American Journal of Educational Science
Vol. 4, No. 4, 2018, pp. 69-80
http://www.aiscience.org/journal/ajes
ISSN: 2381-7127 (Print); ISSN: 2381-7135 (Online)

Potential Influences of Gender on NAEP Scores in an 8th Grade U.S. History Class

Erika Sharp, Mingyuan Zhang*

Department of Teacher Education and Professional Development, Central Michigan University, Mount Pleasant, USA

Abstract

This study presented a secondary analysis of the National Assessment of Educational Progress (NAEP) dataset. The paper examined if a gender gap existed in the non-STEM subject of eighth-grade U.S. history. This study used a quantitative descriptive and hypothesis testing research design to analyze data extracted from the 2014 NAEP Data Explorer. The findings include (1) the average scale score of female students was significantly lower than their male counterparts, (2) the average score of female students did not start to become lower than their male counterparts until the 2001 NAEP scores, and (3) first-year students who remained full time had better student score outcomes than those that became part-time. These findings may indicate that the gender gap that exists is not purely based on intrinsic factors, but other factors such as teacher gender, and school community support may be an indication of causes of gender gaps for both male and female students.

Keywords

Gender, NAEP, History Assessment, National Data, Data Mining

Received: June 21, 2018 / Accepted: July 6, 2018 / Published online: August 10, 2018

1. Introduction

There are much conversation and research regarding the gender gap in STEM subjects, and then further in STEM-based careers. A noticeable lack of females in STEM careers has been noted [1]. This conversation has not been extended to whether there is a gender gap in non-STEM subjects [2]. Much of the research regarding the STEM gender gap focuses on the intrinsic interest of males and female. Females are more interested in people-based subjects that require a connection with feelings and communication, such as English, Psychology, and Sociology. Although U.S. history is not a STEM subject, the subject could not be categorized as people based, or requiring any amount of communication skills. With all the evidence being compiled in the gender gap for STEM, some of the issues being identified in that line of research may apply to non-STEM subjects that bare resemblances to STEM subjects regarding the mechanics of how the subjects are studied and taught, and how students may consume the information mentally [3, 38].

The primary research problem for the present study was to examine if there is a gender gap in grade 8 U.S. history scores. The reasoning behind this problem is, studies have shown that by age six girls feel as though they are not as smart as boys [4]. In a non-STEM subject could this gender gap be seen in scores between boys and girls?

The evidence is abundant that males have a higher incidence of being in STEM-based careers, and females have a higher incidence of being in non-STEM careers. This is tied back to scores received in school. In the United States, 75% of teachers are female while 25% are male [5]. This is on trend with females being intrinsically interested in people-based subjects. Within the subject of history, 64.1% of the teachers are male, and 35.9% are female, which goes against national trends [5]. Research has shown that males go into STEM careers because of higher scores in those subjects in school.

* Corresponding author
E-mail address: zhang1m@cmich.edu (Mingyuan Zhang)

Based on the percentage of male versus female teachers in history, it may be of note that males have higher scores in history and there is a gender gap present in the non-STEM subject of history.

While it is shown that males are a higher percentage of history teachers overall, the overwhelming statistics is that females are the higher percentage of teachers in all subjects. What is not known is if the higher percentage of male history teachers is directly correlated with scores from secondary school, or if the higher percentage of female teachers choose to engage with other subjects. Furthermore, the evidence has not shown a direct correlation between 8th grade U.S. history scores and a gender gap, which is what this study aims to research.

The research presented in this study should be of interest to K-12 educators, and administrators to understand the influence that gender has on learning outcomes, to bridge the gaps that exist between genders. This is important to all an equal opportunity for both males and females to reach their fullest potential in education. Outside of academia fellow researchers will find this research of interest to juxtapose against current research regarding STEM subjects and use as a template to discuss if there are any gender gaps present outside of STEM.

This study will aim to answer the following research questions:

1. Does gender influence scores in history?
2. Outside of gender what are some influencing factors that could contribute to a student's performance in 8th-grade history?
3. Is there a correlation between STEM scores and non-STEM scores in relation to gender?
4. Is there a correlation between school setting, gender, and history scores?

