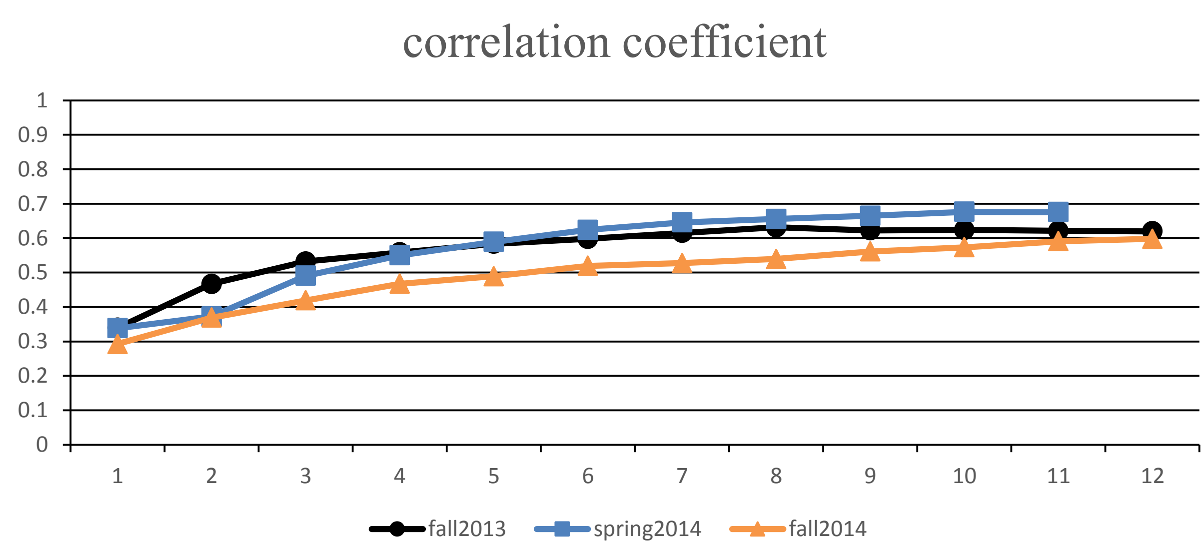


Figure 3.2. Correlation analysis plot



Correlation of cumulative average of transformed clicker scores by week with the final grade

The correlation coefficients of the cumulative average of transformed clicker scores with final grade were calculated by week and shown in figure 3.2 for each semester. The coefficients generally increased as the semester progressed. The correlation coefficient was 0.5 or larger for any time past the third week of the semester in the fall 2013 and spring 2014. In the fall 2014, the correlation coefficient at 3 weeks was 0.42, and it was over 0.5 at 5 weeks. The correlation coefficient of 8 weeks was 0.63 in the fall 2013, 0.65 in the spring 2014, and 0.54 in the fall 2014. They were 0.61 and 0.68 at the end of the two semesters respectively.

Table 3.2. Correlation coefficient of clicker scores at time points

Semester / Timepoint	Fall 2013	Spring 2014	Fall 2014
Week (3)	0.53	0.50	0.42
Week (8)	0.63	0.65	0.54
Compete semester	0.61	0.68	0.60

Results of Trajectory analysis for clicker scores of three semesters

Table 3.3 shows the BIC scores when the number of trajectory groups ranged from 1 to 7 for each semester. When the smallest absolute value of BIC (i.e., largest BIC score) was used for model selection, the fall 2013 and spring 2014 semesters appeared to have three trajectory groups and the fall 2014 semester appeared to have four trajectory groups.

Table 3.3. BIC score of trajectory groups from 1 to 7 of each complete semester

Group \ Semester	Fall 2013	Spring 2014	Fall 2014
1	-12143.55	-14630.62	-12828.51
2	-11673.19	-13926.36	-12036.30
3	-11585.05	-13856.94	-11901.58
4	-11608.23	-13880.65	-11889.81
5	-11596.66	-13869.56	-11873.18
6	-11601.66	-13893.26	-11896.44
7	-11603.32	-13908.20	-11908.93

Figure 3.3 showed the trajectory patterns for each semester. The same patterns held for the fall 2013 and spring 2014 semesters. There were three trajectory groups: (1) the high group was almost constant or slightly increasing over time, (2) the intermediate group was constant, and (3) the low group was decreasing. While the difference between initial trajectory points was not large, the differences increased as time passed. There were four trajectory groups in the fall 2014 semester: (1) the high group was consistently higher than the rest of trajectory groups, (2) high-intermediate group started at the same point as high group, but decreased to the average value of the low-intermediate group at the last session, (3) the low-intermediate group started at the lowest score, but became equal to the low-intermediate, and (4) the low group was decreasing over the whole semester.

Table 3.4. Distribution of students in clicker trajectory group of complete semester

<table>
<tr><th>Semester / Trajectory group</th><th>Fall 2013</th><th>Spring 2014</th><th>Fall 2014</th></tr>
<tr><td>High (G1)</td><td>64.5%</td><td>66.5%</td><td>52.2%</td></tr>
<tr><td>High-Intermediate (G2)</td><td rowspan="2">28.7%</td><td rowspan="2">23.9%</td><td>28.1%</td></tr>
<tr><td>Low-intermediate (G3)</td><td>9.8%</td></tr>
<tr><td>Low (G4)</td><td>6.8%</td><td>9.7%</td><td>9.9%</td></tr>
</table>

Table 3.4 displayed the distribution of students in the clicker trajectory group of the complete semester. It was possible to identify how many students in the course were assigned to each trajectory group. The highest group had over 50% of the class and the lowest group had 6.8%~9.9% of the class. I then examined whether there were distinct trajectory groups apparent in the first 3 weeks and first 8 weeks of the semester in order to identify the fraction of variance of the final course grade explained by BPPs for each time point trajectory. The prevalence of the highest trajectory group at 3 weeks was over at least 73.0% and at 8 weeks was over at least 59.8%, which was the largest proportion. On the other hand, the prevalence of the lowest trajectory group was similar proportion to the prevalence of the lowest whole semester trajectory group. Table 3.5 and table 3.6 showed the distribution of students in clicker trajectory group using 3 weeks and 8 weeks of clicker data.

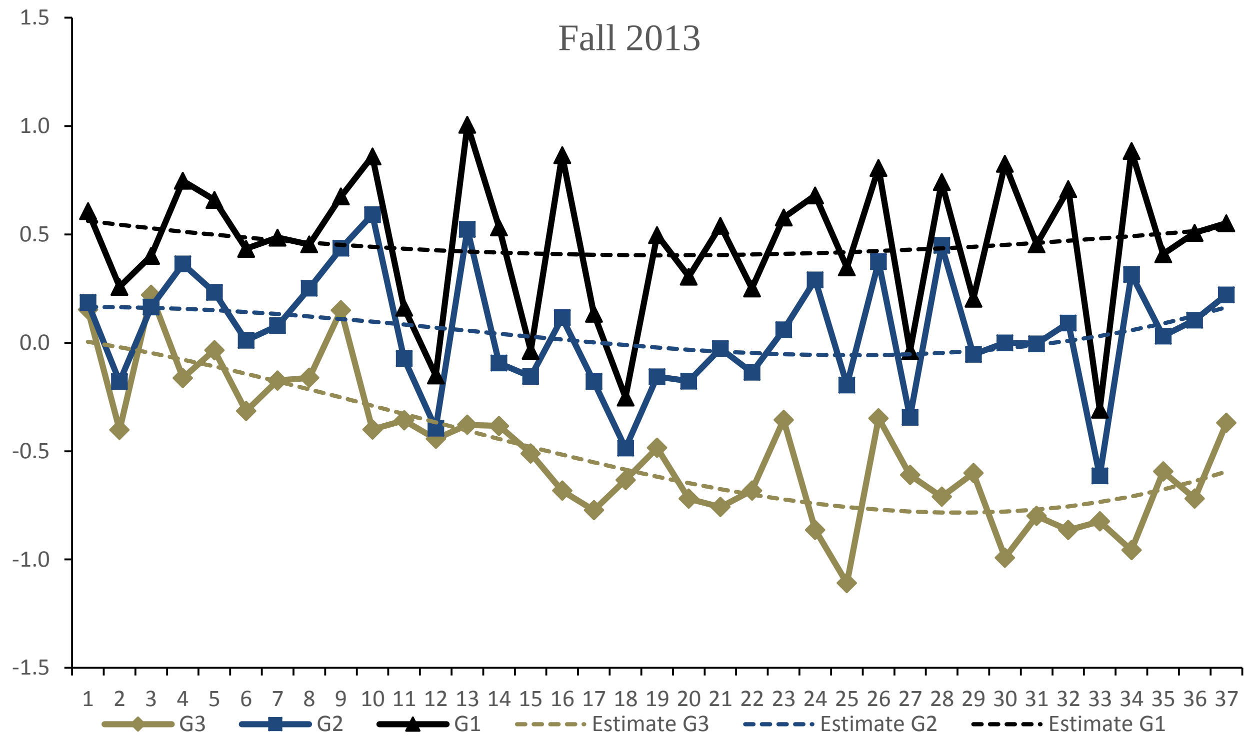
Table 3.5. Distribution of students in clicker trajectory group of 3 weeks

	Fall 2013	Spring 2014	Fall 2014
High	73.0%	95.2%	86.2%
Intermediate	14.5%		6.4%
Low	12.5%	4.8%	7.4%

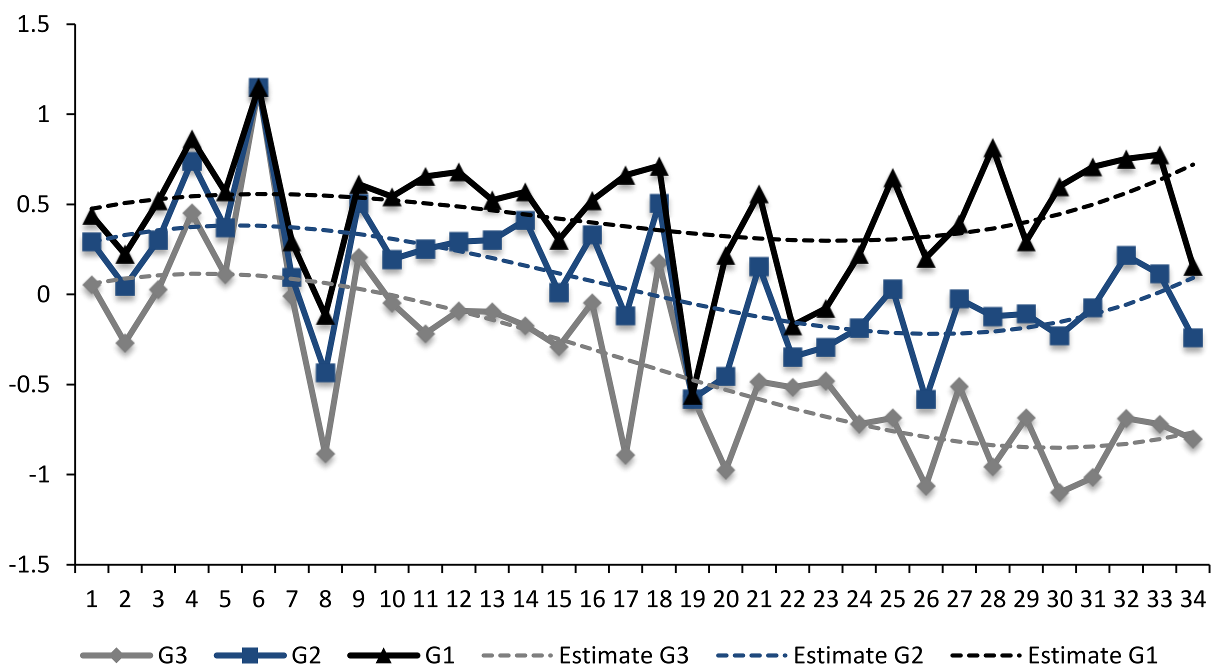
Table 3.6. Distribution of students in clicker trajectory group of 8 weeks

	Fall 2013	Spring 2014	Fall 2014
High	59.8%	86.9%	64.4%
Intermediate	21.6%		26.7%
	13.1%		
Low	5.5%	13.1%	8.9%

Figure 3.3. Estimated trajectory patterns - (CNORM model: Censored Normal Distribution)



Spring 2014



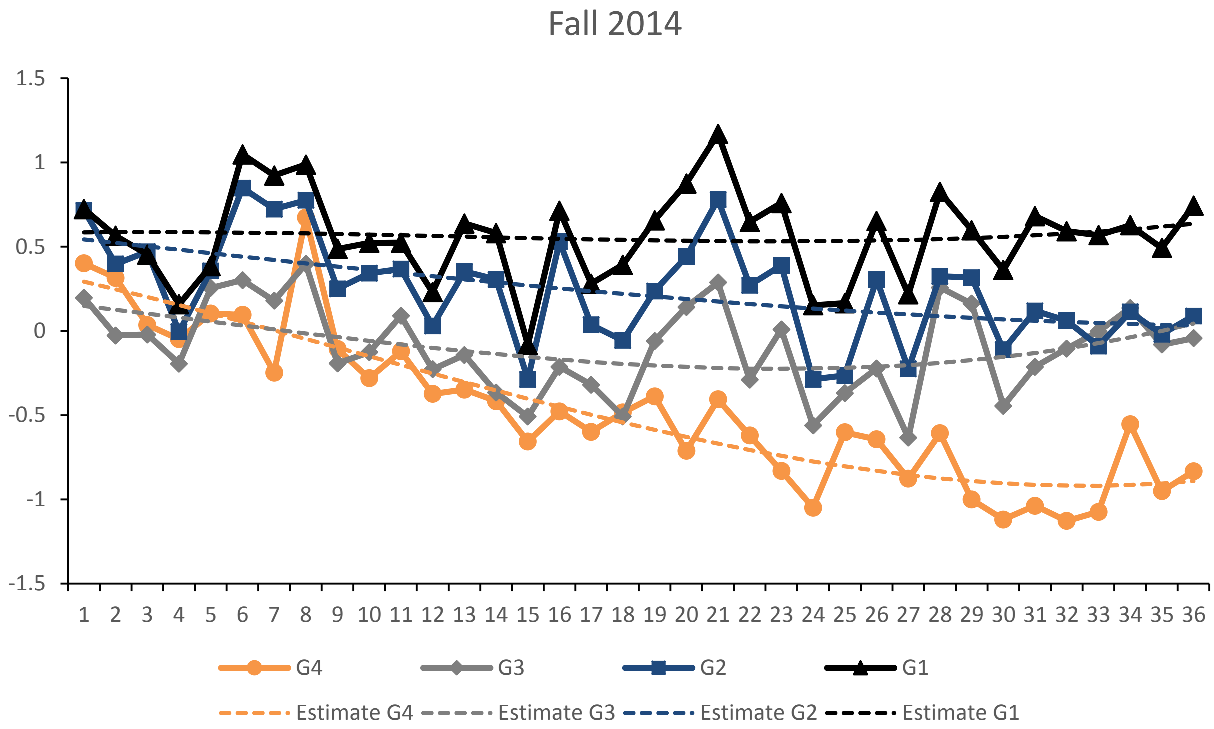


Table 3.7 displays the estimated intercept, the estimated coefficients for the linear, quadratic, and cubic terms of trajectory plots, standard error of each coefficient, and t-statistic value of the coefficient for each trajectory group. The statistic is for the test of the null hypothesis that the estimated parameter was equal to 0.

