

Using Markov Chain Model to Evaluate Medical Students' Trajectory on Progress Tests and Predict USMLE Step 1 Scores

Ling Wang (✉ wangli35@msu.edu)

Michigan State University <https://orcid.org/0000-0003-2452-6691>

Heather S. Laird-Fick

Michigan State University

Carol J. Parker

Michigan State University

David Solomon

Michigan State University

Research article

Keywords: Human Medicine, Markov chain, Novice

DOI: <https://doi.org/10.21203/rs.3.rs-147714/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Medical students must meet curricular expectations and pass national licensing examinations to become physicians. The Michigan State University College of Human Medicine implemented progress testing in place of discipline-specific examinations as its primary assessment of knowledge in 2016. Ideally this innovative assessment strategy will characterize students' growth in basic science knowledge over time and predict licensing examination performance.

Markov chain method was employed to: 1) identify latent states of acquiring scientific knowledge based on progress tests, 2) estimate students' transition probabilities between states, and 3) predict United States Medical Licensing Examination Step 1 results based on the students' predicted probabilities in each state.

A total of 358 students were included in the analysis. Four latent states were identified based on students' progress test results: Novice, Advanced Beginner I, Advanced Beginner II and Competent States. At the end of the first year, students predicted to remain in the Novice state had lower mean Step 1 scores compared to those in the Competent state (209, SD = 14.8 versus 255, SD = 10.8 respectively) and had more first attempt failures (11.5% versus 0%). On regression analysis, it is found that at the end of the first year, if there was 10% higher chance staying in Novice State, Step 1 scores will be predicted 2.0 points lower ($P < .01$); while 10% higher chance in Competent State, Step 1 scores will be predicted 4.3 points higher ($P < .01$). Similar findings were also found at the end of second year medical school.

Using the Markov chain model to analyze longitudinal progress test performance offers a flexible and effective estimation method to identify students' transitions across latent stages for acquiring scientific knowledge. The results can help identify students who are at-risk for licensing examination failure and may benefit from targeted academic support.

Introduction

Medical education has evolved from a focus on the process of education to a focus on outcomes and demonstration of competence. This shift is, in part, founded on the work of Stuart E. Dreyfus and Hubert L. Dreyfus. They developed a model of skill acquisition through formal instruction and practice¹. The Dreyfus model proposes that a student passes through five distinct stages: novice, competence, proficiency, expertise, and mastery. Modeling growth as changes in developmental stages has proven to be useful in many fields. Examples include Piaget's stages of cognitive development², Kohlberg's stages of moral development³, stage-sequential models for reading development⁴ and paired associate learning⁵.

Progress testing assesses learner growth over time through the administration of examinations of similar content and difficulty across the curriculum. Tests are anchored to end-curricular expectations. Progress testing can be used for formative feedback or summative decision making. Progress tests can

complement other summative assessments that are given at endpoints of training, such as the United States Medical Licensing Examination (USMLE) Step 1 and Step 2 Clinical Knowledge examinations administered nationally to medical students.

In 2016 the Michigan State University College of Human Medicine adopted an innovative use of the National Board of Medical Examiners (NBME) Comprehensive Basic Science Examination (CBSE) and Customized Assessment Services (CAS) tests for progress testing twice per semester for the five semesters of the pre-clerkship curriculum. Minimum expectations for examination performance are established for each semester. The examinations contribute to students' grades and inform decisions about progression within the curriculum. Students must meet all curricular expectations, including performance on these progress tests, before sitting for the United States Medical Licensing Examination (USMLE) Step 1 and entering clerkships.

Methodological issues may limit the generalizability of progress tests to larger scale contexts, and their ability to predict future performance. For example, previous studies correlated scores on each iteration of a progress test with USMLE Step 1 results independently^{6,7} and found that later progress tests' scores were highly correlated with Step 1 performance. In these studies, the growth paths of performance on the progress tests were ignored. Another branch of studies modelled the growth of medical knowledge using progress tests⁸, but the growth of medical knowledge was not used to predict USMLE Step 1 results. Thus, it was unclear to medical educators how to best use the tests to confirm the effectiveness of the curriculum and predict student performance on the USMLE Step 1.

