World Journal of Educational Studies
Print ISSN: 2959-9989
Online ISSN: 2959-9997
DOI: https://doi.org/10.61784/wjes3195

PRACTICAL APPLICATION OF THE STATIONARY DISTRIBUTION OF MARKOV CHAINS IN TEACHING-EFFECT EVALUATION

Jun Fu
School of Financial Management, Chengdu Ginkgo Hotel Management College, Chengdu 611743, Sichuan, China.

Abstract: To explore the application of Markov-chain models in process-oriented teaching evaluation, this study analyzes grade transitions among students in two Advanced Mathematics classes. A total of 80 students were included, and their scores from two stage assessments were classified into five levels: excellent, good, medium, pass, and fail. Based on the paired assessment records, transition-frequency tables and transition-probability matrices were constructed for the two classes. The short-term grade distributions and stationary distributions were subsequently calculated to examine the dynamic characteristics and potential evolution of students' learning states. The results show that the two classes exhibited different grade-transition patterns. Class A demonstrated a relatively concentrated and stable transition structure, with most students tending toward the good and medium levels and a comparatively low proportion remaining at the failing level. Class B maintained a relatively prominent excellent group, but its lower-performing students showed greater dispersion and a stronger tendency to remain in or move toward the lower grade levels. The Markov-chain model integrates individual grade transitions with class-level score distributions, enabling a quantitative and visual description of changes in students' learning states. This approach provides a methodological reference for stage-based learning assessment, differentiated instructional intervention, and data-informed optimization of Advanced Mathematics teaching.
Keywords: Markov chain; Stationary distribution; Grade transition; Learning analytics; Teaching evaluation

1 INTRODUCTION

The learning and development of university students have long received extensive attention. As smart education continues to deepen, effectively improving student learning is inseparable from research on teaching effectiveness. Evaluation should not focus solely on final examination scores; it should also consider effects during the learning process. Changes in test performance during learning can influence final achievement. Introducing a Markov-chain model enables observation of the actual effect of instruction throughout the teaching process. The model relates changes in student performance to stage-based learning and uses process orientation to explain final learning outcomes, thereby supporting a more objective evaluation of teaching and adjustment of difficult or key teaching content.

A single final score records only the result reached at one point in time and cannot directly describe how a student moved from one performance level to another. By contrast, two stage tests provide information about both the initial distribution and subsequent transitions. When these transitions are organized as probabilities, the teacher can distinguish stable performance, upward movement, and downward movement among different student groups. The resulting process-based description is consistent with the original aim of observing teaching effects dynamically rather than judging instruction only from an endpoint.

The practical value of this approach lies in connecting class-level evaluation with individual learning changes. Students who begin at the same grade do not necessarily follow the same path after instruction. Some remain at their original level, some improve, and others decline. A transition matrix retains these differences while summarizing them in a form that can be compared across classes. The stationary distribution then extends the observed transition pattern to a long-term structure, providing an interpretable reference for identifying teaching strengths and groups requiring additional support.

Learning analytics and educational data mining increasingly use student behavior and performance data to generate actionable evidence for teaching and learning [1]. Their role in higher-education assessment has expanded with the availability of large and longitudinal datasets [2]. Explainable analytics can identify disengaged students and support timely intervention [3], while meta-analytic evidence indicates that learning-analytics interventions can improve learning achievement [4]. Sequential process-mining research further shows that high- and low-performing groups can exhibit different regulatory-transition patterns [5]. Recent studies have predicted online performance from multidimensional time-series data [6], examined fairness in student-success prediction [7], connected prescriptive analytics with institutional planning [8], developed explainable rule-based performance models [9], and applied deep ensemble learning to academic-performance prediction [10]. Together, these studies support process-oriented, transparent, and data-based evaluation of changes in student learning states.

© By the Author(s) 2026, under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0).

