Research Article Open Access

Application of Markov Chains to Analyze and Predict the Mathematical Achievement Gap between African American and White American Students

Vivian R Moody* and Kanita K DuCloux

Department of Mathematics, Western Kentucky University, Bowling Green, KY, USA

Abstract

A stochastic, mathematical model known as a discrete Markov Chain was used to show how to estimate the probability that the mathematical achievement gap between African Americans and White Americans would close during a particular calendar year. The implications of race in the achievement of mathematics in the United States are profound and well-documented in mathematics education research literature. The authors used historical data drawn from the National Assessment of Educational Progress (NAEP) to examine trends of mathematical achievement between African Americans and White Americans during the assessment years of 1973–2012. The authors provide a discussion of NAEP data in the context of the discrete Markov Chain model and describe how specific properties of the Markov process were used to estimate the probability that the mathematical achievement gap will close within the next 50 years.

Keywords: Markov chains; Transition matrix; Probabilities; Mathematics achievement gap

Introduction

Mathematical achievement gap

Using data drawn from the National Assessment of Educational Progress (NAEP) much research [1-4] have reported the alarming underachievement in mathematics of African American youths, ages 9, 13 and 17 as compared to their White counterparts. Reform efforts ([5-8], from the National Council of Teachers' of Mathematics) in mathematics education in the United States have called for an agenda of mathematics teaching and learning that emphasizes students' engagement in the processes of critical mathematical thinking and problem solving analogous to that of mathematicians [9]. However, despite these efforts, African American students continue to exhibit underachievement and underrepresentation in mathematics.

What has not been delineated in educational policy or reform efforts in mathematics is a clear objective for the mathematical achievement of African American students. For example, public policy in the United States has not determined specifically how African Americans should achieve mathematically compared to their White counterparts and by what date or when should this achievement be realized. To add to the literature on the mathematical achievement of African American students in this regard, this research examined the trends of mathematical achievement between African American and White American students over the past forty years. The authors' goal was to generate a probability statistic that would predict the future achievement of African Americans given historical, mathematical achievement gains and declines. Specifically, this study sought to estimate the probability that the mathematical achievement gap that has historically existed between African Americans and White Americans will close within the next 50 years.

Mathematical model

In 1907, Andrei Markov, a Russian mathematician, began the study of an important new type of chance process. In this process, the outcome of a given experiment can affect the outcome of the next experiment.

This type of process is called a Markov Process or Markov Chain [10]. A Markov-Chain process is a stochastic, mathematical model with transition probabilities (defined below) that provide information about how to relate one stage of a process to the next [10].

To create a mathematical model for this research, the authors used a discrete Markov Chain. As delineated in the work of Kemeny and Snell [11], a discrete or finite Markov chain is a stochastic process with finitely many states on a nominal scale. Moreover, the stochastic process is a random process evolving in time. Since this research sought to determine the probability that the mathematical achievement gap between African American and White American students will close in a finite number of years, a discrete Markov chain was appropriate in modeling this phenomenon. Further, mathematical achievement gap scores examined during discrete times in history lend themselves better to stochastic modeling rather than deterministic modeling in which the impending state of the process completely depends on the past and the present states of the process. The authors used the following properties of a discrete Markov process in developing the framework of the stochastic model:

- The number of possible outcomes or states is finite.
- The outcome at any stage depends only on the outcome of the previous stage.
 - The probabilities are constant over time.

*Corresponding author: Vivian R Moody, Department of Mathematics, Western Kentucky University, Bowling Green, KY, USA, Tel: 1 270 745 6209, E-mail: vivian.moody@wku.edu

Received March 12, 2014; Accepted April 07, 2014; Published April 14, 2014

Citation: Moody VR, DuCloux KK (2014) Application of Markov Chains to Analyze and Predict the Mathematical Achievement Gap between African American and White American Students. J Appl Computat Math 3: 161 doi:10.4172/2168-9679.1000161

Copyright: © 2014 Moody VR, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

J Appl Computat Math ISSN: 2168-9679 JACM, an open access journal

Mathematically, we describe a discrete Markov chain with the following formula:

$$M\{S^{(t+1)} = j | S^{(t)} = i, S^{(t-1)} = i_{t-1,...}\} = M\{S^{(t+1)} = j | S^{(t)} = i\} = m_{ii}(t)$$
 [12]

In the formula, t denotes discrete time. In this paper, we define t in terms of years. As denoted in the above formula, we have a set of states $S^{(t)} = \{s_{1t}, s_{2t}, \dots s_{nt}\}$. The process starts in one of these states and moves successively from one state to another. Each move is called a step. If the chain is currently in state s_{it} , then it moves to state s_{ij} at the next step with a probability denoted by $m_{ij}(t)$. This probability does not depend upon which states the chain was in before the current state.