Table 3.7. Coefficients of clicker scores trajectory models (complete semester)

Group		Fall 2013	Spring 2014	Fall 2014
High (G1)	Constant	0.581±0.040 (14.362)	0.437±0.035 (12.649)	0.58256±0.04213 (13.827)
	Linear	-0.019±0.009 (-2.195)	0.043±0.008 (5.229)	0.00298±0.00969 (0.307)
	Quadratic	0.00057±0.0005 (1.072)	-0.004±0.0005 (-8.192)	-0.00055±0.0006 (-0.905)
	Cubic	0±0.00001 (-0.259)	0.0001±0.00001 (9.883)	0.00001±0.00001 (1.303)
High-Intermediate (G2)	Constant	0.163±0.064 (2.547)	0.241±0.069 (3.489)	0.56454±0.06576 (8.585)
	Linear	0.003±0.014 (0.226)	0.056±0.016 (3.469)	-0.02093±0.01441 (-1.452)
	Quadratic	-0.001±0.0008 (-1.56)	-0.006±0.001 (-5.85)	0.00003±0.0009 (0.031)
	Cubic	0.00003±0.00001 (2.284)	0.0001±0.00001 (6.699)	0±0.00002 (0.245)
Low-Intermediate(G3)	Constant			0.17118±0.12922 (1.325)
	Linear			-0.02213±0.02676 (-0.827)
	Quadratic			-0.00032±0.00163 (-0.197)
	Cubic			0.00002±0.00003 (0.797)
Low (G4)	Constant	0.028±0.122 (0.226)	0.019±0.097 (0.196)	0.336±0.09938 (3.381)
	Linear	-0.021±0.027 (-0.783)	0.046±0.026 (1.74)	-0.04247±0.02223 (-1.91)
	Quadratic	-0.001±0.002 (-0.935)	-0.006±0.002 (-3.508)	-0.00094±0.00139 (-0.674)
	Cubic	0.000040±0.00003 (1.557)	0.0001±0.00003 (3.666)	0.00003±0.00002 (1.311)

Note: the estimated coefficient ± standard error (t-statistic)

Composite predictive power

The BPPs from the trajectory analysis of all clicker scores explained 30.5% in fall 2013, 39.2% in spring 2014, and 32% in the fall 2014 of the variation in final course grades (multiple regression, $p < 0.001$). The combination of BPPs with the pre-course diagnostic assessments, and student characteristics explained 54.9% in fall 2013, 50.1% in spring 2014 and 45% in fall 2014 of the variation in final grade (multiple regression, $p < 0.001$). This is a 25.9% in fall 2013, 35.4% in spring 2014 and 31% in fall 2014 increase over pre-course diagnostic assessments alone and a 24.4% in fall 2013, 10.9% in spring 2014, 13% in fall 2014 increase over BPPs alone.

The BPPs from the trajectory analysis of the first three weeks of clicker scores combined with pre-course diagnostic assessments and student characteristics explained 47.3% ($p < 0.001$) in fall 2014, 21.3% in spring 2014 and 24% in fall 2014 of the variation in final grades. The BPPs from the trajectory analysis of the eighth three weeks of clicker scores combined with pre-course diagnostic assessments and student characteristics explained 56.3% ($p < 0.001$) in fall 2014, 36.8% in spring 2014 and 38% in fall 2014 of the variation in final grades.

Table 3.8. Fraction variance explained by each predictor of the final course grade

Classes of variables	Fraction of variance (R^2) explained		
semester	Fall 2013	Spring 2014	Fall 2014
Student characteristics	14.0%	3.7%	5.4%
Pre-course diagnostic assessments alone	29.0%	15.4%	14.23%
BPPs for 3 weeks clicker sessions	17.0%	4.3%	6.0%
BPPs for 8 weeks clicker sessions	34.0%	23.7%	22.6%
BPPs for all clicker sessions	30.5%	39.2%	32.0%

Table 3.9. Fraction variance explained by classes of variables of the final course grade

Classes of variables	Fraction of variance (R^2) explained		
semester	Fall 2013	Spring 2014	Fall 2014
Students characteristics + pre-course diagnostic assessments + BPPs for 3 weeks clicker sessions	47.3%	21.1%	30%
Students characteristics + pre-course diagnostic assessments + BPPs for 8 weeks clicker sessions	56.3%	37.6%	40.5%
Students characteristics + pre-course diagnostic assessments+ BPPs for all clicker sessions	54.9%	50.1%	47%

Note: Results are from multiple regression analyses using PROC REG and PROC GLM.

Discussion

All predictors including student characteristics were strongly associated with the final grade (R^2 = 54.9% in fall 2013, 50.1% in spring 2014 and 47% in fall 2014). Student characteristics explained 14% of the final course grade in the fall 2013 which was the higher than the 3.7% to 5.7% explained in the other semesters. Several published reports have noted that adding demographic characteristics (gender, ethnicity, family background) with academic predictors generally offers little additional predictive power (Allensworth & Easton, 2005; Balfanz & Neild, 2006; National Research Council and National Academy of Education, 2011). It may be that the fall 2013 participant sample was an exception to general educational patterns. However, large, introductory science courses have been shown to be associated with student alienation and low performance, particularly in students currently underrepresented in science careers (e.g., females, Black and Hispanic students; see PCAST, 2012). There were some differences between genders or among ethnicities in our data. White students had higher average final grades than others, and male students had higher average final grade than females. However, the relevance of these findings to an EWS is not clear.

In another study of the predictive power of formative assessments, Lesisko et al. (2012) reported how the "4Sight" test was capable of predicting students' performance in reading and mathematics. The inclusion of 4Sight exams explained 59.3% of the variability in reading and 54.5% of the variability in mathematics performance. Likewise, pre-diagnostic assessments scores used here were predictive of the final grade and explained higher variation than student characteristics, but less variance explained than the BPPs of 8 weeks clicker sessions and all clicker sessions. In addition, the content of the pre-diagnostic assessments in the fall 2013 was different from the spring 2014 and the fall 2014 semesters. That is, the variance explained of pre-diagnostic assessments was variable. The BPPs of clicker scores had the highest variance explained of the final course grade among all predictors. The variance explained by the BPPs for all clicker sessions ranged from 30.5% to 39.2%.

Research Question 2 was about the optimal time point to make robust prediction. The answer to Research Question 2 was that the average cumulative clicker score $\overline{S_{it}}$ of student *i* had practical predictive value. Two time points were considered for EWS. The first 3 weeks was regarded as the warning time before the first midterm, and the first 8 weeks as the deadline for dropping the course. As shown in the prediction and correlation analyses, the values of prediction and correlation for the first 8 week clicker scores were almost as high as the values of all clicker scores. If confirmed in future studies, this result could provide more data-driven decision about academic course add and drop deadlines, as well as more informed decision making by students. For example, students could be provided with information about their relative clicker score standing at the third week of the semester, and course add and drop deadlines could be adjusted accordingly. The finding that the correlation of $\overline{S_{it}}$ increased with *t* supported instructors providing students with information on likely performance outcomes as the course progressed. Clearly, many aspects of the educational system could benefit from robust predictive data. Those results demonstrated that clicker data could be used for these purposes.

The trajectory analysis of clicker data was a longitudinal analysis approach that provided information pertinent to understanding student performance patterns. Group-Based Trajectory Models (GBTM) placed students into distinguishable subgroups and reveal how these subgroups performed throughout the semester. Fall 2013 and spring 2014 semesters had three distinguishable trajectory groups while the fall 2014 semester had four trajectory groups. Student's probability of being in a particular trajectory group was a significant predictor of their final course grade. Thus, GBTM offered additional information on clicker performance patterns that could be a valuable input into EWS.

The trajectory analyses revealed heterogeneity in longitudinal performance patterns of the students in this course. The lowest-performing trajectory group continued to decline throughout the course, and the other two groups did not display dramatic relative changes in performance (Figure 3.3). The lowest-performing group could be the focus of additional efforts to examine whether they had the necessary prerequisites to master the material, or if specific aspects of the learning environment contributed to their systematic performance declines. In addition to identifying groups in need of targeted interventions, trajectory analysis may also stimulate investigations of learning patterns within very large STEM classes.

The answer to the fourth research question was that each of the sources of variables contributed incrementally to final grade prediction. That is, clicker scores explained more variation in final grade than pre-course diagnostic assessments scores. These results suggest that clicker scores were useful indicators of final course performance. However, combining both data sources--clicker scores and pre-course diagnostic assessments scores--generated the most robust predictions of student performance early in all semester. It was useful to use 3-week clicker scores before the first midterm and 8-week clicker scores before the course drop deadline. With this early warning system, student in the lowest group could prepare for the midterm and the final examination. Therefore, both approaches should be considered as useful inputs to EWS.

Chapter 4. Bivariate Trajectories for clicker scores and review scores to predict the final grade

Introduction

In the previous chapter, univariate trajectory analysis of clicker scores was used to predict final course grade. Clicker scores were collected in every class except for those cancelled because of weather in the spring 2014. In the spring 2014 and the fall 2014 offerings, students were also given a review quiz online for each scheduled class session. Review scores were collected after every class scheduled using online modules so that there were more review scores than clicker scores. I used only those review quizzes that had a matching clicker score. In this chapter, I examine the predictive value of the review quizzes, the trajectory groups describing the review quizzes, and the trajectory groups describing the bivariate trajectories of clicker scores and review quizzes. I hypothesize that using both clicker trajectory and review quiz trajectory BPPs will explain more variation of final course grade than using clicker trajectory BPPs alone. I used the PROC TRAJ CNORM model option for two variables and estimated the simultaneous trajectories for the review and clicker scores.

Results

Predictive value of the review quizzes

In chapter 2, I defined the fraction of correct predictions. Figure 4.1 displays the fraction of correct predictions using review quizzes for the two semesters. The correct prediction rate roughly increased to week 8 when it was 78% in the spring 2014 and remained roughly constant afterward. The rate at 3 weeks was 72.5% in the spring 2014. The prediction rate of review quizzes in the fall 2014 was lower than the prediction rate in the spring 2014. In this semester, the prediction rate at 3 weeks was 61.3% and remained roughly consistent afterward.

Figure 4.1. Fraction of correct predictions made by review quiz by week

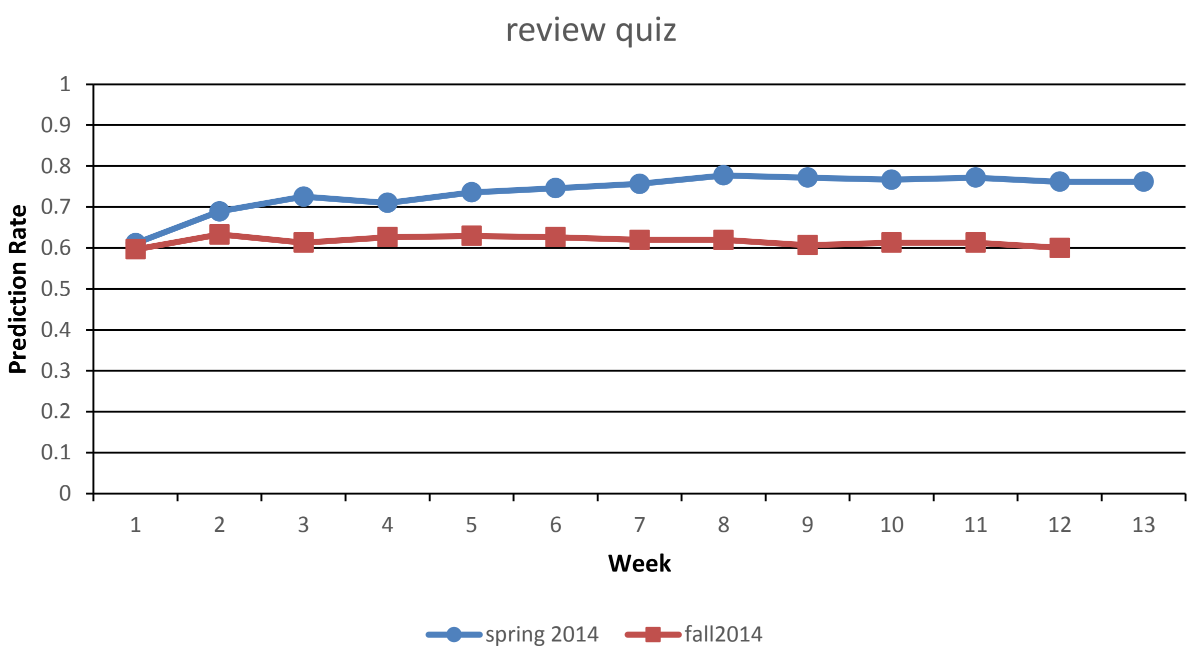
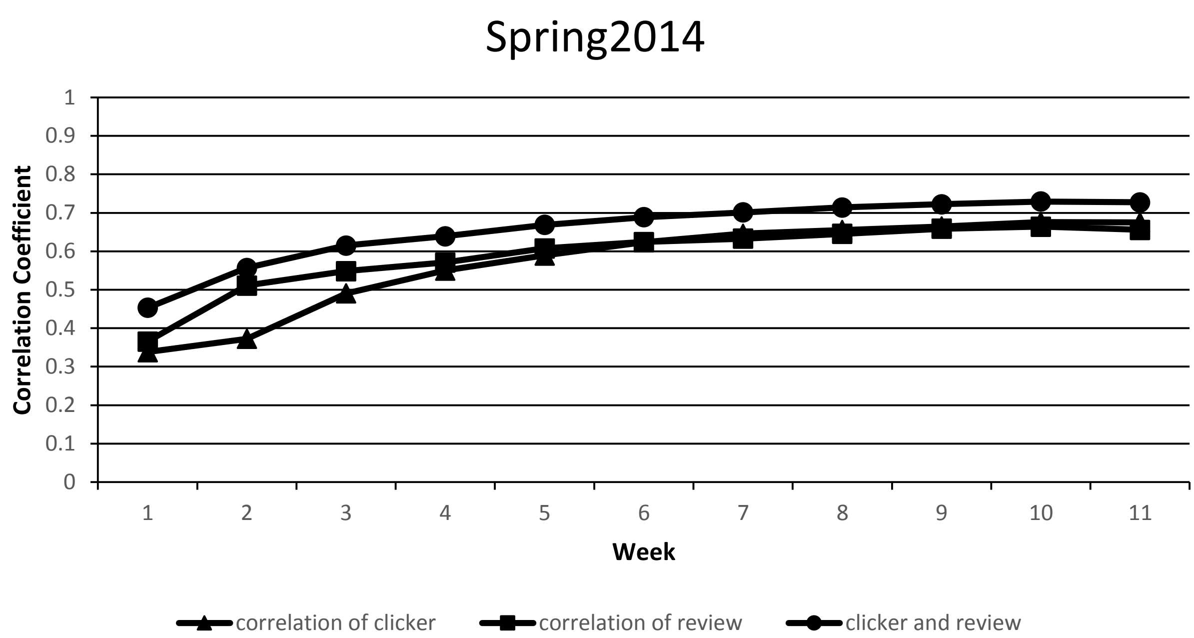


Figure 4.2. Correlation of cumulative average transformed review quiz scores with final grade, and clicker scores by week.



