Analyzing Changes in Student Graph Reasoning and Comprehension Regarding Graph Axis Presentation

Mr. Justin Cory Willis, University Of Maine- Orono

Justin Willis is a Graduate Instructor at the University of Maine, for the Mechanical Engineering Technology department. He is also a graduate student in UMaine's Master of Science in Teaching program, and a volunteer math tutor for students and adults in Old Town, ME. Research interests include statistics education in engineering applications, and career and adult education.

Dr. Brett D. Ellis, University of Maine

Dr. Brett Ellis, P.E. is an Assistant Professor in Mechanical Engineering Technology at the University of Maine and has worked in the mechanical engineering field for 22 years, with approximately 14 years of industrial experience. His industrial experience includes open hole testing in the oil and gas industry; failure analysis consulting; and extensive experience in the plastic processing industry, in which he designed plastic preforms and bottles, designed injection- and blow-molding tooling, designed and optimized polymer processing equipment, and led continuous improvement activities (e.g., Lean Six Sigma, SMED, and Gage R&R). Dr. Ellis's professional interests include stress analysis, solid mechanics, continuous improvement, design, and education. He is a licensed Professional Engineer and a Certified Six Sigma Black Belt.

Analyzing Changes in Student Graph Reasoning and Comprehension Regarding Graph Axis Presentation

Abstract

This study analyzes the effects of truncated or unlabeled graph axis presentations on student-drawn conclusions. The research subjects in question were students in natural science, forestry, medicine, or engineering technology majors in their second or third year. Students were provided survey questions that had different methods of axes labeling on the dependent variable (y-axis) and were scored and coded based on their correct or incorrect response. These multiple-choice survey questions included control questions from the National Assessment of Educational Progress (NAEP) 8th grade math standardized exam, along with experimental questions of similar format having either truncated or unlabeled axes. Students also reported their perceived confidence in their answer on a 0-100% scale. Analyses of student responses and confidence percentages were completed for each question, for all students and for students self-reporting as educated within Maine's K-12 school system. Results indicate that truncated and unlabeled axes decreased correct response levels by 20% and 55%, respectively, compared to control questions. Interestingly, self-reported student confidence for the truncated and unlabeled axes questions decreased by 10% and 2%, respectively, compared to control questions. Based upon the results, it is hypothesized that students receive mixed messages regarding visual and numerical presentation of a graph.

Introduction

This study seeks to understand and quantify statistical literacy of students, namely their comprehension of graphs and pictorial depictions of information. This issue is important to analyze due to its real-world implications. Society uses graphical methods to quickly convey information, sometimes in manners intended to mislead or misinform. This use of graphical methods has increased in recent years as demonstrated by 72% of worldwide working professionals reporting they are working with more data in making decisions then they did three years ago [1]. However, 55% of the same professionals felt as if they had inadequate education and insufficient tools to draw conclusions and make decisions upon graphical data [1]. Understanding and interpreting graphical data are also competencies quantified in 1st-through 5th-grade outcomes in the Data and Measurement section of the Common Core Standards for Mathematics [2]. Understanding how current mathematics education prepares students to navigate and draw conclusions based on these graphical methods allows researchers to locate and address gaps in graphical literacy.

This research seeks to characterize rates of recognition for common misleading graph presentations, including alteration of axes scales, deformation of scales, and unlabeled axes. A question form of this inquiry could be "Do students interpret and recognize characteristics of potentially misleading bar and line graph axes?" The methods employed included having subjects draw conclusions based on complete or incomplete bar and line graphs and provide the confidence in their answer. Sub-questions included "Do students accurately measure their confidence and self-efficacy regarding their ability to interpret and recognize characteristics of potentially misleading bar and line graph axes?" and "What, if any, differences exist between students from Maine and the general population regarding ability to interpret and recognize characteristics of potentially misleading bar and line graph axes?"