In this study we employed Markov chain methodology^{9,10} to evaluate medical students' dynamic trajectories on NBME CBSE and CAS examinations given as progress tests to predict their USMLE Step 1 performance. We selected this method because, in contrast to traditional ANOVA models, the Markov chain model considers the correlation of the previous state to the next one, naturally generating each student's growth pattern based on estimated steady-states. This is in contrast to Growth Mixture Modeling (GMM)¹¹, another approach to modeling growth over time, which estimates subgroup, not individual, growth patterns. These individual growth patterns, in turn, can be used to predict Step 1 performance parametrically.

The Markov chain approach to assessing growth in medical knowledge can be described as moving through several different states of knowledge as proposed by Dreyfus model. In the beginning, students have limited knowledge of medicine despite completing prerequisite science courses, and hence their performance is expected to be well below expectations for passing USMLE Step 1. This can be modeled by means of a **Novice state**, in which the probability of providing a correct answer is low. At the end of a course of study, students have attained a depth of medical knowledge, and hence having a very high probability of passing USMLE Step 1, which is called **Competent state**. Depending on their learning strategies, students may pass through several intermediate states, dubbed **Advanced Beginner states**, in which they have a growing but incomplete medical knowledge base. The number of latent states and the thresholds of each latent state can be estimated by the Latent Markov model¹². The transition dynamics

from Novice state to Advanced Beginner states and to Competent state of each student can be used to predict USMLE Step 1 performance. We hypothesized that students with higher transition probabilities to the Competent state would have better performance on Step 1. This study has three aims: (1) to identify the latent stages medical students go through in the first two years of medical school using progressive tests results, (2) to identify students' transition probabilities among different stages, and (3) to predict USMLE Step 1 results based on their transition probabilities. Findings obtained from our study could be valuable for individual medical student diagnostic and remediation purposes.

Methods

Student sample and measurement instruments

The sample was comprised of 358 medical students who matriculated to Michigan State University College of Human Medicine in Fall 2016 or Fall 2017 and finished Step 1 at the end of the second year of their program. NBME progress tests were administered twice per semester for five semesters. Thus, up to ten NBME test scores per student were collected for this cohort. Most students (86.2%) completed all ten NBME tests. Missing NBME test scores were imputed using multiple imputation method¹³. We also controlled for students' Medical College Admissions Test (MCAT) scores in the analysis, as prior studies have shown an association between MCAT and Step 1 scores^{14,15}. Linked, deidentified data were obtained using our college's honest broker (<https://omerad.msu.edu/research/honest-broker-for-educational-scholarship>), a methodology previously determined to be exempt by the Michigan State University Human Subjects Review Program.

NBME progress tests and USMLE Step 1 Results

The CBSE is a 200-item multiple-choice question (MCQ) exam made available by the NBME as a representative test blueprinted against the Step 1 content outline. Students receive scaled scores for the total test. Because there are a limited number of forms of the CBSE available at any given time, we also utilized the NBME's Customized Assessment Services (CAS) to create examinations. These examinations were blueprinted to provide subscores for categories corresponding to disciplines integrated within our curriculum. NBME generated both total test percent correct scores and locally scaled scores for the total test and individual content areas we defined during the test construction phase. Percent correct scores from CAS tests were used for our analysis. Among the ten NBME tests, there were six CAS and four CBSE tests for the 2016 cohort and five CAS and five CBSE test for the 2017 cohort.

The USMLE Step 1 is a large multiple choice question (MCQ) exam given nationally to medical students to test their basic science and, to some extent, clinical knowledge. Total score is reported, with a maximum of 300 and a minimum passing score of 194 established as of January 1, 2018. Step 1 scores on first attempts in the cohort ranged from 154 to 269, with mean score 230.5. A total of seventeen students failed (score < 194). Figure 1 shows the scatterplots between each iteration of NBME test versus Step 1 results.

Markov Chain Analysis

Markov chain is a stochastic model describing a sequence of possible events in which the probability of each event depends on the state attained in the previous event. To describe a Markov chain, we first define a set of states, $S = \{s_1, s_2, \dots, s_r\}$ where r indicates the number states at one time point. The process starts with one of the states and moves successively to the same state or another one at the next time point. The probabilities of moving from state s_i to s_j are called transition probabilities.