2 OVERVIEW OF THE MARKOV-CHAIN MODEL

2.1 Definition of a Markov Chain

Let $\{X_t|t = 0,1,2, \ldots\}$ be the value set of a specified random variable. Its state space S consists of performance grades such as A, B, C, D, and E. If, for any $t \geq 0$ and any $j, i_0, i_1, \ldots, i_t \in S$, the following condition holds:

$$P(X_{t+1} = j|X_t = i_t, X_{t-1} = i_{t-1}, \ldots, X_0 = i_0) = P(X_{t+1} = j|X_t = i_t) \quad (1)$$

Then $\{X_t\}$ is called a Markov chain. The next state $X_{t+1} = j$ depends only on the current state $X_t = i_t$, and is independent of earlier states. This is the memoryless property. In the study of learning effects, the next test score is related only to the current test score and not to the student's initial learning condition, providing a basis for subsequent teaching improvement.

In this application, the memoryless property is a modeling description of score transitions. It does not deny the existence of earlier learning experiences; instead, it states that the current grade summarizes the information used to predict the next grade. Once a student is classified into A, B, C, D, or E at the current stage, the probability of the next outcome is determined from that current state. This assumption makes it possible to use repeated tests as consecutive observations of one learning process and to express teaching effects through changes in state probabilities.

The five grades also give the state space a direct educational interpretation. Remaining on the same grade corresponds to stability, movement toward A represents improvement, and movement toward E indicates deterioration. Because every student occupies one and only one state in each test, the model can aggregate individual trajectories without losing the direction of change. The same framework can therefore describe both the overall class distribution and the transition behavior of students with different initial foundations.

2.2 Transition Probabilities and Transition Matrix

The transition probability p_{ij} denotes the probability that a student's grade changes from level i in one test to level j in the next test; that is, $p_{ij} = P(X_{n+1} = j|X_n = i)$. It satisfies two conditions: $p_{ij} \geq 0$ and, for every level i, $\sum_{j\in S} p_{ij} = 1$.

All transition probabilities are arranged with the current grade as the row and the next grade as the column. If the state space is $S = \{s_1, s_2, \ldots, s_m\}$, where m is the number of grade levels, then P is an $m \times m$ matrix:

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{pmatrix} \quad (2)$$

Each row contains the probabilities of transition from one initial grade to all target grades and directly reflects how teaching intervention affects students with different initial foundations.

The row-sum condition ensures that every possible destination from one initial grade is included. Diagonal entries describe the probability that students remain at their current level. Entries to the left of the diagonal, under the A-to-E ordering used here, describe movement toward higher grades, whereas entries to the right describe movement toward lower grades. Reading a matrix by rows therefore reveals which initial groups respond steadily to instruction and which groups show greater dispersion after teaching.

This interpretation is especially useful when two classes have similar overall averages but different internal movements. A class may obtain a reasonable second-test distribution because high-performing students remain strong, while some lower-performing students continue to struggle. Another class may show fewer excellent students but more concentrated movement into the good range. Transition probabilities make these patterns visible and prevent the overall score from concealing differences among initial grade groups.

2.3 Stationary Distribution

Let $\pi_0 = (\pi_0(1), \pi_0(2), \ldots, \pi_0(m))$ denote the grade distribution in the initial test at $t = 0$. Here $\pi_0(i) = P(X_0 = i)$ is the proportion of students in grade i initially.

$$\sum_{i\in S} \pi_0(i) = 1, \pi_n = \pi_{n-1}P, \pi = \pi P \quad (3)$$

The recurrence $\pi_n = \pi_{n-1}P$ gives future state distributions. As time tends to infinity, π_n converges to a stationary distribution π, where $\pi = \pi P$. At that point, transition behavior remains stable and supports prediction of process evolution.

The short-term distribution answers how the observed class structure changes after one transition. Repeated multiplication by the same transition matrix describes the distribution after additional stages under an unchanged transition pattern. When the sequence converges, the stationary distribution no longer changes after another multiplication by the matrix. Its components can then be interpreted as the long-term proportions expected at grades A through E under the transition structure estimated from the two tests.