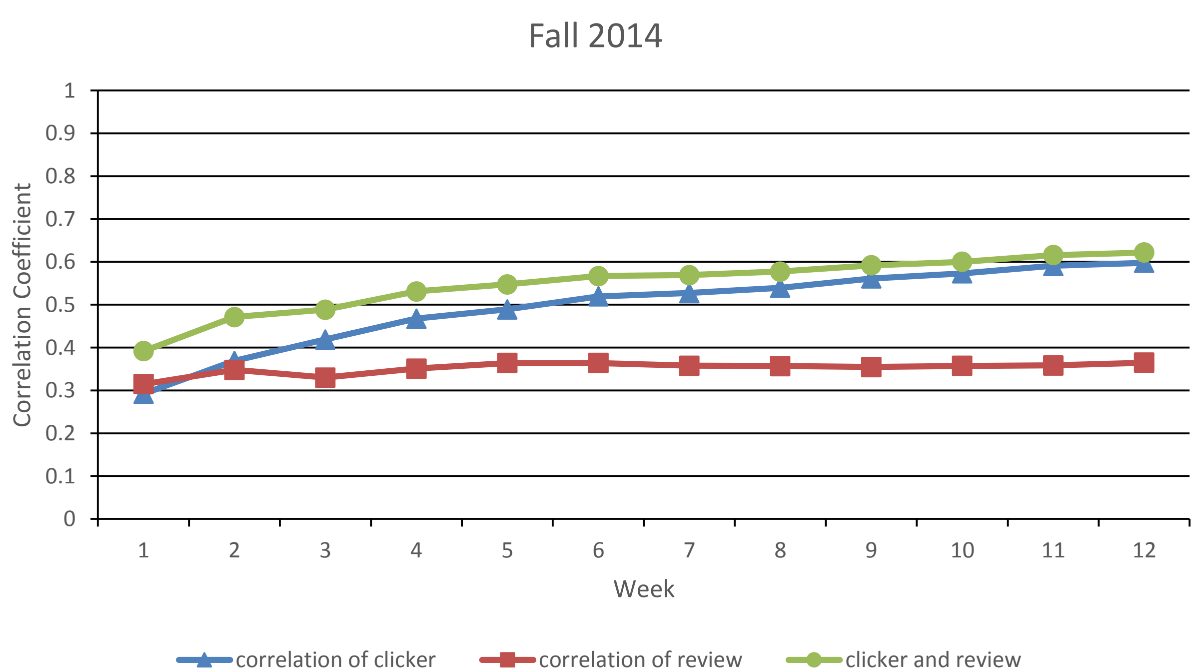


Figure 4.2 displays the correlation coefficients of cumulative average transformed clicker score, cumulative average transformed review quiz and the multiple correlation of these two variables with final grade for the spring and fall 2014 semester. The multiple correlation of the total average of clicker scores and review quiz scores was 0.73 in the fall 2014. This was larger than the correlation of the total average of clicker scores, which was 0.67, and the correlation of the total average of review quiz scores, which was 0.65. Throughout the spring 2014 semester, the correlation coefficient of the review quiz scores was higher than the correlation coefficient of clicker scores from the first week to the fifth week, but the correlation coefficient of clicker scores was equal to or higher than the correlation coefficient of review quiz scores from the sixth week to the final week. The correlation coefficient of the cumulative review quiz at 3 weeks was 0.55 and at 8 weeks was 0.64 at 8 weeks.

Throughout the fall 2014 semester the correlation coefficient of the review quiz scores was lower than the correlation coefficient of clicker scores from the first week to the final week. The multiple correlation of the total average of transformed clicker scores and review quiz scores was 0.61, the correlation of the average of all transformed clicker scores was 0.60, and the correlation of the average of all transformed review quiz scores was 0.36. The correlation coefficient of review quiz at 3 weeks was 0.33 and at 8 weeks was 0.36. The multiple correlation of transformed clicker scores and review quiz scores were higher than the correlations of the transformed clicker score by a relatively small amount compared to spring 2014. That is, the review quizzes did not substantially improve prediction of the final course grading this semester.

Univariate review quiz trajectory analysis

Table 4.1 contains the BIC scores of review quiz trajectory groups *G* ranging from 1 to 7 in the spring 2014 and the fall 2014. The 4-group model had the highest BIC score, -14781.11 in the spring 2014 and was used here. The highest BIC scores in the fall 2014 was 7 groups, but one prevalence of the trajectory groups was below 5%. Consequently, I selected the 5 trajectory group model because the prevalence of each trajectory group was above 5%.

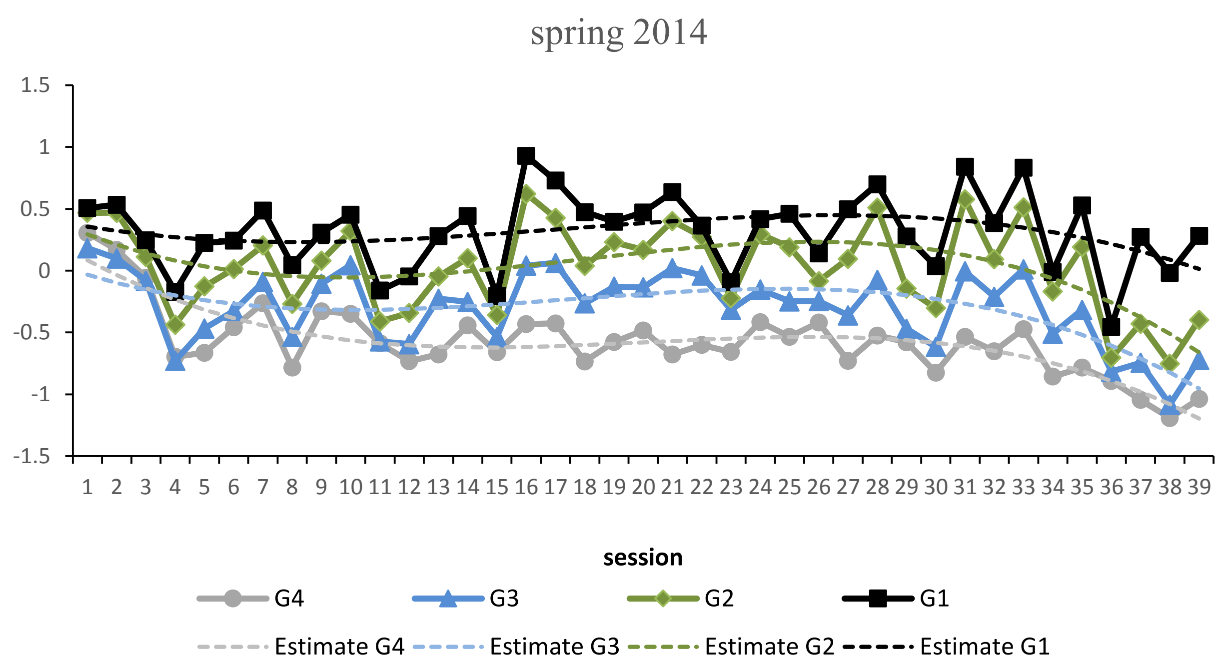
Table 4.1. BIC score of review quiz scores trajectory groups *G* ranging from 1 to 7

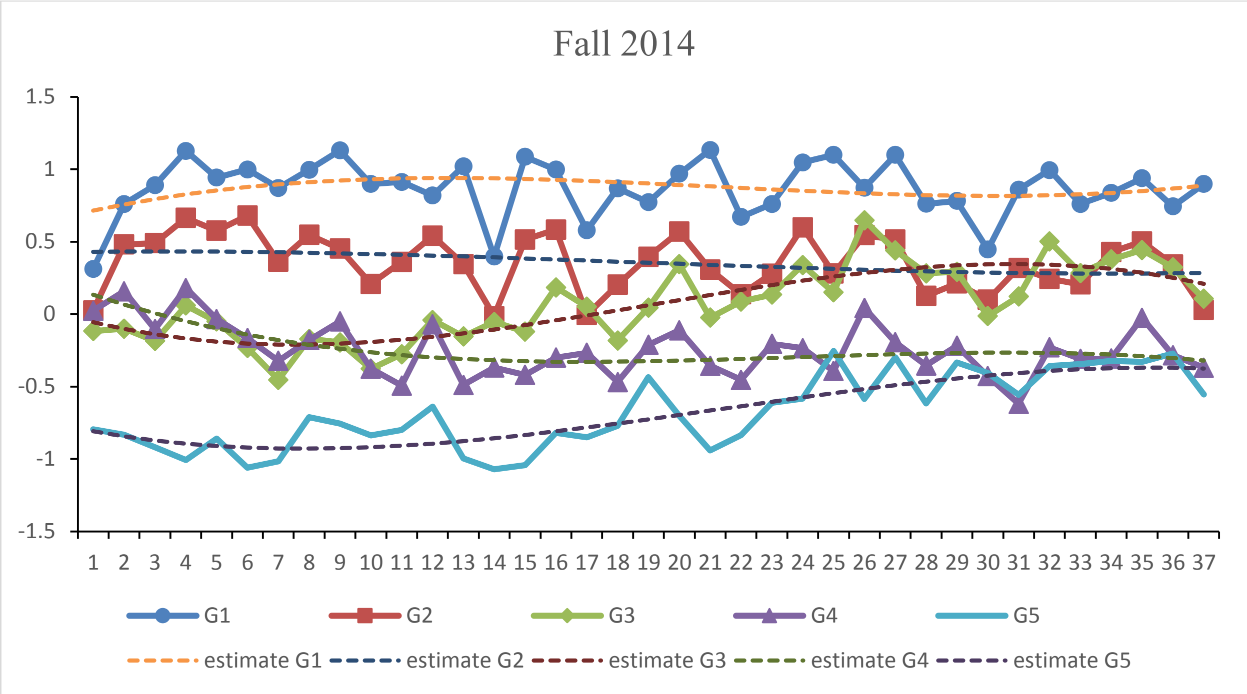
Semester / Group number	Spring 2014	Fall 2014
1	-15871.24	-14713.64
2	-14951.45	-13408.79
3	-14805.48	-13099.00
4	-14781.11	-13008.36
5	-14805.16	-12978.42
6	-14812.10	-12952.74
7	-14815.83	-12962.90

Figure 4.3 is the plot of the average review quiz score for each session of the four trajectory groups. As with the clicker score trajectories reported in chapter 3, the four average trajectories of review quizzes did not cross in the spring 2014. The group labeled 1 consistently had the highest average review quiz. The estimated prevalence of this group was 21.1%. The prevalence of the group labeled 2 was 39.2%. This group started at the similar point to the group 1, but decreased as the semester progressed. The prevalence of the group labeled 3 was 25.0%. This group started at the lower point than the group 4, but increased, being higher than group 4

after 4 weeks. The prevalence of the group labeled 4 which had the lowest review score was 14.7%. Students in group 1 had consistently highest review quiz scores through the semester, while students in group 4 consistently had the lowest review quiz scores. While the gap among all groups was not large in the beginning of the semester, the gap between the highest group and the lowest group gradually increased over the semester.

Figure 4.3. Estimated trajectory patterns of review quiz scores - (CNORM model: Censored Normal Distribution).





The fall 2014 review quizzes had five trajectory groups. The group labeled 1 consistently had the highest average review quiz score. The estimated prevalence of this group was 22.6%. The prevalence of the group labeled 2 was 23.6%. This group had consistently high average of review quiz, but was lower than the group 1. The prevalence of the group labeled 3 was 18.5%. This group started at lower point to the group 4 for first two weeks, but increased as the semester progressed and was similar to the group 2 at the final point. On the other hands, the prevalence of the group labeled 4 was 23.6%. This group started at a higher point than group 3, but its trajectory decreased and was similar to the group 5 at the final point. The prevalence of the group labeled 5 was 11.6%. The lowest trajectory of review quizzes in the fall 2014 semester did not decrease over time.

Table 4.3 displays the estimated coefficients for the intercept linear, quadratic, and cubic terms of trajectory plots, the standard errors of the coefficients, and the t-statistic values of the coefficients for each trajectory group. The T-statistic was for the test of the null hypothesis that the estimated parameter was equal to 0.