The probabilities $m_{ij}(t)$ are called transition probabilities. The matrix $M_i = \left| m_{ij}(t) \right|$ is a one-step transition matrix [12]. Henceforth in this paper, M denotes the transition matrix for our discrete Markov process, such that $\sum_{j}^{n} m_{ij} = 1$ and $m_{ij} \geq 0$ for all i and j [13]. Each entry $m_{ij} \in M$ is defined as the probability of transitioning (moving) from state i to state j. Further, we denote the states: $\{s_1, s_2, \dots s_n\}$. Thus, m_{ij} is the probability that an object in state s_j transitions to state s_j [14].

Further describing our mathematical model, we let M be the transition matrix for our discrete Markov process such that M^k has only positive entries for some k. Then there exists a steady-state probability x_s vector such that $Mx_s = x_s$ [10]. Moreover, $\lim_{k \to \infty} M^k x_0 = x_s$ where x_0 is an initial-state probability vector [15]. The mathematical representation of our steady-state vector can be denoted with the following chain of equivalences:

$$Mx_s = x_s \Leftrightarrow Mx_s - x_s = 0 \Leftrightarrow Mx_s - Ix_s = 0 \Leftrightarrow (M-I)x_s = 0 \Leftrightarrow x_s \in N(M-I)$$

Thus, the steady-state vector x_s is in the null space of M–I. Since M^k has only positive entries for some k, then [dim (N (M-I))=1]. Hence,

any vector in N(M - 1) is a scalar multiple of x_s . Specifically, if $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ is any non-zero vector in N (*M*–*I*), then

$$x_s = \frac{1}{z}x$$
 where or we write $x_s = \frac{x}{\sum_{k=1}^{n} x}$ [15].

Hence, in our discrete Markov process, the steady-state vector x_s is defined as the limiting vector or eigenvector of the transition matrix M, and corresponds to the eigenvalue of 1.

Data

The National Assessment of Educational Progress (NAEP) is the largest nationally representative of what students in the United States know and can do in mathematics, as well as other subject areas. NAEP assessments are conducted periodically during particular calendar years and are administered uniformly across the United States. NAEP assessments are administered to US students at ages 9, 13, and 17. Results from NAEP data show the average achievement score at each age level and for various demographic groups based on descriptors such as race, socioeconomic status, and gender.

Provided in Figure 1 are NAEP data trends in mathematics achievement for African American and White American 9-year olds. The data reveal a wide achievement gap during the 1970s and early 1980s. However, the gap narrows during the 1990s but widens again during the late 2000s.

Similarly, as seen in Figures 2 and 3 for youths ages 13 and 17, although there is a persistent achievement gap in mathematics between African Americans and White Americans, the gap has widen and narrowed during particular time periods.

Results

To analyze the NAEP data, the authors developed tables of the mathematical achievement gaps between African American and White American youths at ages 9, 13, and 17 during the assessment years 1973–2012. To create our mathematical model, we defined g to be the random variable representing the score gaps between African American and White American students. As posited in the work of Anderson and Goodman [13], our model is created under the assumption that g can be grouped into classes that we define as states in our discrete Markov chain. Each of these classes or states represents the range of potential values for g. Thus, using this assumption we developed ranges of score gaps, during particular time periods. Given this analysis, we found it necessary to generate three separate probability statistics based on the

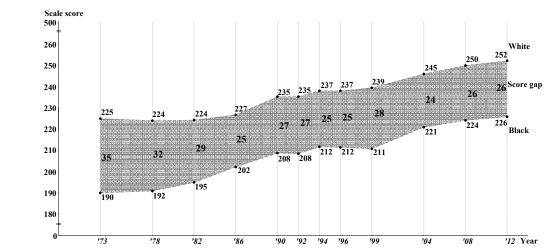


Figure 1: Trend in NAEP mathematics average scores and score gaps for African American and White American 9-year-old

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), various years, 1973-2012 Long-Term Trend Mathematics Assessments.