We first identified the number of latent states and the thresholds for each state using the Latent Markov Model². Grid search method was employed to find the best model fit using three to six latent states. The model likelihood for each latent state was obtained based on maximum likelihood estimators. The two most commonly used criteria [Akaike information criterion (AIC) and Bayesian information criterion (BIC)]¹⁶ were used to find the optimal number of latent states. The cutoffs of NBME test scores in each latent state were obtained based on estimated mean and standard deviation of scores in each state. A detailed setup and estimation algorithm of this Latent Markov Model is shown in the Appendix.

Once the number of latent states and score ranges for each latent stage were identified, the transition probabilities to each state were calculated for each student. Students' steady-states were estimated based on their transition probabilities. The steady state of the Markov chain is the probability distribution of each of the equilibrium states in the long run. In our study, a student's steady state is the likelihood (i.e. predicted probability) of the student attaining each latent state in the long run. If a student has a high likelihood of remaining in the Novice state, that suggests that the student has not been able to acquire, consolidate or apply knowledge, presumably because of inadequate learning strategies. Lastly students' probability distributions in steady-states were regressed against their Step 1 results to see if performance on progress tests predicted Step 1 scores.

Two scenarios were considered in this study: 1) steady-states obtained from the first six NBME tests were examined to determine whether it is possible to identify academically at-risk students at the end of the first year, and 2) steady-states obtained from all ten NBME tests were examined to determine if it is possible to identify students at risk of not passing Step 1. All statistical analyses were performed using R version 3.5.3.11¹⁷.

Results

Latent States

Table 1 shows the results of the model fit using different numbers of latent states. Based on the BIC criterion, four latent states gave the lowest BIC values and provided the optimal model fit: s_1 : **Novice state**; s_2 : **Advanced Beginner I state**, s_3 : **Advanced Beginner II state**, and s_4 : **Competent state**. The range of NBME test scores in each state were s_1 : [0, 44], s_2 : (44, 56], s_3 : (56, 69] and s_4 : (69, 100].

Table 1
Selection of number of latent states using Latent Markov Model

Number of Latent States	Test score ranges in each latent state	AIC ¹	BIC ²
3	[0,45], (45,59), (59,100]	25851.63	26014.38
4	[0, 44], (44, 56], (56, 69], (69, 100]	25875.68	25991.92
5	[0, 44], (44, 55], (56, 62], (62, 71], (71, 100]	26197.04	26243.54
6	[0, 43], (43, 50], (50, 56], (56, 64], (64, 72], (72,100]	25928.72	26006.22

Prediction of USMLE Step 1 based on first six progress tests

Students' transition probability based on the first six NBME tests are displayed in Fig. 2. Students with NBME scores in the category of Novice state (s_1) have a 58% of chance staying in the same state and 42% of chance improving to Advanced Beginner I state. Students in Advanced Beginner I state have a 76% chance staying in the same state, 14% chance improving to Advanced Beginner II state, and 9% chance of degrading to the Novice state. Students in Advanced Beginner II State have a 60% chance staying in the same category, 21% chance of improving, and 19% chance of degrading. The students in the Competent state have a 69% chance of staying in the same category and 31% chance of degrading at the end of the first year of medical school.

Based on the transition probabilities, the predicted probability in each state at the end of first year in medical school for all 358 students were: 8.5% Novice, 39.9% Advanced Beginner I, 30.4% Advanced Beginner II, and 21.2% Competent state. Each student's predicted probability in each state was estimated based on their own transition probability. Figure 4 shows the spaghetti plot of six NBME tests results grouped by the students' predicted probability in each state, with Advanced Beginner I and II states combined. We combined Advanced Beginner I and II states because the spaghetti plots of these two states didn't have strong differences. Among the 78 students who had predicted probability in the Novice state at the end of first year, the mean Step 1 score was 209 (SD = 14.8), and 7 failed Step 1 on the first attempt. On the other hand, 24 students had predicted probability to be in the Competent state at the end of first year, their mean Step 1 score was 255.1 (SD = 10.8).