Study of factors influencing student graph comprehension

Culbertson and Powers [3] describes a graph comprehension study of a 100 agricultural vocational students and 250 high school students in Wisconsin. Students were shown graphs with varying characteristics (*e.g.*, bar versus line graphs, labeling individual bar values on a bar chart versus providing grid squares and axis labels). The students were then tasked with determining data values based on the graphs. Their levels of successful graph analysis and comprehension were compared for two similar graphs, and each similar pair was analyzed to determine which graphical elements were more intuitive, *i.e.*, which graphical elements have the greater probability of being read correctly by students. The study then produced a list of "easier" and "difficult" graphical elements that increased and decreased comprehension, respectively. An analysis of student comprehension to student mental aptitude indicated a weak correlation, with comprehension of "difficult" graphical elements increasing with increased student aptitude.

In 1987, Curcio [4] reported the ability of 4^{th}- and 7^{th}-graders to read and interpret graphs. The 4^{th}- and 7^{th}-graders read data directly off the graph, made inferences for points between data points, and extrapolated additional data sets (*e.g.*, Where would additional, randomly selected data points fit in the data set?). Different variables such as form of graph presentation and type of graph were plotted against ability to answer the comprehension questions. Student comprehension was analyzed by demographic data, mathematical aptitude, English aptitude, and gender. Student graphical comprehension increased with increased grade level, mathematical aptitude, and English aptitude, and was correlated to graphical form (*e.g.*, student comprehension of line and bar graphs was greater than student comprehension of tables). Student graphical comprehension was independent of student gender.

Yolcu [5] investigated graphical comprehension of middle school students via a standardized Statistical Literacy Test (SLT). Based upon the statistical literacy framework developed by Watson [6], the SLT seeks to test and measure three tiers of statistical knowledge: (1) reading data from a graph, (2) making inferences between data points, and (3) extrapolating to additional data sets. The SLT was administered to 6^{th}- through 8^{th}-grade students in several middle schools in Ankara, Turkey. Grade level and gender factors were found to be statistically insignificant on SLT scores. The statistical insignificance for grade level was hypothesized to be due to: (1) the cyclic nature of Turkish middle-school mathematics curricula, and (2) lack of statistics education in 8^{th} grade instruction since instructors often focused on arithmetic to prepare students for an arithmetic-weighted 8^{th} grade exit exam.

Applications of current study

Culbertson and Powers [3] and Curcio [4] both identified labeling and choice of axis labels as a difficult graphical element. This study assesses: (1) the first and second tiers of understanding identified by Watson [6] in more detail regarding labeling and choice of axis labels, and (2) how much of the difficulties seen in previous work carry forward from middle and high school to early career college students. Such findings can illuminate: (1) if graph comprehension abilities observed in Culbertson and Powers [3] and Yolcu [5] evolve from middle and high school to college, and (2) what types of graphical misrepresentations are most difficult for students to overcome.

Maine students entering college are of particular interest to this study due to some of the unique challenges within the Maine school system. Due to Maine's rural nature, many primary, middle, and high schools in Maine are small. Accordingly, many of Maine's schools have limited personnel to teach mathematics and statistics. Further, course offerings are guided by graduation requirements and standardized testing [7], causing less required and less tested subjects such as statistical and data literacy to be taught infrequently. Combining this with 63% of preservice mathematics teachers listing statistics as the concept in mathematics they feel least comfortable teaching [8], Maine schools may provide fewer statistical and data literacy opportunities taught by less confident instructors than larger schools.

Research design

This study included a total of 219 students enrolled in a college-level general physics course at a public, northeastern United States university. Most of these students were natural science, forestry, medicine, or engineering technology majors in their second or third year. This cohort was selected due to the hypothesized high percentage of students completing their K-12 education in Maine. The criterion for a "Maine-educated" subject is defined as a subject having completed at least 9 of their 13 years of schooling between kindergarten and 12th grade at a public or private Maine school. Approximately 35% of those surveyed in the class self-reported as meeting this criterion.