The stationary distribution is not used to replace the observed test results. It complements them by showing the direction implied by the current pattern. Comparison between the initial, short-term, and stationary distributions separates immediate improvement from long-term balance. In teaching evaluation, this distinction helps identify whether a favorable second test reflects a stable transition mechanism or only a temporary concentration of scores.

3 APPLICATION OF THE MARKOV-CHAIN MODEL TO TEACHING EFFECTIVENESS

3.1 Selection of Research Subjects and Grade Classification

Two classes taught by the author were selected: Class 1 of the 2025 Engineering Cost major, referred to as Class A, and Class 2, referred to as Class B. Each class contained 40 students. The subject was Advanced Mathematics. Scores from two tests were recorded, covering modules such as functions and limits and derivatives and differentials. Scores were divided into five levels: A (90-100, excellent), B (80-89, good), C (70-79, medium), D (60-69, pass), and E (0-59, fail). The first test was treated as the initial state, and the number and proportion of students at each level were calculated.

Using the same class size and the same five grade intervals makes the two initial distributions directly comparable. Class A begins with 40% of students at B and 30% at C, while Class B begins with 25% at B and 37.5% at C. Class A has four students at E and Class B has two, but Class B has a larger D group. These differences establish the baseline from which the subsequent transitions and teaching effects are evaluated.

$$\pi_A(0) = \left(\frac{4}{40}, \frac{16}{40}, \frac{12}{40}, \frac{4}{40}, \frac{4}{40}\right) = (0.1, 0.4, 0.3, 0.1, 0.1) \tag{4}$$

$$\pi_B(0) = \left(\frac{5}{40}, \frac{10}{40}, \frac{15}{40}, \frac{8}{40}, \frac{2}{40}\right) = (0.125, 0.25, 0.375, 0.2, 0.05) \tag{5}$$

The corresponding student counts for the initial distributions are reported in Table 1.

Table 1 Distribution of First-Test Scores in the Two Classes

Class	A (Excellent)	B (Good)	C (Medium)	D (Pass)	E (Fail)
Class A (40 students)	4	16	12	4	4
Class B (40 students)	5	10	15	8	2

3.2 Construction of Grade-Transition Matrices

The transition matrix is the core of the Markov chain and describes the probability distribution from the current state to the next state. After teaching the functions and limits module, a second test was administered and scores for both classes were counted in Table 2. The grade reached by each student in the second test was then traced from the student's first-test grade to form the transition-frequency tables in Tables 3 and 4. Transition probabilities were calculated as the transition frequency divided by the total number of students in the initial grade.

Table 2 Distribution of Second-Test Scores in the Two Classes

Class	A (Excellent)	B (Good)	C (Medium)	D (Pass)	E (Fail)
Class A (40 students)	2	20	10	6	2
Class B (40 students)	4	15	12	5	4

Table 3 Grade Transitions for Class A

First-Test Grade	A	B	C	D	E	Total
A	1	2	1	0	0	4
B	1	15	0	0	0	16
C	0	3	8	1	0	12
D	0	0	1	2	1	4
E	0	0	0	3	1	4

Table 4 Grade Transitions for Class B

First-Test Grade	A	B	C	D	E	Total
A	4	1	0	0	0	5
B	0	8	2	0	0	10
C	0	4	9	2	0	15
D	0	2	1	2	3	8
E	0	0	0	1	1	2

The transition-frequency tables can be checked in two directions. The row totals reproduce the first-test grade counts because every student in an initial grade must move to exactly one second-test grade. In Table 3, the row totals 4, 16, 12, 4, and 4 match the Class A counts in Table 1. In Table 4, the totals 5, 10, 15, 8, and 2 match the Class B initial counts. This confirms that all 40 students in each class are represented once in the transition process.

The column totals reproduce the second-test distributions. For Class A, summing each destination column in Table 3 gives 2, 20, 10, 6, and 2, exactly matching Table 2. For Class B, the corresponding column sums are 4, 15, 12, 5, and 4.