The theoretical framework for this research adopts a scientific inquiry-based approach. The scientific inquiry-based approach, according to the National Science Education Standards [6], the Benchmarks of Science Literacy [7], International Society for Technology in Education, [8], and Next Generation Science Standards [9], is investigation driven and science process initiated. It also goes beyond the mere development of process skills such as observing, inferring, questioning, interpreting, and analyzing data. It combines these processes with scientific knowledge, scientific reasoning, and critical thinking to develop scientific knowledge [10]. Scientific inquiry will guide us in examining the nature of the data [11]. By following the scientific inquiry, the authors will begin with an extensive exploration of the dataset (NAEP) and then design data-based research questions to mine the data systematically since there is no single set of sequential steps to follow in the scientific inquiry [12]. The focus of the research is the goal of the data mining – the discovery of knowledge from data [13]. With the scientific inquiry-based approach and data mining focus, the authors established this theoretical framework for the present quantitative data mining research [14].

2. Literature Review

In recent years a spotlight has been placed on the role that gender plays in student scores when looking at STEM subjects. In addition to student scores, the number of female participants versus male participants in these subjects has also been the subject of research. The reasons behind these disparities have been at the center of intense debates that have gone far beyond the realm of academia. Much research has looked at why female students tend to have lower STEM scores than their male counterparts.

Not much research is available for non-STEM subjects. This literature review will focus on reviewing whether non-STEM subjects, in particular, 8th grade U.S. History tend to follow the same patterns of gender disparities in scoring as the STEM subjects. Also, what factors beyond gender may have impacts on the achievement of boys and girls.

2.1. The Influences of Gender on Scores, Intrinsic or Extrinsic

When first reviewing if gender influences non-STEM scores, it is essential to understand how gender may play a role in any capacity on school scores. While the talk regarding STEM score disparities focuses on extrinsic factors, some researchers when looking at gender gaps holistically study intrinsic factors. Su, Rounds, and Armstrong [15] studied the interest factor in academia. What the researchers found was that intrinsic factors were a significant reason for interest in a particular subject over another, with female students tending to be interested in non-STEM subjects and the gender gap may be less ability and more motivation based [15].

In keeping with the look at motivation, Wang, Eccles, and Kenny [16] took the same approach and looked at interest-based rather than lack of ability on the part of women. They also hypothesized that lack of interest coupled with the higher verbal ability of female students, tend to open more career-oriented field, than those of their male counterparts with high STEM-based scores, but low non-STEM scores.

There were breaks from looking at the intrinsic motivation of a student in determining scores, and how societal norms, may factor into scores for each of the genders. Legewie and DiPrete [17] reject the notion that intrinsic factors influence scores and instead looks extrinsically, and more specifically at the high school community. The authors instead of relying on Gender Identify Theory, or the interest factor, instead surmise that scores are more related to the individual high school communities, than any intrinsic factors [17]. They believe that through interaction with other students, teachers, school administrators, other school staff, that these communities and conditioning of the students are more impactful than any gender-related intrinsic factors [17].

2.2. Teacher's Gender Influences Outcomes, Not Student's Gender

It is understood that a teacher can have a broad and profound impact on a student's achievement, but what was found in these articles is the impact teachers had on students based on their own inherent gender bias as well as the impact that the teachers' gender had on the student. These factors are separate from the ability of the instructor. Thomas Dee [18, 19] studied the impact that a teacher's gender had on student achievement and what he found was that boys were most affected by female teachers negatively. The negative effect of the female teachers was more widely seen in the subjects of history and English, more so than in STEM-related subjects [18].

While Escardíbul and Mora [20] agree that the teacher gender is impactful of students, these authors instead argue that female teachers have a more significant positive effect on students than male teachers. The leading prescription to this argument is the manner in which female teachers interacted with students versus male teachers, creating greater trust, and a stronger bond and therefore coxing a better performance out of the students [20].

Thomas Dee [18, 19] found that both male and female students were impacted negatively by opposite gender teachers. He surmised that for non-STEM subjects such as reading where boys tend to have lower scores than girls, if half the English teachers were male in grades, 6, 7, and 8 then the grade disparity between males and females would be cut by a third [18, 19]. The author could not account for why these gaps exist, but did surmise the effects could be as a result of the "role-model" effects, and not primarily in relation to the interactions within the classroom [18, 19].

2.3. Non-Educational Factors Are Influential

Most of the research caters to the notion that for extrinsic educational factors, such as the school, the teachers, and even fellow students have the primary impacts on a student's achievement. The research has shown that some extrinsic factors are outside of the realm of the school community. Singh and Mukherjee [21] found that among students, girls were more likely to achieve lower scores, because of the expectations on the home front. Due to the engagement of housework by girls in secondary school, this is one of the most significant contributing factors to lower achievement [21].