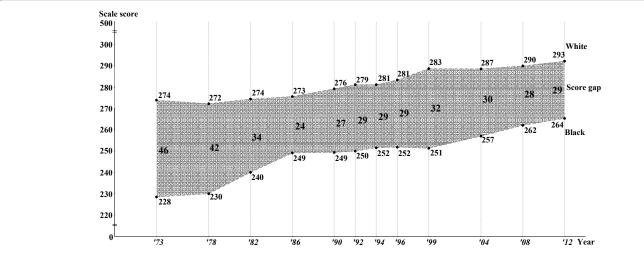


Figure 2: Trend in NAEP mathematics average scores and score gaps for African American and White American 13-year-old students

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), various years, 1973-2012 Long-Term Trend Mathematics Assessments.

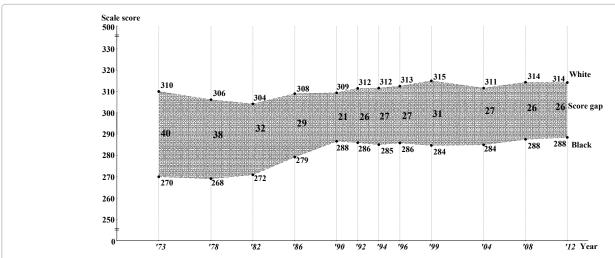


Figure 3: Trend in NAEP mathematics average scores and score gaps for African American and White American 17-year-old students

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), various years, 1973-2012 Long-Term Trend Mathematics Assessments.

mathematical achievement gap at each age level-9, 13, and 17. Thus, we find a transition matrix for each age level-9, 13, and 17-based on our analysis of the score gaps for each of these groups.

As described above, one of the properties of the discrete Markov process is that the set of states S_{ij} are finite. Therefore, we first examined the mathematical achievement gaps at age 9 and assigned each set of gap ranges to a corresponding transition state S_{ij} to create the discrete Markov-Chain model (Tables 1 and 2).

As shown in Table 2, although smaller or larger intervals of the movement of g could be created, we create four states $\{s_1, s_2, s_3, s_4\}$ since gap scores have historically moved within these ranges. We now define x_{ij} as the number of times g transitions from class i to class j [13]. Thus, we count the number of transitions or movements of g in Table 1 between the classes or states in Table 2. Based on our

analysis, we describe the changes between the score gaps as a stochastic process. Hence, this stochastic process of counting the movements of *g* result in the relative frequency of times *g* began in a particular state and transitioned to each of the other states. Therefore, we define

the transition probability
$$m_{ij} = \frac{x_{ij}}{\sum_{n=1}^{4} x_{ij}}$$
. We further describe m_{ij} as

the conditional probability of transitioning into one state, given the immediately preceding state. Hence, the outcome at any stage depends only on the outcome of the previous stage.

To complete the development of our discrete Markov-chain model, we must establish the dependence of against the test hypothesis that *g* is statistically independent. According to Billingsley [16], the Chi-Square test provides a systematic way of addressing the statistical analysis of Markov chains. Therefore, we used a Chi-Square test of

NAEP Assessment Year	Score gap (g)
1973	35
1978	32
1982	29
1986	25
1990	27
1992	27
1994	25
1996	25
1999	28
2004	24
2008	26
2012	26

Table 1: Mathematical achievement score gaps (g) between African American and White American students at Age 9 during the NAEP Assessment Years 1973-2012.

S_1	$23 \le g \le 26$
s_2	$26 \le g \le 29$
s_3	$29 \le g \le 32$
S_4	$32 \le g \le 35$

Table 2: Transition states and g classes for Age 9 students.

the data in Table 1, specifically treating a matrix of transition counts as if it were a contingency table [13]. We found that the result of the Chi-Square test was 13.063 (p=0.042) and was significant at the 95.800 percent confidence level. Thus, based on these results we proceeded in creating our discrete Markov-Chain model for age 9.