This consistency links the individual transition records to the class-level distributions and shows that the probability matrices are constructed from the same observed score data rather than from separate estimates.
For Class A, representative probabilities are calculated as follows; transitions from A to D and from A to E both equal zero. Other probabilities and the Class B values are obtained in the same manner.

$$p_A(A \to A) = \frac{1}{4} = 0.25, p_A(A \to B) = \frac{2}{4} = 0.5, p_A(A \to C) = \frac{1}{4} = 0.25 \tag{6}$$

$$P_A = \begin{pmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 \\ \frac{1}{16} & \frac{15}{16} & 0 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{2}{3} & \frac{1}{12} & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 0 & \frac{3}{4} & \frac{1}{4} \end{pmatrix} \tag{7}$$

$$P_B = \begin{pmatrix} \frac{4}{5} & \frac{1}{5} & 0 & 0 & 0 \\ 0 & \frac{4}{5} & \frac{1}{5} & 0 & 0 \\ 0 & \frac{4}{15} & \frac{9}{15} & \frac{2}{15} & 0 \\ 0 & \frac{1}{4} & \frac{1}{8} & \frac{1}{4} & \frac{3}{8} \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} \tag{8}$$

The Class A matrix shows that one quarter of the students initially at A remain at A, one half move to B, and one quarter move to C. Among the 16 students initially at B, 15 remain at B and one moves to A, producing a strong diagonal probability of 15/16. For the students initially at C, two thirds remain at C, one quarter move to B, and one twelfth move to D. The D and E rows show that movement among the lower grades is more dispersed: students initially at E mainly move to D, while one quarter remain at E.
The Class B matrix presents a different pattern. Four fifths of the students initially at A remain at A, and one fifth move to B. Students initially at B show the same four-fifths stability, with one fifth moving to C. The C group is distributed across B, C, and D, with probabilities 4/15, 9/15, and 2/15. The D row contains transitions to B, C, D, and E, and three eighths remain or move into E. Of the two students initially at E, one moves to D and one remains at E. These entries explain why Class B retains a stronger excellent group while still facing difficulty at the lower end.
The contrast between the two matrices is more informative than comparing only the second-test totals. Class A has a highly concentrated B group and substantial upward movement from E to D. Class B preserves more of its A group, but the D and E rows indicate a greater probability of remaining in or moving toward the failing level. The matrices therefore support different teaching priorities for the two classes even though both contain the same number of students and are evaluated with the same grade scale.
Converting frequencies into probabilities also removes the effect of different initial group sizes. For example, 15 Class A students remain at B, which is a large count because the initial B group contains 16 students; the probability 15/16 expresses this stability directly. By comparison, four Class B students remain at A out of an initial group of five, giving 4/5. Probability form makes these two patterns comparable and clarifies whether a result reflects a large initial group or a genuinely high transition tendency.

3.3 Prediction and Analysis of Teaching Effectiveness

Short-term prediction uses the initial distribution and transition matrix. The recurrence $\pi_n = \pi_{n-1}P$ gives the score distribution in the second test at $n = 1$ and reflects teaching effectiveness in the middle of the semester.

$$\pi_n = \pi_{n-1}P \tag{9}$$

$$\pi_A(1) = \pi_A(0)P_A = \left(\frac{1}{20}, \frac{1}{2}, \frac{1}{4}, \frac{3}{20}, \frac{1}{20}\right) \tag{10}$$

For Class A, the resulting proportions are 5% at level A, 50% at B, 25% at C, 15% at D, and 5% at E. The proportion at B increases by 10 percentage points from the initial distribution, and 80% of students reach level C or above.
The short-term Class A distribution indicates that the main concentration shifts toward B. Although the proportion at A decreases relative to the initial 10%, the combined share at B and C reaches 75%, and the failing share falls from 10% to 5%. The movement is therefore characterized by greater concentration in the middle-to-upper grades and a reduction in the lowest grade. This supports the original assessment that the teaching effect in Class A is stable and balanced rather than dependent on a small excellent group.

$$\pi_B(1) = \pi_B(0)P_B = \left(\frac{1}{10}, \frac{3}{8}, \frac{3}{10}, \frac{1}{8}, \frac{1}{10}\right) \tag{11}$$

For Class B, the proportions are 10% at level A, 37.5% at B, 30% at C, 12.5% at D, and 10% at E. The proportion at A is higher than in Class A, but the proportion at E is twice that of Class A.