Table 4.2. The average of the final course grade of review quiz trajectory groups

Semester / Group	Spring 2014	Fall 2014
Group 1	3.8	2.6
Group 2	3.3	2.4
Group 3	2.5	2.2
Group 4	1.9	1.9
Group 5		1.4

Table 4.3. Parameters estimated by trajectory plot (complete semester)

Group		Spring 2014	Fall 2014
G1	Constant	0.39859 ± 0.05369 (7.423)	0.66615±0.06606 (10.084)
	Linear	-0.04565 ± 0.01194 (-3.823)	0.05078±0.01466 (3.465)
	Quadratic	0.00365 ± 0.0007 (5.203)	-0.00287±0.00089 (-3.238)
	Cubic	-0.00007 ±0.00001 (-5.956)	0.00004±0.00002 (2.92)
G2	Constant	0.39118 ± 0.04211 (9.289)	0.42542±0.07255 (5.864)
	Linear	-0.10493 ± 0.00839 (-12.502)	0.00402±0.01874 (0.215)
	Quadratic	0.00741 ± 0.00049 (15.217)	-0.00061±0.0012 (-0.512)
	Cubic	-0.00014 ±0.00001 (-17.156)	0.00001±0.00002 (0.519)
G3	Constant	0.04509 ± 0.05779 (0.78)	-0.00078±0.12694 (-0.006)
	Linear	-0.08277 ± 0.01176 (-7.037)	-0.06102±0.02184 (-2.794)
	Quadratic	0.00575 ± 0.00064 (9.039)	0.00505±0.00122 (4.126)
	Cubic	-0.00011 ± 0.00001 (-10.662)	-0.00009±0.00002 (-4.182)
G4	Constant	0.21761 ± 0.06885 (3.161)	0.20795±0.07856 (2.647)
	Linear	-0.14141 ± 0.01407 (-10.053)	-0.07779±0.01635 (-4.759)
	Quadratic	0.00758 ± 0.00079 (9.56)	0.00357±0.00101 (3.542)
	Cubic	-0.00013 ± 0.00001 (-9.717)	-0.00005±0.00002 (-2.866)
G5	Constant		-0.76858±0.10087 (-7.62)
	Linear		-0.04447±0.02095 (-2.122)
	Quadratic		0.00348±0.00127 (2.733)
	Cubic		-0.00005±0.00002 (-2.444)

Bivariate trajectory results

Table 4.4 contains the BIC score of bivariate trajectory groups *G* ranging from 1 to 7 in 2014 spring and 2014 fall semesters. The largest value of BIC was 4 groups in the spring 2014 and 7 groups in the fall 2014. However, I excluded the 7 group model in the fall 2014 because the prevalence of some trajectory groups was below 5 %. The, 5 bivariate trajectories groups in the fall 2014 had both high BIC and prevalence of each group above 5%.

Table 4.4. BIC scores of bivariate trajectory groups

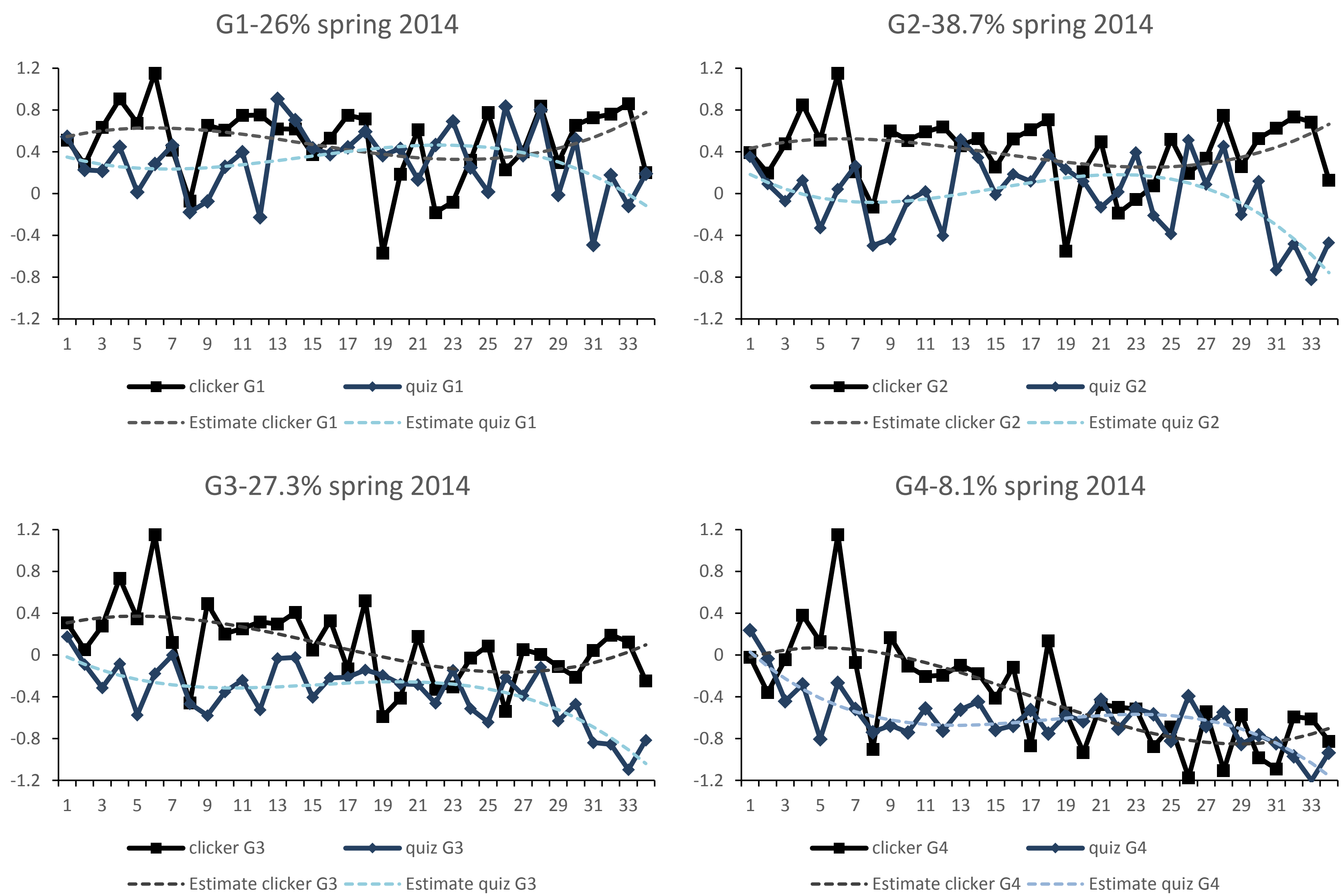
Semester	Spring 2014	Fall 2014
1	-28421.35	-27124.07
2	-27010.32	-25771.02
3	-26750.84	-25253.21
4	-26669.26	-25053.57
5	-26715.05	-24817.20
6	-26760.84	-24862.19
7	-26698.37	-24832.03

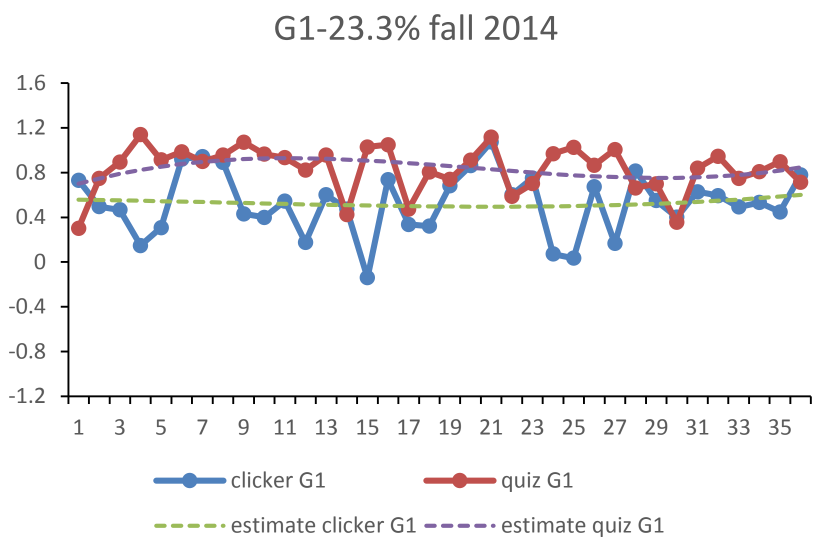
Figure 4.4 represents the bivariate trajectory patterns of clicker scores and review quiz scores. The prevalence of the group labeled 1 was 26.0%. Students in this group had consistently higher average score on both clicker and review quizzes throughout the semester. The prevalence of the group labeled 2 was 38.7%. Students in group 2 had high clicker scores, while their review quiz scores went down a little over time. The prevalence of the group labeled 3 was 27.3%. Students in group 3 had middle clicker scores and lower review quiz scores, compared to the

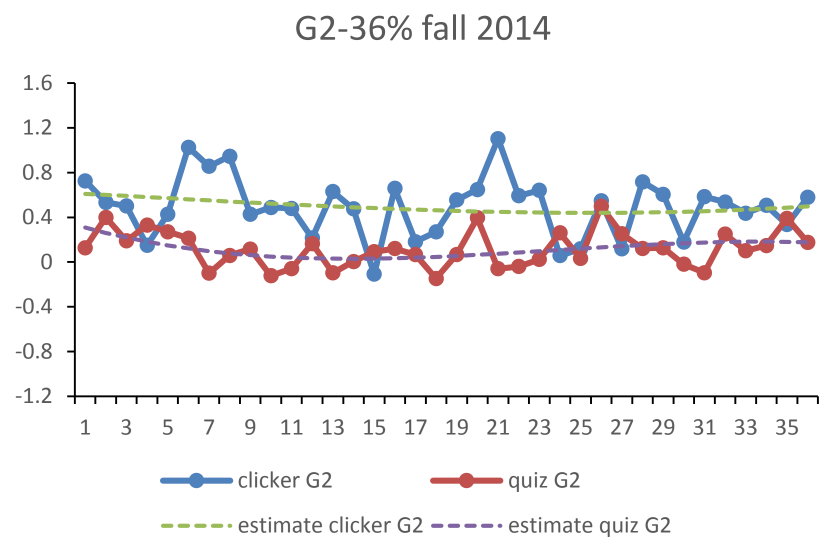
review quiz scores of group 2. The prevalence of the group labeled 4 was 8.1%. The scores of both clicker and review quizzes in group 4 were lower from the beginning of the semester to the end of the semester. This group generally had lowest transformed score.

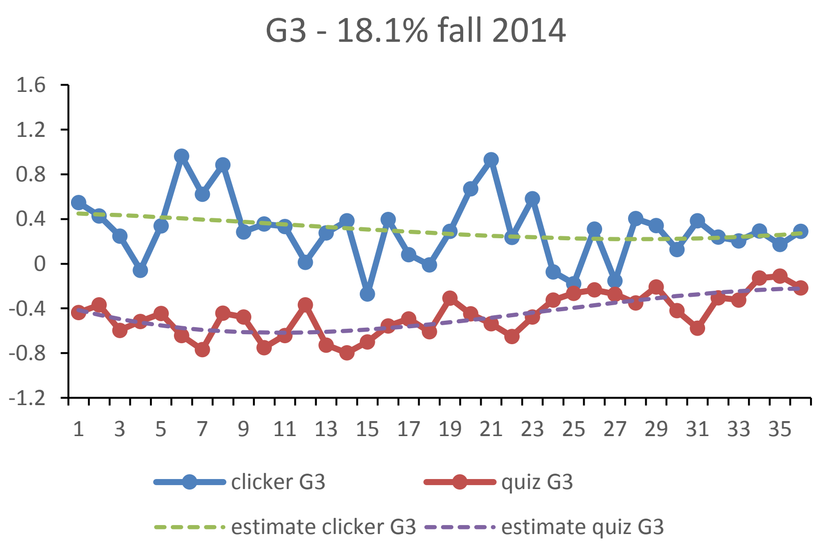
In the fall 2014, there were five trajectory groups. The group 1 consistently had the highest clicker and review quiz scores through the semester. The group 2 had the similarly high trajectory of clicker scores to the group 1, but the review quiz scores were consistently intermediate. The group 3 had the intermediate trajectory pattern of clicker scores and the lowest trajectory pattern of review quiz scores in the fall 2014. The group 4 had decreasing trajectory of clicker score but increasing trajectory of review quiz scores. The group 5 had both lowest trajectories of clicker and review quiz scores, and the trajectory of clicker scores was decreasing as the semester progressed.

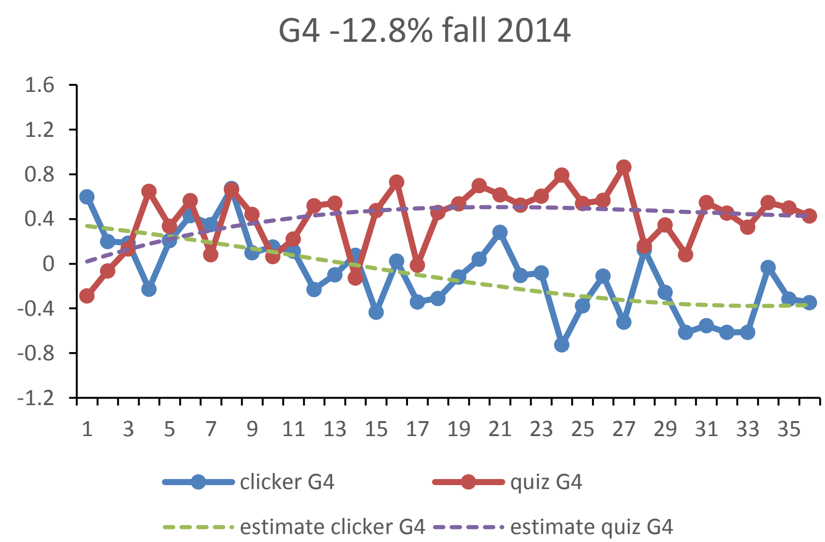
Figure 4.4. Bivariate trajectories plot in the spring 2014 and the fall 2014

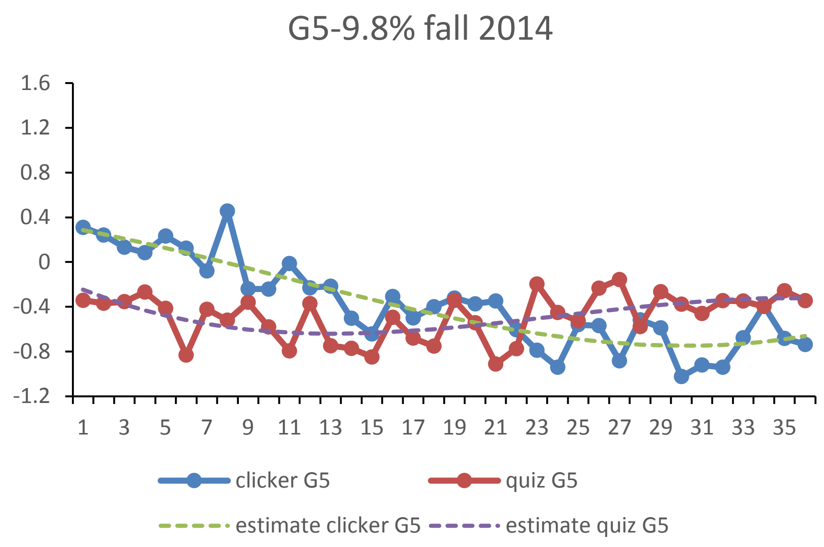



G1-23.3% fall 2014
1.6
1.2
0.8
0.4
0
-0.4
-0.8
-1.2
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35
clicker G1
quiz G1
estimate clicker G1
estimate quiz G1


G2-36% fall 2014
1.6
1.2
0.8
0.4
0
-0.4
-0.8
-1.2
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35
clicker G2
quiz G2
estimate clicker G2
estimate quiz G2


G3 - 18.1% fall 2014
1.6
1.2
0.8
0.4
0
-0.4
-0.8
-1.2
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35
clicker G3
quiz G3
estimate clicker G3
estimate quiz G3


G4 -12.8% fall 2014
1.6
1.2
0.8
0.4
0
-0.4
-0.8
-1.2
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35
clicker G4
quiz G4
estimate clicker G4
estimate quiz G4


G5-9.8% fall 2014
1.6
1.2
0.8
0.4
0
-0.4
-0.8
-1.2
1
3
5
7
9
11
13
15
17
19
21
23
25
27
29
31
33
35
clicker G5
quiz G5
estimate clicker G5
estimate quiz G5

The average final grade differed among the bivariate trajectory groups. The average final grade for group 1 was 3.78, for group 2 was 3.33, for group 3 was 2.43, and for group 4 was 1.72 in spring 2014. These differences were significant (one way ANOVA, $F = 112.7$, $p < 2e\text{-}16$). In the fall 2014, the average final grade for group 1 was 2.79, for group 2 was 2.57, for group 3 was 1.75, for group 4 was 1.54, and for group 5 was 1.24. The differences among groups were significant (one way ANOVA, $F = 36.51$, $p < 2e\text{-}16$).