Given that the score gaps described above transition through four states $\{S_1, S_2, S_3, S_4\}$ for 9-year old African American and White American students during the years 1973-2012, we define our transition matrix of the Markov process for this age level by determining the initial class in which g existed and count g's movement from this class or state to each of the other classes or states. Mathematically, we describe this process as:

$$M=m_{ij}=Mx_s=x_s=rac{x}{\displaystyle\sum_{n=1}^4 x}$$
 . Thus, we have

$$M = m_{ij} = Mx_s = x_s = \frac{x}{\frac{4}{4}} \quad \text{. Thus, we have}$$

$$M = \begin{pmatrix} \frac{3}{5} & \frac{2}{5} & 0 & 0\\ \frac{3}{4} & \frac{1}{4} & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0 \end{pmatrix}, \text{ where the entries in } M \text{ are the transition}$$

probabilities in our discrete Markov-Chain model.

Further, in our discrete Markov chain, the sequence of all successive vectors x_n is linked by our transition matrix, M. Eventually, the successive probabilities stabilize or reach an equilibrium state, and converge over time [10]. To generate our probability statistic to predict whether the mathematical achievement gap will close within the next 50 years for students at age 9, we first find our transition matrix after a

¹Mathematica is a computational software program based on symbolic mathematics.

particular number of steps less than fifty. Choosing ten years (or steps) and performing the power calculations of M with Mathematica1, we have:

$$M^{10} = \begin{pmatrix} 0.652174 & 0.347826 & 0 & 0 \\ 0.652174 & 0.347826 & 0 & 0 \\ 0.652174 & 0.347826 & 0 & 0 \\ 0.652174 & 0.347826 & 0 & 0 \end{pmatrix}$$

We see that after ten years (or steps) the transition matrix reaches an equilibrium state or stabilizes, in which all rows of the matrix are the same. Thus, this matrix is our steady-state matrix, and we conclude that based on our Markov process, the achievement gap will likely close within the next 10 years for the 9-year old age level.

Note that since our data analysis ends at year 2012, our ten-year prediction includes the years 2013 and 2014. Hence, it is important to investigate whether the steady-state matrix can be determined in less than ten years (or steps). We find that after nine years (or steps) our transition matrix for this age level is as follows:

$$M^9 = \begin{pmatrix} 0.652174 & 0.347826 & 0 & 0 \\ 0.652174 & 0.347826 & 0 & 0 \\ 0.652174 & 0.347826 & 0 & 0 \\ 0.652175 & 0.347825 & 0 & 0 \end{pmatrix}.$$

Thus, to six decimal places, the transition matrix does not reach an equilibrium state until after ten steps.

Similar to the process described above for 9-year olds, the authors created a discrete Markov-Chain model based on NAEP data score gaps for 13-year olds (Tables 3 and 4).

As shown in Table 4, although smaller or larger intervals of the movement of g could be created, the authors created four states $\{s_1, s_2, s_3, s_4\}$ since gap scores have historically moved within these ranges and defined x_{ii} as the number of times g transitions from class ito class *j* [13].

To complete the development of the discrete Markov-chain model for age 13, as described above for age 9, we established the dependence of g against the test hypothesis that g is statistically independent by using the Chi-Square test. We found that the result of the Chi-Square test was 16.107 (p=0.013) and was significant at the 98.700 percent confidence level. Thus, based on these results we proceeded in creating our discrete Markov-Chain model for age 13.

NAEP Assessment Year	Score gap (g)
1973	46
1978	42
1982	34
1986	24
1990	27
1992	29
1994	29
1996	29
1999	32
2004	30
2008	28
2012	29

Table 3: Mathematical achievement score gaps (g) between African American and White American students at Age 13 during the NAEP Assessment Years 1973-

s_1	$23 \le g \le 30$
s_2	$30 \le g \le 37$
s_3	$37 \le g \le 44$
S_4	44 ≤ <i>g</i> ≤ 51

Table 4: Transition states and *g* classes for Age 13 students.

For the 13-year old age level, counting the transitions of the score gaps between the four states, $\{s_1, s_2, s_3, s_4\}$ defined above, we find the transition matrix to be:

$$M = m_{ij} = Mx_s = x_s = \frac{x}{\sum_{s=1}^{4} x}$$
. Thus, we have

$$M = m_{ij} = Mx_s = x_s = \frac{x}{\sum_{n=1}^{4} x}$$
. Thus, we have
$$M = \begin{pmatrix} \frac{6}{7} & \frac{1}{7} & 0 & 0\\ 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0 \end{pmatrix}$$
, where the entries in M are the transition

probabilities in our discrete Markov-Chain model.