Class B also shows short-term improvement in the middle range because the share at B rises from 25% to 37.5%, while the share at D falls from 20% to 12.5%. At the same time, the share at E rises from 5% to 10%. The coexistence of a relatively strong A group and an enlarged E group suggests a more polarized result. Teaching adjustment for this class should therefore retain support for high-performing students while directing additional attention to students who remain at or move into the failing grade.

The short-term vectors can also be verified against the second-test counts. Multiplying the Class A proportions by 40 gives 2, 20, 10, 6, and 2 students, while multiplying the Class B proportions by 40 gives 4, 15, 12, 5, and 4. The equality between the model output and Table 2 demonstrates that one-step matrix multiplication correctly reproduces the observed second test. This verification is important before the same matrices are used to describe longer-term tendencies.

For long-term stationary-state prediction, let the Class A stationary distribution be $\pi_A = (\pi_1, \pi_2, \pi_3, \pi_4, \pi_5)$. Substituting $\pi_A = \pi_A P_A$ and $\pi_1 + \pi_2 + \pi_3 + \pi_4 + \pi_5 = 1$ produces a linear system.

$$\pi_A = \pi_A P_A, \pi_1 + \pi_2 + \pi_3 + \pi_4 + \pi_5 = 1 \tag{12}$$

$$\pi_A = (0.058, 0.682, 0.185, 0.055, 0.020) \tag{13}$$

$$\pi_B = (0.092, 0.456, 0.288, 0.124, 0.040) \tag{14}$$

Taking 95, 85, 75, 65, and 55 as representative evaluation scores for the five grade intervals, the comprehensive indicators are calculated as follows.

$$I_A = 0.058 \times 95 + 0.682 \times 85 + 0.185 \times 75 + 0.055 \times 65 + 0.020 \times 55 \approx 82.3 \tag{15}$$

$$I_B = 0.092 \times 95 + 0.456 \times 85 + 0.288 \times 75 + 0.124 \times 65 + 0.040 \times 55 \approx 80.1 \tag{16}$$

In the long run, Class A forms a stable structure dominated by level B, which accounts for 68.2%, while level E accounts for only 2%. Long-term teaching therefore places the class as a whole at a stable medium-to-high level. Class B has a higher proportion of level A students than Class A, at 9.2% versus 5.8%, but its level E proportion is also higher, at 4%. This indicates a more prominent high-achieving group in Class B, together with a need to strengthen the improvement of lower-achieving students.

The comprehensive indicators summarize the five-component stationary distributions with representative scores for the grade intervals. Class A reaches approximately 82.3, compared with 80.1 for Class B. This difference is consistent with the stationary structures: Class A has a much larger B component and a smaller E component, whereas Class B has a larger A component but also greater proportions at D and E. The weighted indicator should therefore be interpreted together with the complete distribution rather than as a substitute for it.

The two-point difference between the comprehensive indicators does not by itself identify where teaching differs. The stationary vectors provide that explanation. Class A places 86.7% of its long-term mass in B and C and only 7.5% in D and E. Class B places 74.4% in B and C and 16.4% in D and E. Class B's larger A share partly offsets its larger lower-grade shares in the weighted score, but the full distribution reveals the continued need to improve the D and E groups.

From a teaching perspective, Class A requires continued consolidation of the B group and support for movement from C toward B and A. Its small stationary E share indicates that the existing approach is comparatively effective in controlling persistent failure. Class B requires a more differentiated strategy. The stable A group can be challenged with higher-level work, while students in D and E need targeted review, additional exercises on difficult modules, and closer stage-based monitoring. These recommendations follow directly from the observed transitions and the stationary proportions.