Each student was presented a total of two multiple-choice graphical-literacy questions – the first question was from a pool of three control questions, and the second question was from a pool of two experimental questions. After answering each graphical-literature question, students were asked to provide their confidence in their answer on a scale between 0% and 100%, with 0% being not confident at all, and 100% being completely certain. The survey was administered by paper questionnaires in three, approximately equal-sized sessions; all students within a session received the same the same two questions. The two experimental questions combined with the sampling protocol meant that two sessions received the same experimental question, while the third session received the other. Future sampling protocols should alleviate the disproportionate sampling via question randomization.

Responses were coded as either correct or incorrect for each question provided, with unanswered questions omitted from tabulation. Confidence values were tabulated as a percentage for each question, and students who self-reported that they attended at least 9 of their years of K-12 schooling were flagged for subgroup analysis. This was the only demographic data taken, and names or other confidential information were not collected. This study complied with the approved IRB research plan; students' responses did not affect any grades in the course.

Control questions

The first multiple-choice graphical-literacy question for each session was randomly selected from pool of three control questions. The three control questions were selected from 8^{th} grade mathematics questions included in the 2015 and 2016 National Assessment of Educational Progress (NAEP) multiple-choice question banks [9] and are reproduced below in Figures 1-3.

According to the graph above, how many times did the yearly increase of the price of a hamburger exceed 10 cents?

A. None

B. One

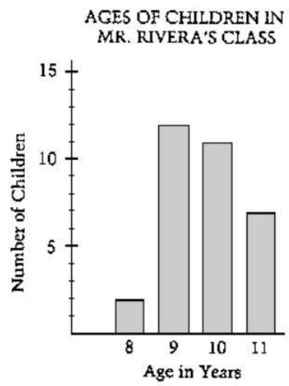
C. Two

D. Three

E. Four

Figure 1: Control Question 1 – Hamburger Prices

Subjects receiving this question were ideally expected to use the 10-cent increments of the vertical axis to determine the three times the graph shows an increase of more than 10 cents in one year, *i.e.*, 1986–1987, 1988–1989, and 1989–1990. Students were then able to use this information to select the correct answer (D).



The graph above shows how many of the 32 children in Mr. Rivera's class are 8, 9, 10, and 11 years old. Which of the following is true?

A. Most are younger than 9.

B. Most are younger than 10.

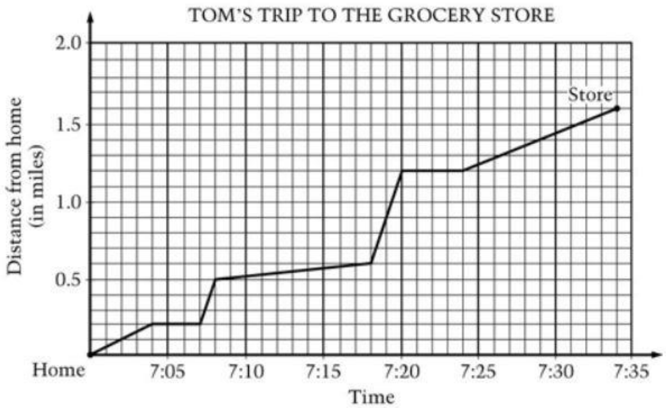
C. Most are 9 or older.

D. None of the above is true.

Figure 2: Control Question 2 – Riviera's Class

Students presented with this control question were ideally expected to be able to use the vertical axis labels to find the total amount of students of ages 8, 9, 10, and 11 years old. With these totals the student would be able to test the possible answers for validity, selecting the only true choice provided, answer (C).

Tom went to the grocery store. The graph below shows Tom's distance from home during his trip.



Tom stopped twice to rest on his trip to the store. What is the total amount of time that he spent resting?