Again, to generate our probability statistic to predict whether the mathematical achievement gap will close within the next 50 years, we first find our transition matrix prior to fifty steps to determine if it reaches a steady state. Using Mathematica to perform power calculations of M, after ten years (or steps) we find the transition matrix to be:

$$M^{10} = \begin{pmatrix} 0.875 & 0.125 & 0 & 0 \\ 0.875 & 0.125 & 0 & 0 \\ 0.875 & 0.125 & 0 & 0 \\ 0.875 & 0.125 & 0 & 0 \end{pmatrix}.$$

Since the transition matrix reaches equilibrium after ten steps, the authors investigated whether the steady-state matrix could be determined prior to ten steps. The authors found:

$$M^9 = \begin{pmatrix} 0.875 & 0.125 & 0 & 0 \\ 0.875 & 0.125 & 0 & 0 \\ 0.875 & 0.125 & 0 & 0 \\ 0.875001 & 0.124999 & 0 & 0 \end{pmatrix}.$$

Thus, to six decimal places, we find that the transition matrix reaches an equilibrium state after ten steps. Therefore, we conclude that based on our Markov process, the achievement gap will likely close within the next ten years for 13-year olds.

Finally, the authors created a discrete Markov-Chain model based on NAEP data score gaps at age 17 (Tables 5 and 6).

As shown in Table 6, we created four states $\{S_1, S_2, S_3, S_4\}$ for age 17 and defined x_{ii} as the number of times g transitions from class i to class j.

To complete the development of our discrete Markov-chain model for age 17, we again used a Chi-Square test. The result was 12.833 (p=0.046) and was significant at the 95.400 percent confidence level.

Thus, based on these results we proceeded in creating our discrete Markov-Chain model for age 17.

For the 17-year old age level, counting the transitions of the score gaps between the four states, $\{S_1, S_2, S_3, S_4\}$ defined below, we find the transition matrix to be:

$$M = m_{ij} = Mx_s = x_s = \frac{x}{\sum_{s=1}^{4} x}$$
. Thus, we have

$$M = m_{ij} = Mx_s = x_s = \frac{x}{\sum_{n=1}^{4} x}.$$
 Thus, we have
$$M = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} & 0 & 0\\ \frac{1}{3} & \frac{2}{3} & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0 \end{pmatrix},$$
 where the entries in M are the transition

probabilities in our discrete Markov-Chain model.

To generate our probability statistic to predict whether the mathematical achievement gap will close within the next 50 years for this age level, we first find our transition matrix prior to fifty steps to determine if it reaches a steady state. Using Mathematica to perform power calculations of M, after ten years (or steps) we found the transition matrix to be:

$$M^{10} = \begin{pmatrix} 0.500008 & 0.499992 & 0 & 0 \\ 0.499992 & 0.500008 & 0 & 0 \\ 0.499995 & 0.500025 & 0 & 0 \\ 0.499924 & 0.500076 & 0 & 0 \end{pmatrix}.$$

Finding that the transition matrix does not stabilize after ten steps,

NAEP Assessment Year	Score gap (g)
1973	40
1978	38
1982	32
1986	29
1990	21
1992	26
1994	27
1996	27
1999	31
2004	27
2008	26
2012	26

Table 5: Mathematical achievement score gaps (g) between African American and White American students at Age 17 during the NAEP Assessment Years 1973-2012

s_1	$20 \le g \le 26$
s_2	$26 \le g \le 32$
<i>S</i> ₃	$32 \le g \le 38$
S_4	$38 \le g \le 44$

Table 6: Transition states and *g* classes for Age 17 students.

the authors examined whether it would stabilize after fifteen years (or steps). The authors found:

$$M^{15} = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \end{pmatrix}.$$

Since the transition matrix reaches equilibrium after fifteen steps, we investigated whether the steady-state matrix could be determined prior to fifteen steps. We found:

$$M^{14} = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0.499999 & 0.500001 & 0 & 0 \end{pmatrix}.$$

Thus, to six decimal places, the transition matrix does not reach an equilibrium state until after fifteen steps. Therefore, we conclude that based on our Markov process, the achievement gap will likely close within the next fifteen years for 17-year olds.