Next the predicted probability distribution in each state were regressed on students' Step 1 results, controlling for MCAT scores. The regression results show that if the probability of steady-states in the Novice state increased by 10%, the predicted Step 1 scores would be 2.0 points lower ($P < .01$); 10% higher probability of steady-states in Advanced Beginner I state led to 0.3 points lower on Step 1 ($P = .53$); 10% higher probability of steady-states in Advanced Beginner II state led to 2.7 points higher on Step 1 ($P < .01$); and 10% higher probability in the Competent state lead to 4.3 points higher on Step 1 ($P < .01$). No one in this final group failed Step 1.

Prediction of USMLE Step 1 based on ten progress tests

The same analysis was replicated using all ten iterations of NBME tests. The transition probabilities of all students are displayed in Fig. 3. Compared with the transition probability based on the first six NBME tests, students had a lower probability of degrading from Competent state to Advanced Beginner II (8%) and a higher probability of staying in the Competent Learned State (92%). The predicted probability for each state at the end of the second year for all students was: 1.1% Novice, 8.9% Advanced Beginner I, 18.3% Advanced Beginner II, and 71.8% for Competent state.

Figure 4 shows the spaghetti plot of students' ten NBME tests by their predicted probability in the latent states. There were four students who had a greater than 50% chance of staying in the Novice state. Among these four students, two (50%) failed Step 1 and one (25%) barely passed (score = 195). Eighty-four students still had a high probability of being in the Novice or Advanced Beginner I state. Their mean Step 1 score was 209 (SD = 18.6), and 15 (18%) of them failed Step 1. One hundred and sixty-eight students had predicted probability to be in the Competent state. Their mean Step 1 score was 243.3 (SD = 11.6) and none of them failed Step 1.

These regression results are similar to those from our six test model: a 10% higher chance in the Novice state led to scoring 3.0 points lower on Step 1 ($P < .01$); 10% higher chance in Advanced Beginner I state led to scoring 0.6 points lower ($P = 0.11$); 10% higher chance in Advanced Beginner II state led to scoring 1.3 points higher ($P < .01$); and 10% higher chance in the Competent state led to scoring 3.4 points higher ($P < .01$).

Discussion

Though progress tests have been used for decades in medical education, most programs suffer from methodological limitations which limit generalizability to other contexts. In this study, we introduce an innovative assessment method to model students' progression in acquiring basic medical knowledge using NBME examinations administered as progress tests. The standard methodological framework for the study of intra-individual differences' change over time in continuously measured variables is growth curve modeling¹⁸⁻²². Growth curve modeling takes as its data source individual empirical growth trajectories. Growth curve modeling can provide an estimate of the average initial level and average rate of growth taken to be estimates of the growth parameters in a defined population. Compared to growth curve methods, Markov chain method not only considers the correlation between the previous and next test performance but also provides an estimated probability distribution of an individual's growth pattern which can be used to predict high-stake test performance, such as on Step 1.

Employing the Markov chain method, we can find each student's transition probability among four stages and their predicted probability in each stage. The regression results show that the students with higher probabilities of staying in the lowest state (Novice) will have significantly lower Step 1 performance, while those with higher probabilities for attaining the Competent state will have significant higher Step 1

results. Based on the transition probability, we can identify students with little growth during the first year of medical school (i.e. the first six examinations) and provide additional academic support to them to enhance their trajectory in the second year.

Our study has two important limitations. First, this study focused only on one medical school and two student cohorts within this medical school. Our school implemented a new competency-based, integrated curriculum in 2016 which may impact our students' trajectories for learning. Our entering students' characteristics, however, were similar to that for other US medical schools (Mean MCAT is 506.2 for these two cohort students). Our model of progress testing twice per semester is relatively unique, but could be readily implemented in other schools. Future studies are necessary to assess how well our findings might generalize to other student cohorts and medical schools. Second, we used two different NBME exam types for our progress testing. Blueprinting for content and overall item difficulty was similar, but not identical, between the two types. The NBME reports scaled scores for the CBSE, normed to a mean of 70 with a standard deviation of 8 for first time Step 1 takers. In contrast, CAS scores are reported as percentage correct. The two types of scores tracked quite well with one another, but they cannot be equated.