A. 5 minutes

B. 7 minutes

C. 8 minutes

D. 10 minutes

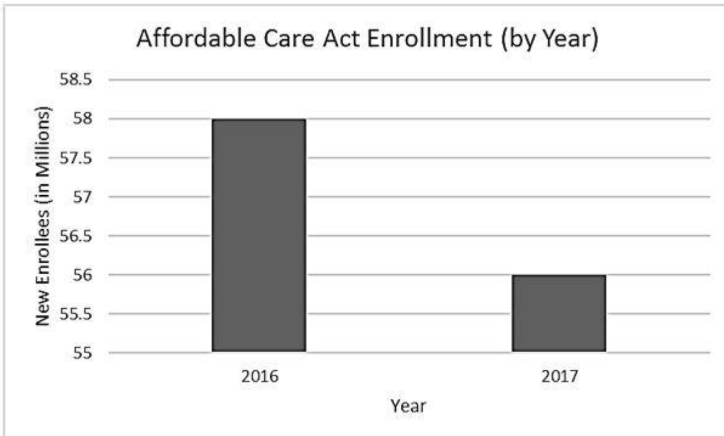
E. 25 minutes

Figure 3: Control Question 3 – Grocery Store

Students presented with this question were expected to recognize that the times Tom spent resting were times when his distance from the store did not change, represented by time intervals where the line of the line graph is horizontal (*i.e.*, slope equal to zero). Once reaching this conclusion, the student would be able to use the two visible horizontal portions of the graph and the scale of the horizontal axis to calculate the total time resting was seven minutes. Students would then select the corresponding correct answer (B).

Experimental questions

The second multiple-choice graphical-literacy question for each testing session was randomly selected form a pool of two experimental questions. The two experimental questions were designed by one of the authors to mimic control question format and employed fictional data and graphical representations. The first experimental question employed an intentionally truncated vertical axis; the second experimental question employed an unlabeled vertical axis. The two experimental graphical-literacy questions are shown in Figures 4 and 5.



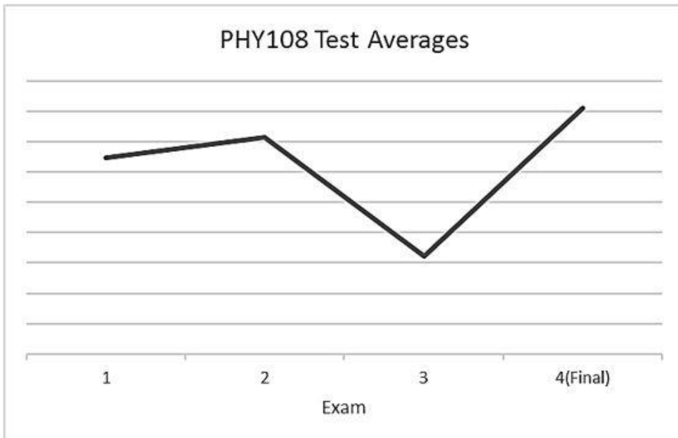
Use the graph above to complete the statement below.

“According to the graph above, enrollment in 2017 was approximately ____ of enrollment in 2016.”

A. 34%
B. 44%
C. 96%
D. 72%
E. There is not enough information to tell

Figure 4: Experimental Question 1 – Affordable Care

Students presented with this experimental question were expected to use the labeled units of the vertical axis to correctly determine 58 million enrollees in 2016, and 56 million enrollees in 2017. They were then able to use these two numbers to calculate the ratio of 2017 enrollment to 2016 enrollment. This was calculated by dividing 56 million by 58 million, for a percentage of approximately 96%. Students were then able to select the correct answer (C). Without noticing the truncated axis, students may have been led to believe that 2017 enrollment was approximately a third of 2016 enrollment, due to initial visual inspection. This may have led them to select the incorrect response (A).



True or False: The exam average on Exam 1 was approximately double the average on Exam 3.

A. True
B. False
C. There in not enough information to tell

Figure 5: Experimental Question 2 – Test Averages

Although visual inspection indicates the true-or-false statement is true, students were expected to notice that the vertical axis is unlabeled, which makes any assertions regarding test averages impossible. Due to the inability to compare numerical values, there is insufficient information to draw conclusions regarding the averages of the two tests. Therefore, students would conclude that there was insufficient information to make a claim and select the correct answer (C).