Figure 4.5. Boxplot of final grade by bivariate trajectory groups

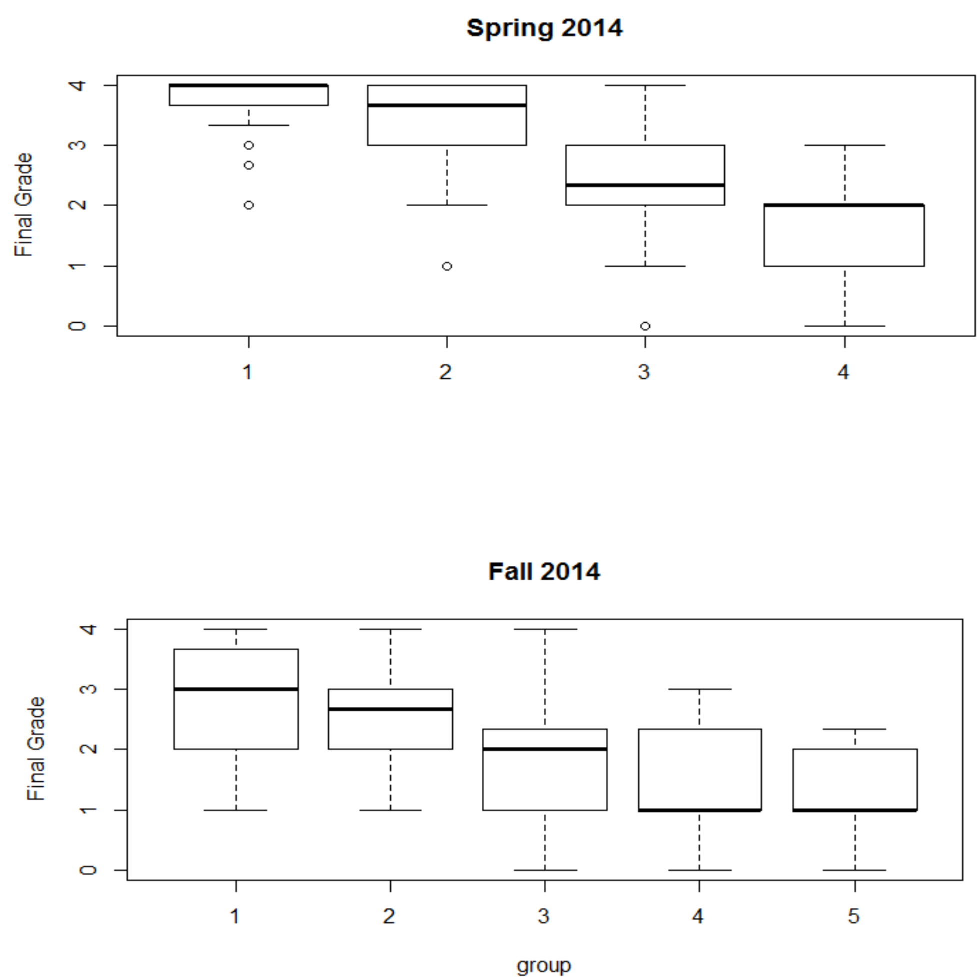


Table 4.5. Parameters estimated by bivariate trajectory plot (whole semester)

Spring 2014

Group		Review	Clicker
G1	Constant	0.39832 ± 0.05077 (2.004)	0.50478 ± 0.05403 (-0.731)
	Linear	-0.05441 ± 0.01303 (-7.546)	0.0459 ± 0.01334 (2.369)
	Quadratic	0.00527 ± 0.00088 (7.055)	-0.00488 ± 0.00089 (-4.091)
	Cubic	-0.00012 ± 0.00002 (-6.992)	0.00011 ± 0.00002 (4.278)
G2	Constant	0.27838 ± 0.04262 (1.287)	0.39073 ± 0.04497 (5.011)
	Linear	-0.1031± 0.01022 (-7.424)	0.04676 ± 0.01108 (3.263)
	Quadratic	0.00879 ± 0.00068 (8.207)	-0.00471 ± 0.00074 (-5.868)
	Cubic	-0.0002 ± 0.00001 (-9.552)	0.0001 ± 0.00001 (6.636)
G3	Constant	0.06223 ± 0.04836 (6.532)	0.27027 ± 0.05393 (8.688)
	Linear	-0.08761 ± 0.0118 (-10.084)	0.04367 ± 0.01339 (4.22)
	Quadratic	0.00639 ± 0.00078 (12.915)	-0.00521 ± 0.00089 (-6.399)
	Cubic	-0.00014 ± 0.00001 (-15.185)	0.00011 ± 0.00002 (7.572)
G4	Constant	0.18189 ± 0.09074 (7.846)	-0.07491 ± 0.10243 (9.343)
	Linear	-0.16458 ± 0.02181 (-4.177)	0.06052 ± 0.02555 (3.441)
	Quadratic	0.01004 ± 0.00142 (5.994)	-0.007 ± 0.00171 (-5.5)
	Cubic	-0.00019 ± 0.00003 (-7.18)	0.00014 ± 0.00003 (6.627)

Fall 2014

Group		Review	Clicker
G1	Constant	-0.16768±0.10726 (-1.563)	0.31995±0.09898 (3.233)
	Linear	-0.08242±0.02388 (-3.451)	-0.03339±0.02308 (-1.447)
	Quadratic	0.00432±0.00148 (2.921)	-0.0013±0.00143 (-0.908)
	Cubic	-0.00006±0.00003 (-2.271)	0.00004±0.00003 (1.608)
G2	Constant	-0.04096±0.09052 (-0.453)	0.35842±0.08542 (4.196)
	Linear	0.06342±0.02083 (3.045)	-0.01947±0.01968 (-0.989)
	Quadratic	-0.00231±0.0013 (-1.776)	-0.00081±0.00123 (-0.656)
	Cubic	0.00003±0.00002 (1.091)	0.00002±0.00002 (0.997)
G3	Constant	-0.36679±0.08367 (-4.384)	0.455±0.07476 (6.087)
	Linear	-0.05165±0.01812 (-2.851)	-0.0057±0.01708 (-0.334)
	Quadratic	0.00311±0.00113 (2.748)	-0.00048±0.00107 (-0.454)
	Cubic	-0.00004±0.00002(-2.148)	0.00001±0.00002 (0.734)
G4	Constant	0.36096±0.05601 (6.445)	0.6195±0.05181 (11.957)
	Linear	-0.05533±0.01284 (-4.311)	-0.00953±0.01189 (-0.802)
	Quadratic	0.00281±0.0008 (3.517)	-0.00008±0.00074 (-0.112)
	Cubic	-0.00004±0.00001 (-2.773)	0.00001±0.00001 (0.537)
G5	Constant	0.64489±0.06785 (9.505)	0.55887±0.06305 (8.863)
	Linear	0.05828±0.01536 (3.795)	-0.00209±0.01456 (-0.143)
	Quadratic	-0.00362±0.00096 (-3.777)	-0.00023±0.00091 (-0.257)
	Cubic	0.00006±0.00002 (3.51)	0.00001±0.00002 (0.558)

Table 4.5 displays the estimated intercept, the estimated coefficients for the linear, quadratic, and cubic terms of trajectory plots, standard error of each coefficient, and t-statistic value of the coefficient for each trajectory group. The statistic is for the test of the null hypothesis that the estimated parameter was equal to 0.

Table 4.6 displays the fractions of variance in the final grade explained using each predictor. In the spring 2014, the BPPs of the review quiz trajectory had higher fraction of variance explained than the BPPs of the clicker trajectory at 3 weeks, 8 weeks, and complete semester. BPPs of bivariate trajectories had high fraction of variance explained in the final course grade. However, in the fall 2014 semester, the variance explained by BPPs of the review quiz trajectory was low, compared to the clicker scores. The variance explained by BPPs of all review session is 13.4% and the variance explained by BPPs of all bivariate trajectories is 33.7%, which is lower than the variance explained by BPPs of clicker and review quiz session, 35.8%, in the fall 2014.

Table 4.6. Fraction variance explained by each predictor of the final course grade

Classes of variables	Variance of final grade (R^2) explained	
semester	Spring 2014	Fall 2014
Student characteristics	3.7%	5.4%
Pre-course diagnostic assessments alone	15.4%	14.23%
BPPs for 3 weeks clicker sessions	4.3%	6.0%
BPPs for 3 weeks review sessions	28.5%	11.2%
BPPs for 3 weeks bivariate trajectories	33.0%	11.5%
BPPs for 3 weeks clicker +review	30.3%	14.8%
BPPs for 8 weeks clicker sessions	23.7%	22.6%
BPPs for 8 weeks review sessions	39.8%	13.8%
BPPs for 8 weeks bivariate trajectories	47.9%	27.5%
BPPs for 8 weeks clicker +review	45.4%	28.3%
BPPs for all clicker sessions	39.2%	32%
BPPs for all review session	40.0%	13.4%
BPPs for all bivariate trajectories	49.0%	33.7%
BPPs for all clicker + BPPs for all review	49.6%	35.8%

Discussion

In this chapter, I used the review quiz scores to identify whether the review quizzes were useful for predicting the final course grade. For the first research question, I estimated the prediction rate and correlation coefficient of review quiz scores for the final course grade. The correlation coefficient of the review quiz in the spring 2014 semester was equal to or higher than correlation coefficient of the clicker scores in week 3, week 8 and complete semester. The prediction rates of the review quiz in the fall 2014 semester were consistently over 60%. When compared to the clicker scores, the correlation coefficient of the review quiz were lower than clicker scores in fall 2014. One possible reason was that students in the fall 2014 semester had a different testing procedure than students in the spring 2014. The standard deviations of the transformed review quiz scores in the fall 2014 were small. This restriction in the range may have reduced the correlation coefficient of review quiz scores in the fall 2014 semester.

As shown in figure 4.3 which was similar to figure 3.3 in chapter 3, the trajectory patterns did not cross from the start to the end of the semester. Also, the difference between review quiz trajectory groups increased over the semester. In the fall 2014, while there were consistently highest and lowest groups, one of intermediate trajectory groups crossed over. Even though the trajectory pattern of group 3 in figure 4.3 increased a little, the average of the final course grade of group 3 was still lower than the average of the final course grade of group 2.

There were 4 bivariate trajectory groups in the spring 2014, and 5 bivariate trajectory group in the fall 2014. Group 1 and group 4 in the spring 2014 and group 1 and group 5 in the fall 2014 had consistently highest and lowest averages for both scores. Other trajectory patterns were more complex. However, the average of the final course grade of each trajectory group followed the trajectory of clicker scores rather than the review quiz scores. For example, group 3 had intermediate trajectory of clicker scores and low trajectory of review quiz scores, while group 4 had intermediate trajectory of review quiz and decreasing trajectory of clicker scores in the fall 2014. The average of the final course grade of group 3 was 1.75, which was higher than the average of the final course grade of group 4, 1.55.

The fractions of variance explained by BPPs of the review quiz trajectory were higher than the fraction of variance explained by BPPs of the clicker scores trajectory in spring 2014, but not in the fall 2014. Therefore, review quiz, clicker scores, and the outcome of trajectory analysis and bivariate trajectories analysis were helpful to predict the final grade in the spring 2014 semester. However, the results showed that the review quiz was not helpful to predict the final course grade at any time points because of the low standard deviation and correlation coefficient of review quiz scores in the fall semester.

Chapter 5. The application of trajectory analysis to autocorrelation and partial autocorrelation

The correlation structure of each student's sequence of clicker scores and review quizzes may affect the distribution of the parameter estimates in the GBTM. Each individual student sequence may have a unique autocorrelation function (ACF) and partial autocorrelation function (PACF). Consequently, I will use GBTM techniques on the sequences of ACFs, the sequences of PACFs, and the bivariate sequences of ACF and PACF to assess the homogeneity and other properties of these functions. I hypothesize that there will be heterogeneity of correlation function structure with one trajectory group showing sequences with virtually no autocorrelation and partial autocorrelation. Calculated values of ACF and PACF functions were used as the longitudinal variables in trajectory analyses to identify their patterns and groups.

Results

ACF of clicker scores

Table 5.1 showed the BIC scores for the ACF sequences when the number of trajectory groups ranged from 1 to 6 for each semester. All BIC scores were positive. I used the maximum BIC value in selecting the optimal trajectory group, provided that the models in each trajectory group had 6 or more members. As a result, I selected two groups in fall 2013, three groups in spring 2014, and two groups in fall 2014.