With these caveats in mind, this study significantly contributes to our understanding of progress testing in at least two areas. First, it has provided strong validity evidence for our internal assessment program and provides an estimation method for faculty members to gauge the progress of students and intervene with additional academic support as needed. Second, the analytical approach proposed in this study provides a flexible method by which medical knowledge growth can be categorized into latent states. It can provide useful information for medical educators interested in pursuing progress examinations, even after the USMLE Step 1 changes to Pass/Fail.

Declarations

Acknowledgement: None

Disclosure: None

Funding/Support: None

Other disclosures: None

Ethical approval:

Per request from the Michigan State University's Human Research Protection Program (MSU-HRPP), a designated honest broker is used to deidentify curricular and student evaluation data collected as a normal part of the medical school's educational programs. By the MSU-HRPP's determination, these data are not considered human subject data. Documentation concerning the honest broker program can be found at <https://omerad.msu.edu/research/honest-broker-for-educational-scholarship>.

Disclaimers: The authors alone are responsible for the content and writing of the article.

Previous presentations: The abstract of an earlier version of this article was accepted to present at American Educational Research Association (2020)

References

1. Dreyfuss SE, Dreyfus HL. A five-stage model of the mental activities involved in directed skill acquisition. *Oper Res Cent.* 1980. doi:ADA084551.
2. PIAGET J, INHELDER B. Diagnosis of mental operations and theory of the intelligence. *Am J Ment Defic.* 1947.
3. Kohlberg L. Stages of Moral Development. *Moral Educ.* 1971.
4. Kaplan D, Walpole S. A stage-sequential model of reading transitions: Evidence from the early childhood longitudinal study. *J Educ Psychol.* 2005. doi:10.1037/0022-0663.97.4.551.
5. Millward RB, Wickens TD. Models for Behavior: Stochastic Processes in Psychology. *Am J Psychol.* 2006. doi:10.2307/1422829.
6. Johnson TR, Khalil MK, Peppler RD, Davey DD, Kibble JD. Use of the NBME Comprehensive Basic Science Examination as a progress test in the preclerkship curriculum of a new medical school. *Adv Physiol Educ.* 2014. doi:10.1152/advan.00047.2014.
7. Morrison CA, Ross LP, Fogle T, Butler A, Miller J, Dillon GF. Relationship between performance on the NBME comprehensive basic sciences self-assessment and USMLE step 1 for U.S. and Canadian medical school students. *Acad Med.* 2010. doi:10.1097/ACM.0b013e3181ed3f5c.
8. Champlain AF, De, Cuddy MM, Scoles PV, et al. Progress testing in clinical science education: Results of a pilot project between the National Board of Medical Examiners and a US Medical Progress testing in clinical science education : Results of a pilot project between the National Board of Medical. 2010. doi:10.3109/01421590903514655.
9. Gagniuc PA. *Markov Chains: From Theory to Implementation and Experimentation.*; 2017. doi:10.1002/9781119387596.
10. Serfozo R. *Basics of Applied Stochastic Processes.*; 2009. doi:10.1007/978-3-540-89332-5.
11. Jung T, Wickrama KAS. An Introduction to Latent Class Growth Analysis and Growth Mixture Modeling. *Soc Personal Psychol Compass.* 2008. doi:10.1111/j.1751-9004.2007.00054.x.
12. Bartolucci F, Farcomeni A, Pennoni F. *Latent Markov Models for Longitudinal Data.*; 2012. doi:10.1201/b13246.
13. Allison PD. Multiple imputation for missing data: A cautionary tale. *Sociol Methods Res.* 2000. doi:10.1177/0049124100028003003.
14. Basco WT, Way DP, Gilbert GE, Hudson A. Undergraduate institutional MCAT scores as predictors of USMLE step 1 performance. *Acad Med.* 2002. doi:10.1097/00001888-200210001-00005.
15. Zhao X, Oppler S, Dunleavy D, Kroopnick M. Validity of four approaches of using repeaters' MCAT scores in medical school admissions to predict USMLE step 1 total scores. *Acad Med.* 2010. doi:10.1097/ACM.0b013e3181ed38fc.