Findings

Table 1 shows a summary of graphical literacy and self-reported confidence for the entire cohort by question. Whereas the control questions were answered correctly by 91% to 97% of the cohort, the experimental questions, with intentional graphical misrepresentations, were answered correctly by 74% and 39% of the cohort. The differences between 93.6% correct for the cumulative control questions and 74% and 39% correct for the Affordable Care and Test Averages experimental questions were statistically significant at an α of 0.05 with *p*-values of <0.001 and <0.001, respectively.

The reduction in correct interpretation of graphical elements is even more striking when considering that that self-reported confidence decreased from 85% to 96% for the control questions to 82% to 90% for the experimental questions. The differences between 92.0% confidence for the cumulative control questions and 82% confidence for the Affordable Care experimental question was statistically significant at an α of 0.05 with a *p*-value of .003. The differences between 92.0% confidence for the cumulative control questions and the 90%

confidence for the Test Averages experimental questions was statistically insignificant with a *p*-value of .567. In summary, the cohort was unable to correctly interpret graphical misrepresentations and, equally important, was for the most part unaware of their inability to correctly interpret graphical misrepresentations, especially in the Test Averages experimental question.

A second observation is that the 39% correct for the Test Averages question was significantly less than the 74% correct for the Affordable Care question. The lack of vertical axis labeling in the Test Averages question may have encouraged students to rely on visual interpretation, even though visual interpretation lacked concrete evidence of being correct. Several students drew their own axis tick marks and units on the vertical axis of this question and used their hand-drawn demarcations to answer the question, thus indicating their discomfort with a lack of axis labels and inability to accept the lack of quantifiable data.

Table 1: Summary of graphical literacy and self-reported confidence for entire cohort by question.

		Graphical literacy		**Self-reported confidence**	
Type	**Question**	***% Correct (95% CI)***	***N***	***% Confident (95% CI)***	***N***
Control	1: Hamburger Prices	91% *(81.5% - 96.1%)*	74	85% *(74.6% - 92.2%)*	73
	2: Riviera's Class	97% *(90.3% - 99.7%)*	72	96% *(88.1% - 99.1%)*	71
	3: Grocery Store	93% *(84.7% - 99.7%)*	73	95% *(83.4% - 97.0%)*	73
	Cumulative	93.6% *(89.5% - 96.5%)*	219	92.0% *(87.8% - 95.4%)*	217
Experimental	4: Affordable Care	74% (66.3% - 81.0%)	147	82% *(74.3% - 87.7%)*	142
	5: Test Averages	39% (27.6% - 51.1%)	72	90% *(80.5% - 95.9%)*	70
	Cumulative	62.5% (55.8% - 69.0%)	219	84.6% *(78.8% - 89.0%)*	212

Formatted: Left

Analysis of Maine-educated student subgroup

Table 2 shows graphical literacy correctness and self-reported confidence for the students who self-reported as members of the Maine-educated subgroup. Similar to the entire cohort, students in the Maine-educated subgroup correctly answered control questions between 94% and 100% correct and answered the experimental questions between 39% and 74% correct. The differences between 95.0% correct for the cumulative control questions and 74% and 39% correct for the Affordable Care and Test Averages experimental questions were statistically significant at an α of 0.05 with *p*-values of 0.001 and <0.001, respectively.

Also, similarly, the Maine-educated subgroup self-reported confidence only decreased slightly from the control questions to the experimental questions. The differences between 92.3% confidence for the cumulative control questions and 81% confidence for the Affordable Care experimental question was statistically significant at an α of 0.05 with a *p*-value of 0.047. The differences between 92.3% confidence for the cumulative control questions and the 87% confidence for the Test Averages experimental questions was statistically insignificant with a *p*-value of 0.407. In summary, the cohort was unable to correctly interpret graphical misrepresentations and, equally important, was for the most part unaware of their inability to correctly interpret graphical misrepresentations, especially in the Test Averages experimental question.