Conclusion

Limitation of the Study

In creating the mathematical model for this study, the authors chose a random variable g, namely the score gaps in mathematical achievement between African American and White American youths, ages 9, 13, and 17, and analyzed the transition of this variable through a unit of time t in years. This type of analysis creates a limitation of the study in that other factors (such as socioeconomic status, enrollment patterns in mathematics courses, gender issues, teacher expectations, test anxiety, and etc.) that may impact the score gaps in mathematical achievement are not considered. Rather than isolating each of these potential impact factors and analyzing separately their effect on the mathematical achievement gap, our random variable g represented the collective outcome of these factors. Moreover, modeling the random variable in this regard was necessary to create the discrete Markov-Chain model. In light of this limitation of the study, the following implications are made.

Implications of the study

As can be discerned from the National Assessment of Educational Progress (NAEP) data, historically there has existed an achievement gap in mathematics between African American students, ages 9, 13, and 17 and their White counterparts. This achievement gap has influenced educational processes in public and private school systems, as well as public policies in the United States. Although programs and policies in the United States such as Algebra for All, No Child Left Behind, and the Common Core Mathematics Standards were implemented to improve the mathematical achievement of all American students, the consistent achievement gap in mathematics among various demographic groups has influenced these implementations. Moreover, professional mathematics education organizations such as the National Council of Teachers of Mathematics have called for equitable mathematics classrooms for all students of mathematics [5-8].

Based on the discrete Markov process, described as a stochastic, mathematical model, presented in this article, the authors found that the mathematical achievement gap that exists between African Americans and White Americans in the United States will likely close

within the next ten years for 9- and 13-year olds and within the next fifteen years for 17-year olds. Given that these predictions are varied for different age levels, they may serve useful as educational policy makers continue to develop public policy to address the mathematical achievement of all American youths, including those minority youths who consistently underachieve in mathematics.

References

- Johnson ML (1984) Blacks in mathematics: A status report. Journal for Research in Mathematics Education 15: 145-153.
- Jones LV, Burton NW, Davenport EC (1984) Monitoring the mathematics achievement of Black students. Journal for Research in Mathematics Education 15: 154-164
- Matthews W, Carpenter TP, Lindquist MM, Silver EA (1984) The Third National Assessment: Minorities and mathematics. Journal for Research in Mathematics Education 15: 165-171.
- Grissmer D, Flanagan A, Kawata J, Williamson S (2000) Improving student achievement: What state NAEP test scores tell us. Santa Monica, CA: RAND.
- National Council of Teachers of Mathematics (1989) Curriculum and evaluation standards for teaching school mathematics. Reston, VA: Author.
- National Council of Teachers of Mathematics (1991) Professional standards for teaching school mathematics. Reston, VA: Author.
- National Council of Teachers of Mathematics (1995) Assessment standards for teaching school mathematics. Reston, VA: Author.
- National Council of Teachers of Mathematics (2000) Curriculum standards for teaching school mathematics. Reston, VA: Author.
- Ladson-Billings G (1997) It doesn't add up: African American students' mathematics achievement. Journal for Research in Mathematics Education 28: 697-708.
- Gordon SP, Gordon FS, Tucker AC, Siegel MJ (2004) Functioning in the real world: A precalculus experience. New York: Pearson Education.
- 11. Kemeny JG, Snell JL (1976) Finite Markov Chains. New York: Springer-Verlag.
- Osiecka K, Stawicki RJ (2008) Markov Set-Chains as a tool for an analysis of household expenditure structure in Poland 1993-2003. Dynamic Econometric Models 8: 67-74.
- Anderson TW, Goodman LA (1957) Statistical Inference about Markov Chains.
 The Annals of Mathematical Statistics 28: 89-110.
- Stawicki RJ, Lawrence KD (1994) Using Markov Chains to identify potential large donors. The Journal of Business Forecasting Methods & Systems 13: 282-291.
- Leon SJ (2005) Linear Algebra with applications, 7th edition. New York: Pearson Education.
- Billingsley P (1960) Statistical Methods in Markov Chains. Santa Monica, CA: RAND Corporation.

J Appl Computat Math
ISSN: 2168-9679 JACM, an open access journal