The same calculations can be repeated after later modules. A new test can replace the second test as the current state, and its transitions can be compared with the matrices reported here. If the probability of moving from E to D or C rises, targeted support is producing an observable effect. If the probability of remaining at E remains high, the relevant teaching content and practice arrangements require further adjustment. This creates a direct link between stage testing, model updating, and instructional response.

4 CONCLUSION

Using Class 1 and Class 2 of the Engineering Cost major as examples, this study evaluates teaching effectiveness from student test scores and applies a Markov-chain model to quantify two test results. The results show that Class A has stable and balanced teaching effects. Its long-term distribution is dominated by level B, and the proportion of failing students at level E is only 2%, indicating effective control of lower-achieving students. Although Class B has a higher proportion of level A students and an advantage in cultivating excellent students, its level E proportion reaches 4%, revealing a weakness in the improvement of lower-achieving students and a need for targeted enhancement. The model overcomes the limitation of traditional evaluation that emphasizes outcomes while neglecting process, enables dynamic monitoring and prediction of teaching effectiveness, and provides precise data support for optimizing Advanced Mathematics teaching strategies, strengthening support for lower-achieving students, and improving overall teaching quality.

The application also shows that teaching-effect evaluation benefits from examining several levels of information together. The frequency tables record actual student movements, the transition matrices normalize these movements into comparable probabilities, the short-term distributions reproduce the immediate class structures, and the stationary distributions reveal the long-term tendency implied by the same transition mechanism. Their combined interpretation makes the evaluation transparent and keeps each conclusion connected to observable score changes.

The model is based on the two tests and five grade intervals used in the original teaching case. Its predictions describe the continuation of the estimated transition pattern. In subsequent teaching practice, new stage-test results can be incorporated to update the transition probabilities and compare whether targeted instruction changes the movement of students in the lower grades. In this way, the Markov-chain framework can support a continuing cycle of observation, adjustment, and reevaluation while preserving the process-oriented logic of the study.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Lemay D J, Baek C, Doleck T. Comparison of learning analytics and educational data mining: A topic modeling approach. Computers and Education: Artificial Intelligence, 2021, 2: 100016.
[2] Caspari-Sadeghi S. Learning assessment in the age of big data: Learning analytics in higher education. Cogent Education, 2023, 10(1): 2162697.
[3] Linden K, van der Ploeg N, Roman N. Explainable learning analytics to identify disengaged students early in semester: An intervention supporting widening participation. Journal of Higher Education Policy and Management, 2023, 45(6): 626-640.
[4] Zheng L Q, Fan Y C, Gao L, et al. The impact of learning analytics interventions on learning achievements: A meta-analysis of research from 2012 to 2021. Interactive Learning Environments, 2024, 32(10): 6267-6282.
[5] Channa F, Dindar M, Nguyen A, et al. Exploring sequential interplay between challenges and regulatory processes in collaborative learning with process mining. Scandinavian Journal of Educational Research, 2024, 68(6): 1320-1342.
[6] Shou Z Y, Xie M Q, Mo J W, et al. Predicting student performance in online learning: A multidimensional time-series data analysis approach. Applied Sciences, 2024, 14(6): 2522.
[7] Nezami N, Haghighat P, Gandara D, et al. Assessing disparities in predictive modeling outcomes for college student success: The impact of imputation techniques on model performance and fairness. Education Sciences, 2024, 14(2): 136.
[8] Manly C A. Connecting prescriptive analytics with student success: Evaluating institutional promise and planning. Education Sciences, 2024, 14(4): 413.
[9] Mai J H, Wei F X, He W, et al. An explainable student performance prediction method based on dual-level progressive classification belief rule base. Electronics, 2024, 13(22): 4358.
[10] Tang B, Li S L, Zhao C H. Predicting the performance of students using deep ensemble learning. Journal of Intelligence, 2024, 12(12): 124.