Table 5.1. BIC scores of the ACF of clicker scores

Semester / Group	Fall 2013	Spring 2014	Fall 2014
1	2116.97	2119.25	2086.79
2	2197.59	2184.29	2307.50
3	2186.87	2201.09	2336.16
4	2188.23	2180.17	2370.37
5	2172.04	2164.24	2359.63
6	2160.69	2143.32	2343.85

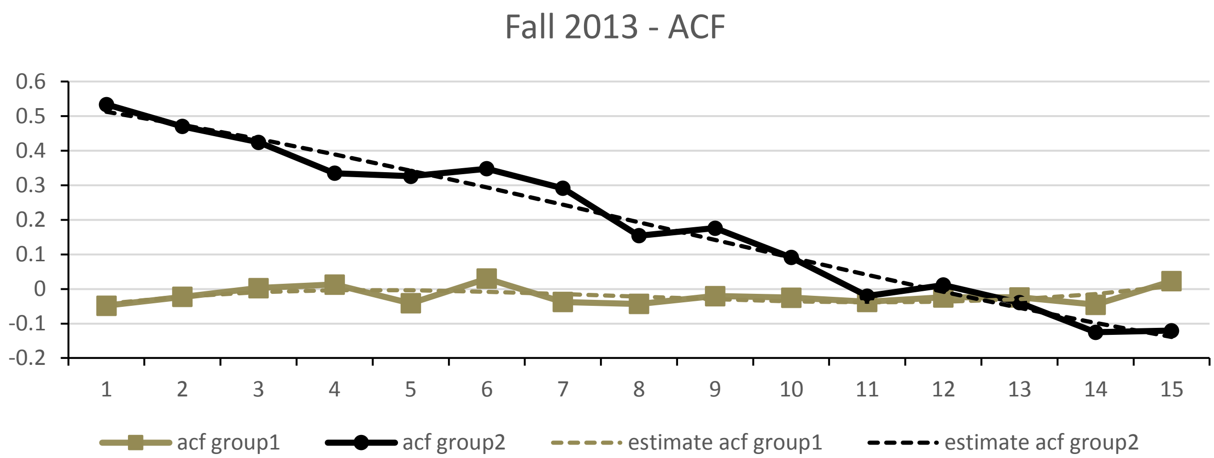
Table 5.2. The prevalence of each trajectory group using selected model

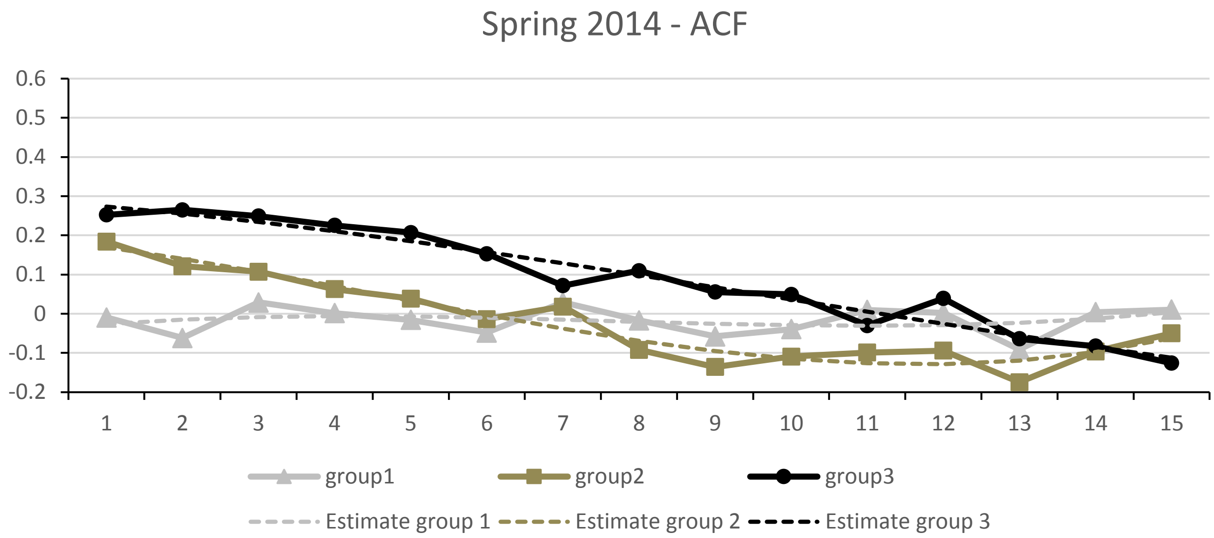
Semester / Group	Fall 2013	Spring 2014	Fall 2014
Essentially zero ACF	98.6%	78.8%	96.0%
Intermediate		15.7%	
Positive decreasing ACF	1.4%	5.5%	4.0%

Figure 5.1 showed the trajectory patterns of ACFs for each semester in spring 2014. In each semester, there was a trajectory group with ACF essentially zero at each lag. This group had prevalence 98.6%, 78.8%, and 96% in the three semesters as shown in table 5.2. Both fall semesters had a small trajectory group with ACF decreasing from 0.25 or 0.50 to 0. In spring

2014, there was the relatively small group (5.5%) with ACF decreasing from 0.25 to 0 as well as a larger group with ACF decreasing from 0.2.

Figure 5.1. Trajectory plot of ACF





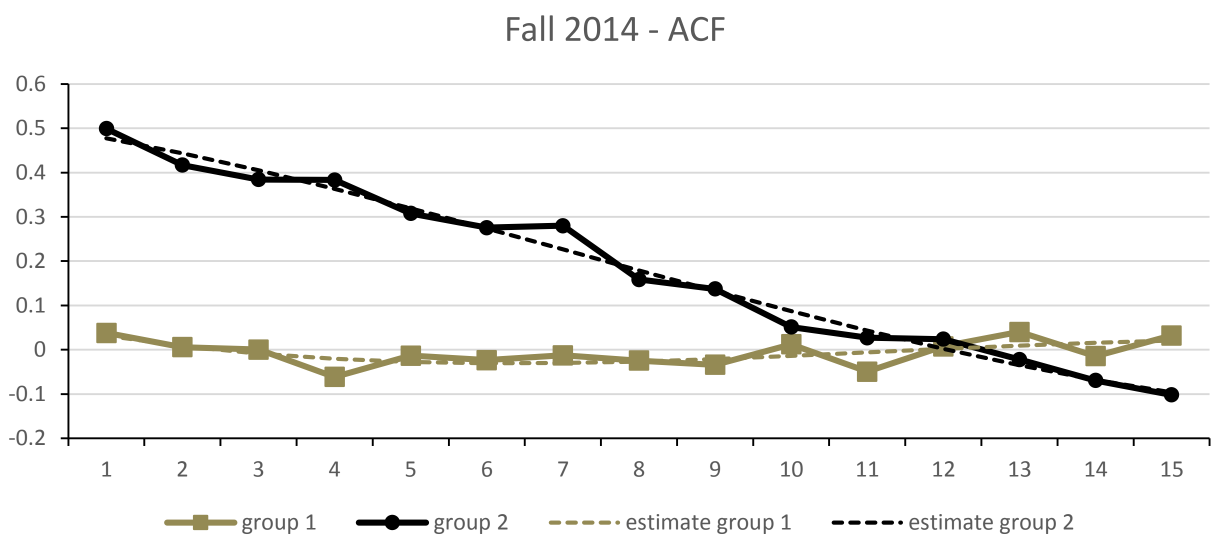


Table 5.3 displays the estimated intercepts, the estimated coefficients for the linear, quadratic, and cubic terms of trajectory plots, standard errors of coefficients, and t-statistic values for each trajectory group. The T-statistics are for the tests of the null hypothesis that the estimated parameters were equal to 0. Trajectory pattern for group 1 (essentially zero ACF function) looked consistent in all three semesters, but the coefficients for constant, linear, quadratic, and cubic were not 0. However, the value of the cubic coefficients were small (< 0.001).

Table 5.3. The coefficient of trajectory patterns

Group		Fall 2013	Spring 2014	Fall 2014
Essentially zero ACF	Constant	-0.07228±0.01155 (-6.257)	-0.04385±0.01381 (-3.174)	0.06138±0.01138 (5.394)
	Linear	0.03539±0.00605 (5.849)	0.02081±0.00682 (3.053)	-0.03342±0.00595 (-5.614)
	Quadratic	-0.00552±0.00086 (-6.384)	-0.00345±0.00096 (-3.591)	0.00367±0.00085 (4.311)
	Cubic	0.00023±0.00004 (6.593)	0.00015±0.00004 (3.839)	-0.00011±0.00004 (-3.082)
Intermediate	Constant		0.19363±0.03384 (5.723)	
	Linear		-0.02056±0.01681 (-1.224)	
	Quadratic		-0.00355±0.00236 (-1.504)	
	Cubic		0.00025±0.0001 (2.607)	
Positive decreasing ACF	Constant	0.54566±0.09722 (5.613)	0.28886±0.05123 (5.639)	0.5064±0.05747 (8.812)
	Linear	-0.03103±0.0509 (-0.61)	-0.01334±0.026 (-0.513)	-0.02649±0.0292 (-0.907)
	Quadratic	-0.00239±0.00727 (-0.329)	-0.00178±0.00372 (-0.477)	-0.00281±0.00417 (-0.674)
	Cubic	0.00009±0.0003 (0.316)	0.00006±0.00015 (0.383)	0.00013±0.00017 (0.735)

One question is whether these different ACF patterns predict student performance. Table 5.4 is the average course grade for each ACF trajectory group using modal BPP assignment.

Table 5.4. The average of the final grade of each trajectory group

Semester / Group	Fall 2013	Spring 2014	Fall 2014
Essentially zero ACF	2.77	3.11	2.25
Intermediate		3.14	
Positive decreasing ACF	2.2	1.99	1.08

Trajectory groups with greater autocorrelation had lower average groups. One possible source of autocorrelation was students receiving clicker scores of 0, probably due to cutting class. Table 5.5 gives the fraction of zero scores in each trajectory group. The trajectory groups with largest declines (group 2 in fall 2013 and fall 2014, and group 3 in spring 2014) had fraction of zero clicker scores near 50%.

Table 5.5. The average number of zero clicker scores

Semester / Trajectory Group	Fall 2013	Spring 2014	Fall 2014
Essentially zero ACF	5.44 (14.7%)	3.76 (11%)	5.9 (16.4%)
Intermediate		3.69 (10.8%)	
Positive decreasing ACF	18.5 (50%)	16.53 (48.6%)	19 (52.7%)

PACF of clicker scores

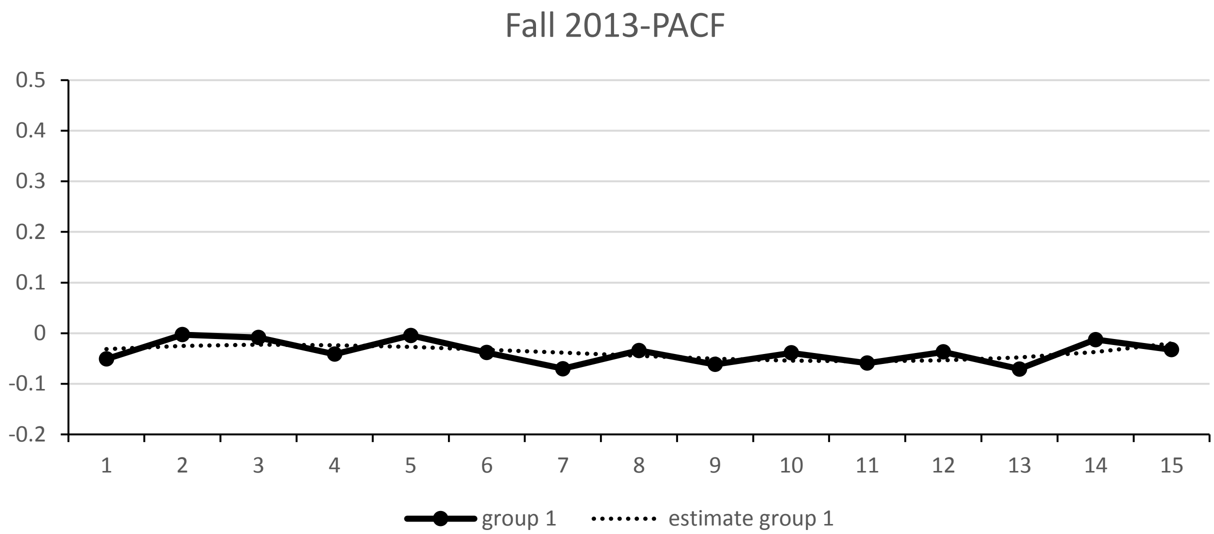
I used the maximum BIC value in selecting the optimal number of trajectory groups for the PACF sequences. As shown in table 5.6, there was only one trajectory group in each of the three semesters.

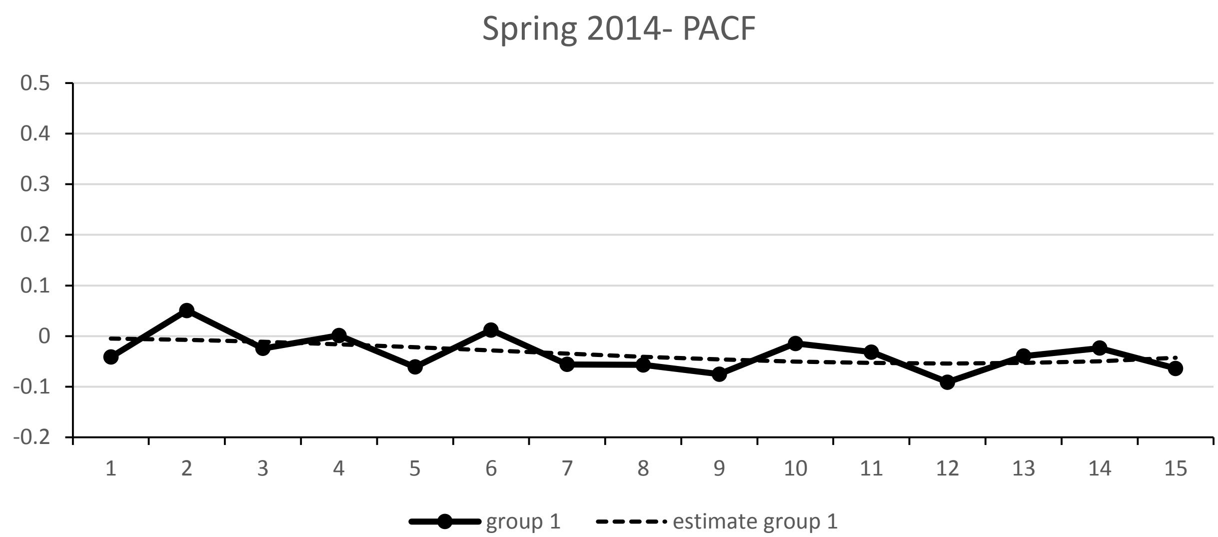
Figure 5.2 displayed the trajectory patterns of PACF. Fundamentally, the values of PACF were not as large as the values of ACF. Each semester had one trajectory group, with trajectory essentially constant at 0. On average, the values of PACF at all lags were within 0 ± 0.1.