Table 2: Summary of graphical literacy and self-reported confidence for Maine-educated subgroup by question.

		Graphical literacy		Self-reported confidence	
Type	**Question**	***% Correct***	***N***	***% Confident***	***N***
Control	1: Hamburger Prices	100% *(75.3% - 100.0%)*	13	79% *(46.2% - 95.0%)*	13
	2: Riviera's Class	94% *(78.6% - 99.2%)*	31	95% *(78.6% - 99.2%)*	31
	3: Grocery Store	94% (79.8% - 99.3%)	33	95% *(79.8% - 99.3%)*	33
	Cumulative	95.0% *(87.2% - 98.6%)*	77	92.3% *(83.8% - 97.1%)*	77
Experimental	4: Affordable Care	74% *(58.9% - 85.7%)*	46	81% *(65.4% - 90.4%)*	45
	5: Test Averages	39% *(21.9% - 57.8%)*	31	87% *(70.2% - 96.4%)*	31
	Cumulative	59.9% *(47.9% - 70.8%)*	77	83.4% *(72.5% - 90.6%)*	76

Overall student population response analysis

Student responses were stratified into four categories as shown in Table 3. With 59.1% of the cohort, the largest category is composed of students who answered the control question and the experimental question correctly. While 93.6% of the overall population correctly completed their control question, only 63.4% of those also completed their experimental question correctly.

Table 3: Response combinations

Category	***N***	**%**
Correct answer to control question; Correct answer to experimental questions	130	59.4%
Correct answer to control question; Incorrect answer to experimental question	75	34.2%
Incorrect answer to control question; Correct answer to experimental question	7	3.2%
Incorrect answer to control question; Incorrect answer to experimental question	7	3.2%

Control versus experimental question correlation comparisons

McNemar's test for independence was employed to determine if a statistically significant difference existed between mean rates of correct answers for control and experimental questions. McNemar's test was conducted [10] using the 2 × 2 contingency table of correctness shown in Table 4.

Table 4: Contingency table of correctness

		Control		**Sum**
		Correct	**Incorrect**	
Experimental	**Correct**	130	7	137
	Incorrect	75	7	82

Sum	205	14	219

Using McNemar's test for independence, an α-value of less than 0.001 was utilized to reject the null hypothesis regarding these two sample means and find that the relationship between success on the control and experimental questions was statistically significant. Therefore, it is expected that rates of correctness for control and experimental questions would continue to remain different with an increasing sample of similar demographic. However, the control questions used in this manuscript are an accurate predictor of experimental question success 62.7% of the time (using the amount of correct-correct combinations and incorrect-incorrect combinations in relation to the total). Given the fact that random combination amounts would be an accurate predictor 50% of the time, it is concluded that control question success is not a strong predictor of experimental question success. This contrasts with the predictability of experimental question success on control question success: 94.9% of those completing the experimental question correctly also correctly answered the control question correctly. Given their statistically different rates of correctness and lack of bilateral predictive relationship, it was concluded the successful completion of control and experimental questions require different levels or skills of graphical comprehension.

Analysis of student confidence and self-efficacy

Differences in student confidence were found to be statistically insignificant between responses on control and experimental questions, with students self-reporting confidences greater than 80%. This follows in line with the studied effect of students overestimating their own ability to solve mathematics problems [11]. Interestingly, students measured their confidence in their answer and indirectly their self-efficacy to answer problems of this type at similar percentages regardless of questions, despite experimental questions being actually answered correctly at a statistically significant lower rate. This large amount of overestimation in calibration of a student's own ability is often seen in students with beginning or novice understanding [12]. This overestimation of self-efficacy can sometimes be beneficial in having students persist in a difficult task they otherwise may give up on, but can cause false security for the student regarding incorrect conclusions [12]. This can be problematic if students are presented with the opportunity to make conclusions from data, which they are likely to do, particularly in those entering a STEM-related field such as engineering technology.