Victor Hiller [22] furthered this notion by suggesting that even in developed worlds, the ability of boy to bring in more money pushes parents to encourage them more in the way of education. Whereas girls are seen as less viable economically, and therefore their educational pursuits are not as encouraged and often put behind those of their male counterparts [22]. He also states that societal norms and coordination cause parents of different households to favor the education of their male children, over those of the females [22]. This notion also holds true for houses not containing any male children, where the value of female education is still not as strong as that for males [22]. This extrinsic non-educational factor does still have a significant impact on the gender gap seen in education, where girls are held back not by the subject, but by the structure of their own homes.

Regarding societal norms, the student may feel pressure to not step out of what is seen as normal. This pressure to conform may cause male students to remain in male-dominated fields such as STEM keep female students in non-STEM subjects. Kate Heddleston [23] wrote that when a person steps outside of societal norms, several consequences could be a result, but the overarching theme of those consequences is a negative reaction from society. These adverse reactions are known as social control whereas the societal participants seek to make the outlier conform in varying degrees of a negative reaction [23]. This fear of societal repercussion can lead to students adhering to what is thought as usual for their gender.

Gender inequality in education, even non-STEM subjects is a multi-pronged issue. While some gender disparity may be biologically related in relation to what the student has an interest, most of the research points to extrinsic factors. These extrinsic factors come from many places within and outside of the school community. In some instances, these factors are the school community itself. In a few cases, the gender of the student is not the determining factor in the impact of gender on achievement, but the gender of the instructor.

3. Methods

The NAEP Data Explorer was used to study the gender gap in 8^{th} Grade U.S. History. The data contained within this database were a good fit for the scientific inquiry-based approach described in the Introduction section of this paper.

The theoretical frame is investigation driven and science process initiated. It also goes beyond the mere development of process skills such as observing, inferring, questioning, interpreting, and analyzing data. The focus of the research is the goal of the data mining – the discovery of knowledge from data [13]. The participants were eight grade students, the location, and type of school was selected by NAEP.

3.1. Participants and Sampling

NCES [24] described the sampling and data collection protocols used for collecting NAEP history data every four years on a sliding grade scale of grades 4, 8, and 12. "According to the U.S. History Framework developed by the National Assessment Governing Board, the assessment should be organized around three components: Themes in U.S. History, Periods of U.S. History, and Ways of Knowing and Thinking about U.S. History" [25].

NCES [24, 26] described the sampling and data collection protocols used for collecting long-term trend assessment. This data uses a complex multi-stage collection. This data collection sampled 8th-grade students from within selected schools. The locations were various geographic areas across the country. "Each assessment cycle, a sample of students in designated grades within both public and private schools throughout the United States (and sometimes specified territories and possessions) is selected for Assessment" [27, 28].

The sample selected for NAEP uses a complex multistage sampling design that involves sampling students from selected schools within selected geographic areas across the country. The sampling design has the following stages: (1) selection of geographic areas (a county, group of counties, or metropolitan statistical area); (2) selection of schools (public and nonpublic) within the selected areas; and (3) random selection of students within the selected schools.

Each selected school that participates in the assessment and each student assessed represents a portion of the population of interest. Some smaller populations are oversampled to ensure sufficient representation. Therefore, sampling weights are needed to make valid inferences between the student samples and the respective populations from which they were drawn. Sampling weights adjust for disproportionate representation due to such oversampling.

3.2. Data Analysis

The NAEP Data Explorer [26] was used to analyze the data from the eighth-grade national public schools reading composite, gain information, and history grade scale scores for the year 2014 by gender. Cohen's *d* effect sizes [29] were calculated by using an online effect size calculator found at http://www.uccs.edu/~lbecker/ [30, 31]. Effect size is a standard measure that can be calculated from any number of statistical outputs. One type of effect size, the standardized mean effect, expresses the mean difference between two groups in standard deviation units [29]. "Though the values calculated for effect size are generally low, they share the same range as standard deviation (-3.0 to 3.0), so can be quite large" [29] ("Effect Size", n.d., para. 8). Cohen [29] defined *d* as the difference between the means, M1-M2, divided by standard deviation, *s*, of either group. The 2014 NAEP eighth-grade U.S. History assessment composite average scale scores and standard deviations were selected for the analyses.

3.3. NAEP Data Explorer

The NAEP Data Explorer [26] is a dynamic, interactive tool used to explore assessment results for various subjects, grades, and jurisdictions. The NAEP data explorer allows a user to search and then to create customizable tables and graphs, that display NAEP results. The results can be seen across multiple years, broken into different student groups, and the results can be searched by subject. The NDE is a statistical tool that features various analytical functions, such as the ability to search the NAEP results, compare NAEP data, and the aforementioned chart and table creation. The three coded questions selected through Data Explorer were:

1. Did you take a United States history course in the following grades? Options: 5th grade; 6th grade; 7th grade; 8th grade, Yes; No; I don't know.
2. How much do you agree that history or social studies is one of your favorite subjects? Options: Not at all; A little; A lot.
3. How important was it to you to do well on this test? Options: Not very important; Somewhat important; Important; Very important.