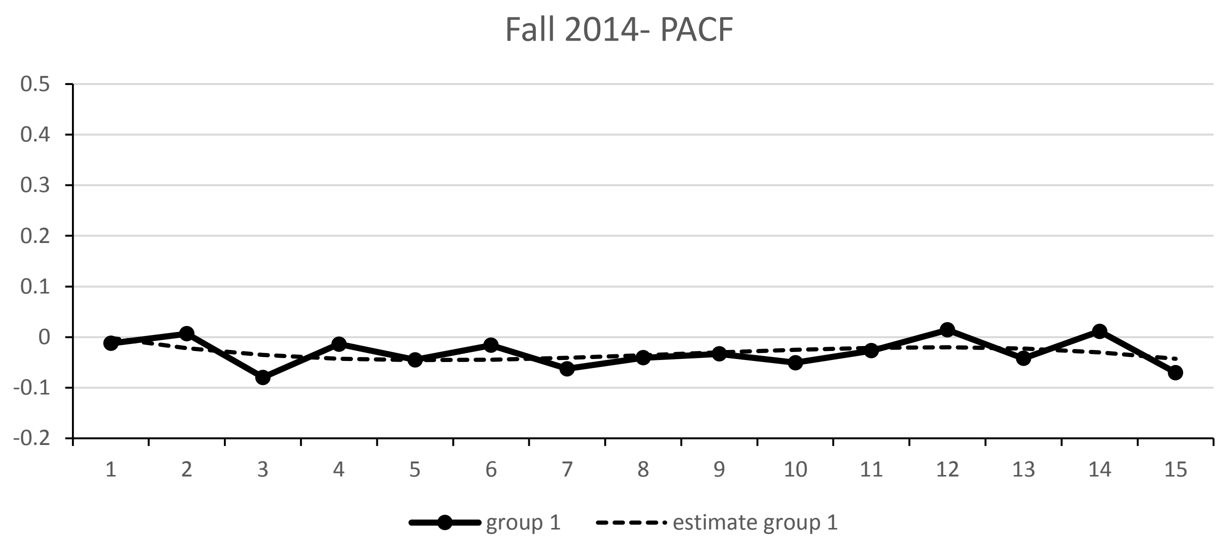
Table 5.6. BIC scores of PACF trajectory group

	Fall 2013	Spring 2014	Fall 2014
1	2443.11	2275.66	2483.89
2	2428.58	2275.24	2482.56
3	2411.06	2254.32	2467.48
4	2391.79	2233.81	2447.67
5	2370.87	2212.89	2427.02
6	2349.95	2191.97	2407.58

Figure 5.2. Trajectory plot of PACF







Trajectory groups for bivariate of (ACF, PACF)

Figure 5.3 displays the bivariate trajectories groups of ACF and PACF. Fall 2013 and fall 2014 semesters had two groups: (1) the patterns of group 1 (essentially zero ACF and PACF) for both functions were consistently around 0 from lag 1 to lag 15. (2) The patterns of group 2 (positive decreasing ACF and PACF) were positive and decreasing as the lag became larger. The pattern of PACF was below the pattern of ACF. The prevalence of the decreasing pattern's group was 4.5% in fall 2013 and 7.4% in fall 2014. The average of the final grade of this group was also lower than the average of the final grade of group 1. There were three groups in spring 2014. The pattern of group 1 (essentially zero ACF and PACF) was consistent with the other semester, i.e. essentially zero. However, PROC TRAJ found two decreasing patterns. Group 2 (intermediate) started at 0.1 for both PACF and ACF and decreased subsequently to 0. In group 3 (positive decreasing ACF and PACF), 6.6% had decreasing patterns of ACF and PACF starting at about 0.2. The trajectory patterns were associated with final course grade as shown in table 5.8. The average of the final grade of group 3 was 2.21, which was lower than the average of the final grade of group 2, 2.99. The average grade for this intermediate group was 2.99, slightly less than the average of group 1, 3.18.

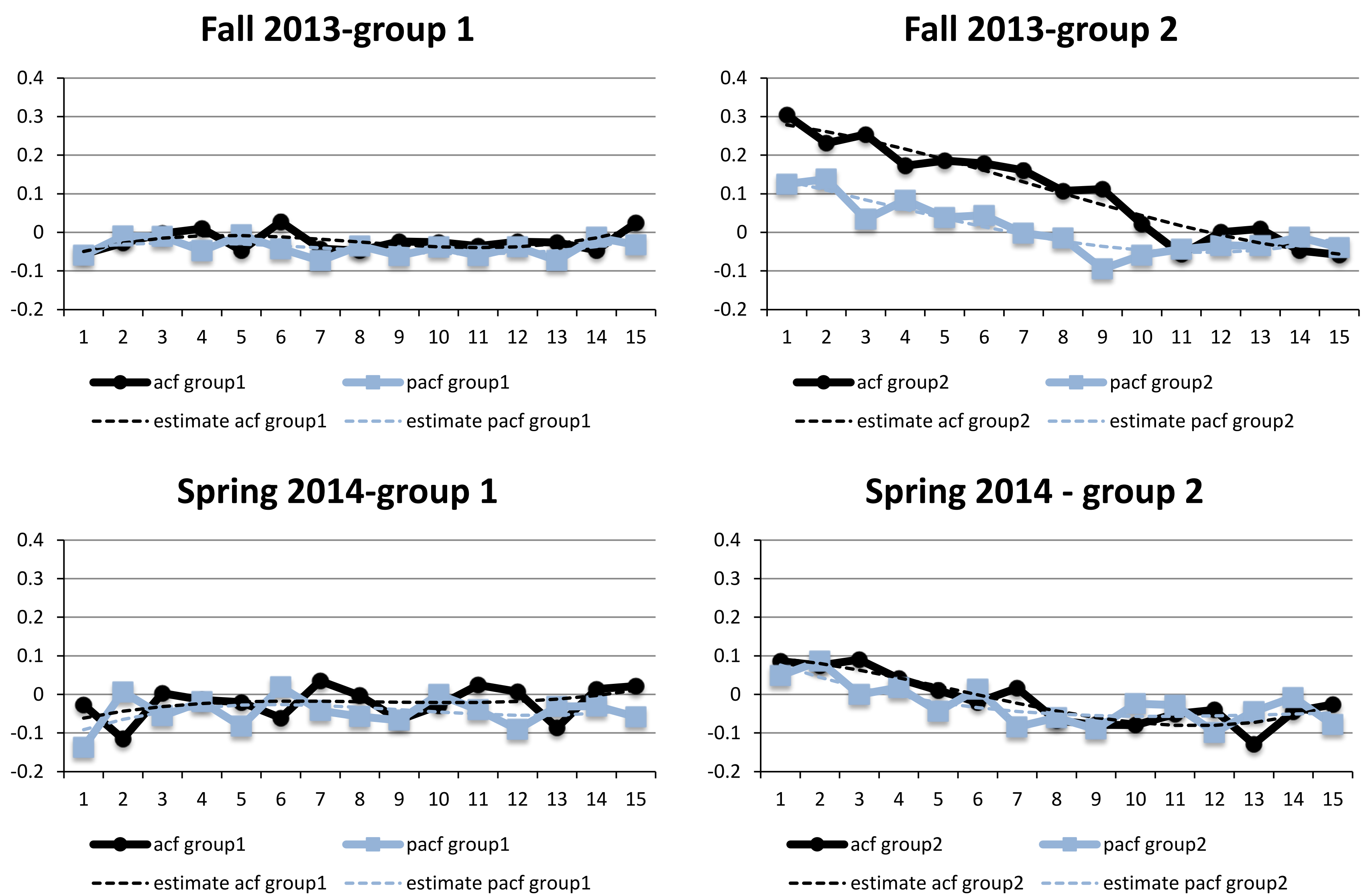
Table 5.7. The prevalence of bivariate trajectories groups

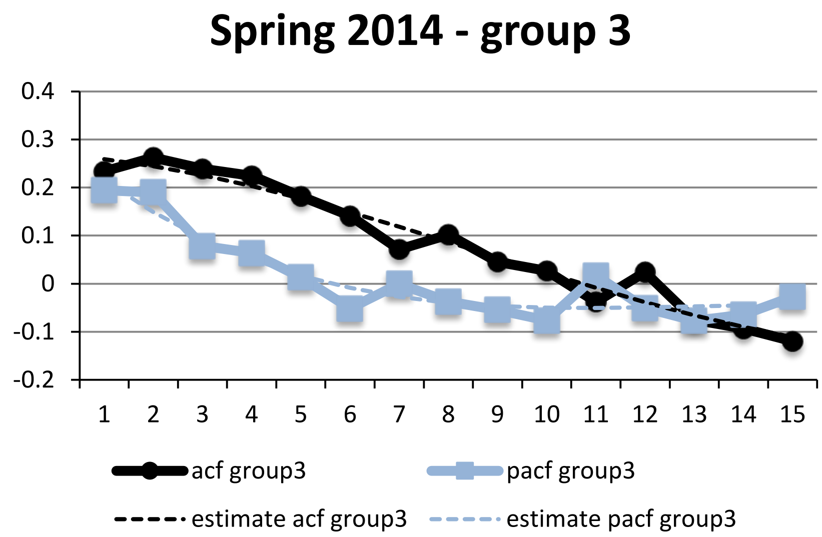
Semester / Group	Fall 2013	Spring 2014	Fall 2014
Essentially zero ACF and PACF	95.5%	53.7%	92.6%
Intermediate		39.7%	
Positive decreasing ACF and PACF	4.5%	6.6%	7.4%

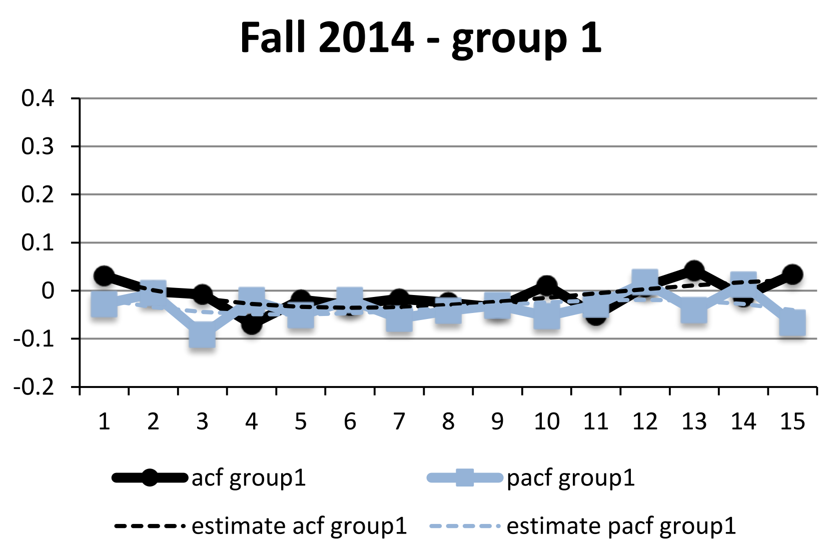
Table 5.8. The average of the final grade

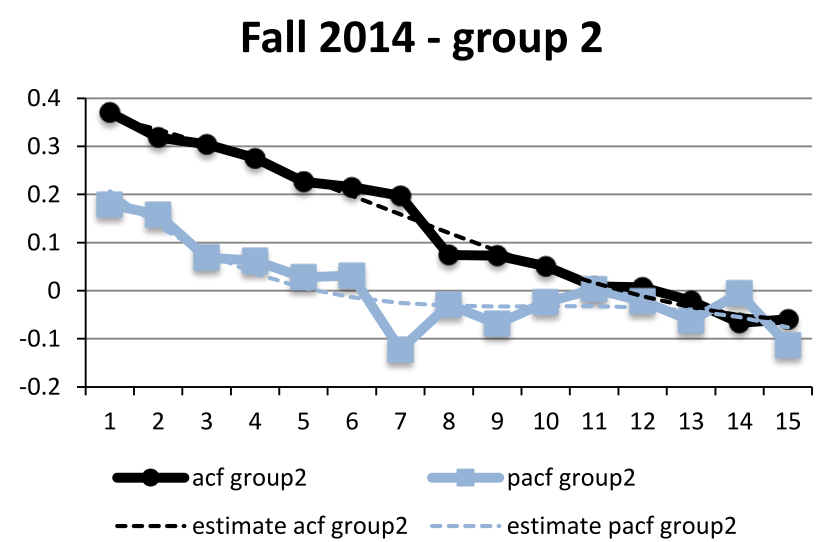
Semester / Group	Fall 2013	Spring 2014	Fall 2014
Essentially zero ACF and PACF	2.78	3.18	2.29
Intermediate		2.99	
Positive decreasing ACF and PACF	2.45	2.21	1.27

Figure 5.3. Bivariate trajectories group of ACF of clicker scores




Spring 2014 - group 3
0.4
0.3
0.2
0.1
0
-0.1
-0.2
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
acf group3
pacf group3
estimate acf group3
estimate pacf group3


Fall 2014 - group 1
0.4
0.3
0.2
0.1
0
-0.1
-0.2
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
acf group1
pacf group1
estimate acf group1
estimate pacf group1


Fall 2014 - group 2
0.4
0.3
0.2
0.1
0
-0.1
-0.2
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
acf group2
pacf group2
estimate acf group2
estimate pacf group2

As before, the fraction of zero scores is associated with both the autocorrelation structure of student clicker score sequence and student final grade. Table 5.9 contains the fraction of zero scores in each had low zero clicker score rates (between 10.6% and 15.6%). Students with notable ACF and PACF values had high zero clicker score rates (41.2% to 46.7%) and lower final course grade.