16. Kuha JAIC. and BIC: Comparisons of assumptions and performance. *Sociol Methods Res.* 2004. doi:10.1177/0049124103262065.
17. Team RC. R: A Language and Environment for Statistical Computing. *R Found Stat Comput.* 2016.
18. Meredith W, Tisak J. Latent curve analysis. *Psychometrika.* 1990. doi:10.1007/BF02294746.
19. Muthén B. Latent Variable Analysis. In: *Handbook of Quantitative Methodology for Social Sciences;* 2004.
20. Rogosa D, Brandt D, Zimowski M. A growth curve approach to the measurement of change. *Psychol Bull.* 1982. doi:10.1037/0033-2909.92.3.726.
21. Singer JD, Willett JB. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence;* 2009. doi:10.1093/acprof:oso/9780195152968.001.0001.
22. Willett JB, Sayer AG. Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychol Bull.* 1994. doi:10.1037/0033-2909.116.2.363.

Figures

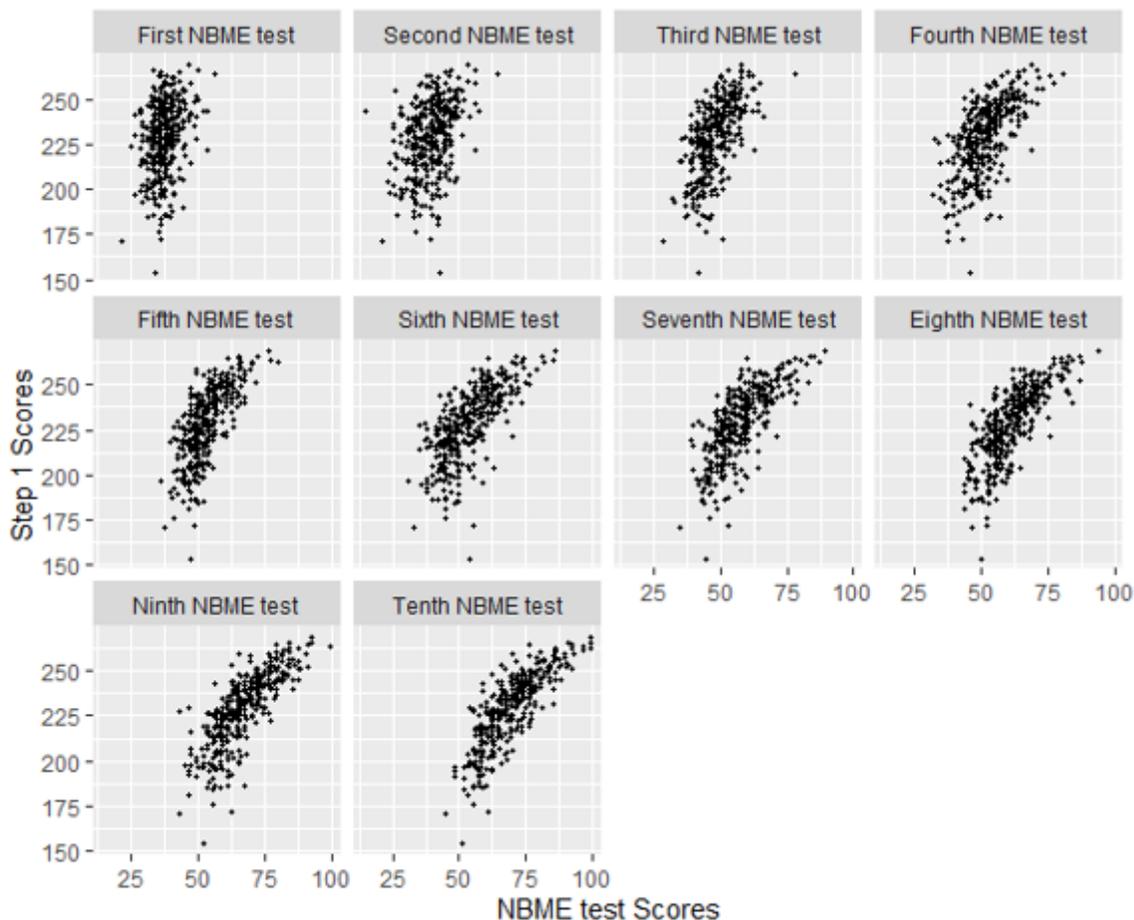


Figure 1

Scatter plots for NBME test versus Step 1 results