Conclusions and implications

The sharp decrease in student success on the experimental questions in relation to control questions shows that students do indeed encounter difficulty when encountering graphs with axes either truncated or unlabeled, which is a common presentation of graphs in media and industry work, sometimes with the intent to mislead. This conclusion may be of use for informing education decisions regarding instruction in interpreting graphical data in K-12 settings and informing those who present data to the public through industry and public outlets. The corresponding lack of change in confidence between the control and experimental questions shows that students are likely unaware that this difficulty exists. This may be problematic as they would be unaware of their incorrect conclusions when presented with certain data representations.

The failure of the control questions to accurately predict the outcomes of the experimental questions suggests that students are using different strategies between the two types of questions. A hypothesis for the behavior being seen in this study is the ability to reconcile the visual representation and perceptions of the graph (*i.e.*, how it looks) with its numeric representation (*i.e.*, the axis labels). For instance, when not provided with full context in a problem a student stops looking at the problem from a mathematic, numerical perspective and instead relies on purely visual cues, which, by design, would lead a student to an incorrect answer on the experimental questions.

Further investigation into this subject could include more detailed and qualitative analysis of students to develop a theory why this difficulty exists in students regarding interpreting graph axes and the information they convey. Some data was collected in this survey asking students to explain their reasoning. Although omitted here, this data may be analyzed to further illuminate student thought processes. A possible research question arising could be "What thought processes are employed by high school graduates and early college individuals to interpret axes labels?" Further data should also be collected from varied demographic groups as well to determine patterns or differences between student groups.

Bibliography

[1] Qlik, "Data Literacy Survey Results," *Data Literacy Campaign*, 2018. [Online]. Available: www.dataliteracy.info/application. [Accessed: 04-Oct-2018].

[2] National Governors Association Center for Best Practices and Council of Chief State School Officers, "Common Core State Standards for Mathematics (Data and Statistics)," 2010. .

[3] H. M. Culbertson and R. D. Powers, "A study of graph comprehension difficulties," *Educ. Commun. Technol.*, vol. 7, no. 3, pp. 97–110, 1959.

[4] F. R. Curcio, "Comprehension of Mathematical Relationships Expressed in Graphs," 1987.

[5] A. Yolcu, "Middle school students' statistical literacy: Role of grade level and gender," *Stat. Educ. Res. J.*, vol. 13, no. 2, pp. 118–131, 2014.

[6] J. Watson, "Assessing statistical thinking using media," in *The Assessment Challenge in Statistics Education*, Amsterdam: IOS Press, 1997, pp. 107–122.

[7] N. Daun-Barnett and E. P. St. John, "Constrained Curriculum in High Schools: The Changing Math Standards and Student Achievement, High School Graduation and College Continuation," *Educ. Policy Anal. Arch.*, vol. 20, p. 5, 2015.

[8] J. N. Lovett and H. S. Lee, "New Standards Require Teaching More Statistics: Are Preservice Secondary Mathematics Teachers Ready?," *J. Teach. Educ.*, vol. 68, no. 3, pp. 299–311, 2017.

[9] National Center for Education Statistics, "National Assessment of Educational Progress (NAEP)," Ann Arbor, 2015.

[10] R. Lowry, "McNemar's Test For Correlated Proportions in the Marginals of a 2x2 Contingency Table," *VassarStats*, 2008. [Online]. Available: http://vassarstats.net/propcorr.html. [Accessed: 17-Dec-2018].

[11] F. Pajares and J. Kranzler, "Role of self-efficacy and general mental ability in mathematical problem solving: A path analysis," in *Annual Meeting of the American*

Educational Research Association, 1995, no. 1986, pp. 3–33.
[12] D. H. Schunk, *Learning Theories: An Educational Perspective*, 6th ed. Pearson Education Inc., 2012.