Table 5.9. The average of number of zero scores of each student

Trajectory group / Semester	Fall 2013	Spring 2014	Fall 2014
Essentially zero ACF and PACF	5.1 (14%)	3.6 (10.6%)	5.6(15.6%)
Intermediate	15.5(42%)	4.0 (11%)	16.8(46.7%)
Positive decreasing ACF and PACF		14.0(41.2%)	

Discussion

In this chapter, I calculated ACF and PACF of clicker scores and applied the trajectory and bivariate trajectory analyses for three semesters. Each semester had a large trajectory group with ACF sequence essentially zero as well as low zero clicker score percentages. Each semester had the different number of trajectory groups, but at least 1 % or at most 5% of students had the decreasing trajectory pattern of ACF in the course. Students with clicker scores trajectory pattern of ACFs starting at a positive value and decreasing were associated with students' fraction of zero clicker scores. The clicker score data does not show student attendance directly since the instructors gave 0 point to students if they did not submit their answers. I used the fraction of zero clicker score as a proxy for attendance rate. The fraction of zero scores of a student was significantly associated with the Bayesian posterior probability of being assigned to the trajectory pattern with decreasing ACF. As shown in the table 5.4 and 5.8, the low final course grade was also associated with the decreasing trajectory group of ACF and bivariate trajectories group of (ACF, PACF). Therefore, trajectory group of ACF and bivariate trajectories group of (ACF, PACF) may be informative about the association of student's final course grade and their attendance.

This analysis suggests an alternative approach to the analysis of this data. When a student's clicker score is zero, researchers could treat the score as missing and impute its value using the missing data algorithms in PROC TRAJ or MPLUS. I hypothesize that these sequences would have essentially zero ACF and PACF sequences. The results of the trajectory analyses may change. The cumulative fraction of zero clicker scores could then be evaluated as a predictor of final course grade in an EWS.

Chapter 6. Conclusions and discussion

I examined the relation of each class of predictors, (i.e, student characteristics, pre-course diagnostics assessment, clicker scores, and review quiz scores) with final course grade. Several published reports have noted that adding demographic characteristics (gender, ethnicity, family background) with academic predictors generally offers little additional predictive power (Allensworth & Easton, 2005; Balfanz & Neild, 2006; National Research Council and National Academy of Education, 2011). My results were essentially consistent. White students had a higher average final course grade than other students, and that male students had a higher average final course grade than females. The fraction variance of final course grade explained by these variables was small here. Diagnostic assessments and clicker quiz scores were predictive of final course grade. However, review quiz showed inconsistent results as a predictor of final course grade. Review quiz scores were powerful predictors in spring 2014, but not fall 2014.

Trajectory analysis identified trajectory groups of clicker and review quiz scores in a STEM course. In univariate trajectory analysis, three semesters, (the fall 2013, the spring 2014, and the fall 2014 semester) had replicable trajectory groups of clicker scores. One of the hopes was that a group of students would have a crossing trajectory pattern, meaning that they were able to improve their relative academic performance through the semester. There was not such a pattern. There was a little difference in the number or pattern of trajectory groups for each semester. There was a highest group (which was most common), a lowest group (which was least prevalent), and an intermediate group. BPPs of clicker trajectory were also predictive of the final grade. Overall, using formative assessment and BPPs should be considered as predictor variables in an EWS for final course performance. Trajectory patterns of review quiz in the spring 2014 did not cross. These BPPs were also predictive of final grade in spring 2014. However, the review quizzes were less predictive in the fall 2014. Therefore, using review quizzes as a predictor needs further research and development.

In addition, trajectory analyses of ACF and PACF sequences documented the association of student's final grade with the trajectory of ACF and PACF. Students whose clicker scores were not associated with the past clicker scores (that is, low ACF and PACF sequences) had higher final grade than others and their attendance rate was also higher, based on the number of zero clicker scores. Therefore, the cumulative fraction of zero clicker quiz scores may be an early and powerful predictor of final course performance because it reflects attendance.

6.1 Answers for research questions

I repeat each of my research questions and give the answers found in my work.

Ch 3. Clicker scores

(RQ1) To what extent does each data source (i.e. clicker scores, concept inventory scores) predict final course performance?

Gender or ethnicity in student characteristics were associated with the final course grade. White or male students were likely to have high performance. However, my research showed that the variance explained by student characteristics are small for the final course grade, compared to the variance explained by pre-diagnostic assessment and clicker scores. Pre-diagnostic assessments were significant predictors, especially CINS, for the final course grade. The correlation coefficients of cumulative transformed clicker scores for the entire semester in all three semesters were over 0.60 and the prediction rates were over 70%.

(RQ2) When in the semester can accurate final course performance predictions be made using clicker scores and other data?

I select two time points within the semester for prediction. Since the correlation coefficient of the 3rd week was the highest before the first midterm in this course, I chose it as the first time point. Since the deadline for dropping the course was the 9th week, I chose the 8th weeks as my second time point.

I concluded that 3rd week and 8th week were effective time points for an EWS, based on the correlation and prediction analyses. The prediction rate and correlation coefficient then was only slightly less than the prediction rate and correlation for the whole semester.

(RQ3) How many distinct trajectory patterns characterize students' clicker performances?

There were three trajectory groups in the fall 2013 and spring 2014 semesters and four groups in the fall 2014. Although the fall 2014 semester had four groups, this semester had non-intersecting trajectories like other semesters. Therefore, this biology course has three or four parallel trajectory groups with no trajectory pattern intersecting another trajectory pattern.

(RQ4) To what extent do academic--including trajectory results—and non-academic variables predict final course performance?

Fundamentally, academic variables for the complete semester and non-academic variables were predictive in all three semesters.

Ch 4. Review quizzes added to clicker scores for prediction

(RQ1) To what extent do review quiz scores predict final course performance?

The review quiz scores in the spring 2014 semester were significant predictors of the final course grade, but less so in the fall 2014.

(RQ2) How many distinct trajectory patterns characterize students' review quiz scores?

The spring 2014 had four trajectory groups. The trajectory patterns of review quiz scores were intersecting similar to the trajectory patterns of clicker scores. The fall 2014 semester had five trajectory groups. The highest group and lowest group were consistent through the semester and some of intermediate groups were consistent, increasing or decreasing a little as the semester progressed.

(RQ3) How many distinct bivariate trajectory patterns characterize student' performance on clicker and review quizzes?

The spring 2014 semester had four bivariate trajectories groups, and the fall 2014 semester had five bivariate trajectories groups. Figure 4.4 shows the patterns of the two semesters. In terms of patterns according to the trajectory groups, the spring 2014 semester had parallel bivariate trajectories patterns among groups. However, the fall 2014 semester showed more variation of combinations of clicker and review quiz trajectory patterns. For example, one of the groups, group 4, had trajectory of clicker scores decreasing while the trajectory of review quiz scores were increasing. However, review quiz scores were less predictive of final course grade in this semester.

(RQ4) To what extent does adding review quiz scores--including trajectory and bivariate trajectory results—improve the prediction of final course performance?

The spring 2014 semester showed that the review quiz scores were good predictors of the final course grade, and explained higher variance than clicker scores. In addition, the results of bivariate trajectories of 3rd week and 8th week session were higher than the results of univariate trajectory. However, the fall 2014 semester had low predictive power of review quiz scores and was not helpful to the bivariate trajectories analysis.

Ch 5. Stochastic properties of time series

(RQ1) Does a clicker score in one session have associations with clicker scores in other sessions? That is, is each student's sequence of clicker scores "white noise"? Similarly, for the sequence of review quizzes?

Roughly 95% of students did not have significantly non-zero ACF or PACF values However, roughly 5% of students were in a trajectory group that had positive ACF.

(RQ2) How many distinct trajectory patterns characterize collection of student ACF and PACF functions?

There were two trajectory groups of ACFs in the fall 2013: (1) a decreasing non-zero ACF trajectory group and (2) consistently zero ACF trajectory group. The same patterns helped for the fall 2014 semester. There were three trajectory groups of the ACFs in the spring 2014 semester. There was one trajectory group of PACF in each of the three semesters.

(RQ3) How many distinct bivariate trajectory patterns of ACF and PACF are there?

There were two bivariate trajectory groups in the fall 2013 and 2014 semesters, and three bivariate trajectory groups in the spring semester. The patterns were similar to univariate trajectory pattern.

Autocorrelation of student clicker scores is not important as it occurs in about 5% of students. The students with significantly non-zero ACFs have many more zero clicker scores. Imputation of these zero values may reduce the ACF for these students. Therefore, formative assessments used here are replicable and may be valuable inputs to predict the final course performance. Trajectory analyses of clicker and review quiz scores can be useful to identify the patterns of students in the course. Trajectory analyses of ACF and PACF of clicker scores are associated with the final course grade and the attendance.

Further research

STEM courses include science, technology, engineering, and mathematics. However, my research used one biology class for three semesters. If the data were collected from various courses in STEM disciplines, the results would be more generalizable for early warning systems. In addition, there might be some reasons the review quizzes in the fall 2014 semester were not helpful in predicting the final course grade. Therefore, I would expect that clicker and review quizzes would have predictive power for the final course grade.

Reference

1. Allensworth EM, Easton JQ (2005). The OnTrack indicator as a predictor of high school graduation. Chicago, IL: Consortium on Chicago School Research.
2. Anderson DL, Fisher KM, Norman GJ (2002) Development and evaluation of the conceptual inventory of natural selection. J Res Sci Teach 39(10):952–978
3. Balfanz R, Neild RC (2006) Unfulfilled promise: The dimensions and characteristics of Philadelphia's dropout crisis, 2000–2005. Philadelphia, PA: Philadelphia Youth Network.
4. Brook JS, Lee JY, Rubenstone E, Brook DW, Finch SJ (2014) Triple comorbid trajectories of tobacco, alcohol, and marijuana use as predictors of antisocial personality disorder and generalized anxiety disorder among urban adults. Am J Pub Health 104(8): 1413-1420
5. Bruff D (2011) Classroom response system ("clickers") bibliography. Center for Teaching, Vanderbilt University.
6. Caldwell JE (2007) Clickers in the large classroom: Current research and best-practice tips. CBE-Life Sci Educ 6(1): 9-20
7. Daempfle, P. A. (2003) An Analysis of the High Attrition Rates Among First Year College Science, Math, and Engineering Majors. Journal of College Student Retention 5, no. 1: 37-52.
8. Dobson JL (2008) The use of formative online quizzes to enhance class preparation and scores on summative exams. Adv Physiol Educ 32(4): 297-302
9. Douglas D. (2006) Clickers: A New Teaching Aid with Exceptional Promise. The Astronomy Education Review, Issue 1, Volume 5:70-88, 2006

10. Eddy SL, Brownell SE, Wenderoth MP (2014) Gender gaps in achievement and participation in multiple introductory biology classrooms. CBE-Life Sci Educ 13(3): 478-492
11. Granger, C. W. J. (1981) Some properties of time series data and their use in econometric model specification, Journal of Econometrics 16, 121—130.
12. Haudek KC, Kaplan JJ, Knight J, Long T, Merrill J, Munn A, Nehm RH, Smith M, Urban-Lurain M (2011) Harnessing technology to improve formative assessment of student conceptions in STEM: forging a national network. CBE-Life Sci Educ 10(2): 149-155
13. Jones B. L., Nagin D. S. (2007) Advances in group-based trajectory modeling and an SAS procedure for estimating them. Sociological Methods & Research 35:542–571.
14. Jones S, Henderson D, Sealover P (2009)"Clickers" in the classroom. Teach Learn Nurs, 4(1):2-5.
15. Koenig, K., Schen, M., Edwards, M., Bao, L. (2012). Addressing STEM retention through a scientific thought and methods course. Journal of College Science Teaching, 41, 23-29.
16. Lee O, Maerten-Rivera J, Penfield RD, LeRoy K, Secada WG (2008) Science achievement of english language learners in urban elementary schools: Results of a first-year professional development intervention. J Res Sci Teach 45(1): 31-52
17. Lesisko LJ, Sraiheen A. (2012) Utilizing a benchmark formative assessment to predict academic achievement in a rural school system. Int J Soc Sci Educ 3(1): 89-97

18. Libarkin JC (2008) Concept inventories in higher education science. Prepared for the National Research Council Promising Practices in Undergraduate STEM Education Workshop 2 (Washington, D.C., Oct. 13-14, 2008).

19. McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., Roediger, H. L. , III. (2011). Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. Journal of Educational Psychology, 103, 399-414. doi.10.1037/a0021782.

20. McDaniel, M. A., Howard, D. C., Einstein, G. O. (2009). The read-recite-review study strategy: Effective and portable. Psychological Science, 20, 516-522.

21. Nagin DS, Land K (1993) Age, Criminal Careers, and Population Heterogeneity: Specification and Estimation of a Nonparametric Mixed Poisson Model. Criminology 31: 327-362.

22. Nagin DS, Odgers CL (2010) Group-Based Trajectory Modeling in Clinical Research. Annu. Rev. Clin. Psychol. 2010. 6:109–38

23. Nagin, D. S., Richard E. T. (2001) Analyzing developmental trajectories of distinct but related behaviors: A group-based method. Psychological Methods 6: 18– 34.

24. National Research Council and National Academy of Education. (2011). High School Dropout, Graduation, and Completion Rates: Better Data, Better Measures, Better Decisions. Washington, D.C.: The National Academies Press.

25. Nehm RH, Beggrow EP, Opfer JE, Ha M (2012) Reasoning about natural selection: diagnosing contextual competency using the ACORNS instrument. Am Biol Teach 74(2):92–98

26. Orr R, Foster S (2013) Increasing student success using online quizzing in introductory (majors) biology. CBE-Life Sci Educ 12(3): 509-514
27. Rath KA, Peterfreund AR, Xenos SP, Bayliss F, Carnal N (2007) Supplemental instruction in introductory biology I: Enhancing the performance and retention of underrepresented minority students. CBE-Life Sci Educ 6(3): 203-216
28. Richardson M, Abraham C, Bond R (2012) Psychological correlates of university students' academic performance: A systematic review and meta-analysis. Psychol Bull 138(2): 353-387
29. Rutledge ML, Warden MA (1999) The development and validation of the measure of acceptance of the theory of evolution instrument. Sch Sci Math 99(1): 13-18
30. Seymour, E., & Hewitt, N. M. (1997). Talking about leaving: Why undergraduates leave the sciences. Boulder, CO: Westview Press.
31. Strenta, A. C., Elliott, R., Adair, R., Matier, M., Scott, J. (1994). Choosing and leaving science in highly selective institutions. Research in Higher Education, 35(5), 513–547.
32. Torre MEL, Allensworth E, Jagesic S, Sebastian J, Salmonowicz M (2013) Turning around low-performing schools in Chicago. Research Report - The university of Chicago Consortium on Chicago school research
33. Watkins J., Mazur E. (2013) Retaining students in science, technology, engineering, and mathematics (STEM) majors J. Coll. Sci. Teach., 42, 36-41