4. Results

The NAEP Data Explorer (NAEP, 2018) does not include the exact number of students who participated in the study. Results reported in this section represent that percentage of the variable without the frequency of the occurrence.

4.1. Research Question #1

Does gender influence scores in 8th grade U.S. history? There is much discussion about the gender gap in STEM subjects, but not many studies have looked at if a gender disparity exists in non-STEM subjects. U.S. history, while being non-STEM does not fit the characteristics of most non-STEM subjects, which are non-technological, non-mathematical, and more human interaction. While U.S. history is both non-technological and non-mathematical in nature, it is not human interactive and thus may not lend itself to the intrinsic preferences that have been identified in female students to explain some of the gap existing in STEM subjects.

Table 1. Average scale scores and standard deviations for grade 8 U.S. history, by gender [GENDER] and jurisdiction: 2014.

Year	Jurisdiction	Gender	Average scale score	Standard deviation
2014	National	Male	270	29
		Female	265	28

NOTE: Some apparent differences between estimates may not be statistically significant.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2014 U.S. History Assessment.

Table 1 presents NAEP eighth-grade U.S. history composite average scale scores by gender across the assessment year 2014. The average scale score for 100% of the students of the eighth-grade students on the 2014 NAEP U.S. history assessment for males was 270 and females 265 (scale-range 0-300) with a standard deviation of 29 for males and 28 for females. Differences in percentages by questions are presented in tables throughout the results section.

Table 2. U.S. History, grade 8 Difference in average scale scores between variables, for gender National, 2014.

	Male	Female
Male		> Diff = 4 P-value = 0.0030
Female	< Diff = -4 P-value = 0.0030	
LEGEND:		
<	Significantly lower.	
>	Significantly higher.	
x	No significant difference.	

NOTE: Within jurisdiction comparisons on any given year are dependent with an alpha level of 0.05.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2014 U.S. History Assessment

Table 2 presents the significance test of male and female eighth-grade U.S. history scores. Male scores were statistically significantly (p. <.001) higher than females' scores with an effect size of d=0.18. The effect sizes are interpreted as small across genders [29].

4.2. Research Question #2

Outside of gender what are some influencing factors that could contribute to a student's performance in 8th-grade history? The literature review presented topics regarding if influences on scores for students were exclusively extrinsic, intrinsic, or a combination of both. History is a specialized subject and whether a teacher has an educational background in the subject could be a contributing factor to a student's performance.

Table 3. Average scale scores and standard deviations for grade 8 U.S. history, by grad major/minor history and jurisdiction: 2014.

Year	Jurisdiction	Grad major/minor history	Average scale score	Standard deviation
2014	National	Major	268	29
		Minor/spec emphasis	272	30
		No	268	29

NOTE: Some apparent differences between estimates may not be statistically significant.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2014 U.S. History Assessment.

Table 3 presents NAEP presents the average scale score of the education background of history teachers. The average scale scores for the three groups of teachers did not show major differences in terms of if a teacher took history as major, minor, or not at all.

Table 4. U.S. History, grade 8 Difference in average scale scores between variables, for grad major/minor history National, 2014.

	Major	Minor/spec emphasis	No
Major		x Diff = -4 P-value = 0.2706 Family size = 3	x Diff = 0 P-value = 0.9838 Family size = 3
Minor/spec emphasis	x Diff = 4 P-value = 0.2706 Family size = 3		x Diff = 4 P-value = 0.1048 Family size = 3
No	x Diff = 0 P-value = 0.9838 Family size = 3	x Diff = -4 P-value = 0.1048 Family size = 3	
LEGEND:			
<	Significantly lower.		
>	Significantly higher.		
x	No significant difference.		

NOTE: Within jurisdiction comparisons on any given year are dependent with an alpha level of 0.05.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2014 U.S. History Assessment

Table 4 presents the significance test of teachers who majored in history, minored in history, and had no historical background. There was statistically no significant difference (p. <.001) between the three variables which had effect sizes that ranged from d=0 to d=0.13. The effect sizes are interpreted as small across education background [29].

4.3. Research Question #3

Is there a correlation between STEM scores and non-STEM scores in relation to gender? Looking at the overall picture of a student's performance could indicate if the student's performance is correlated beyond the subject and is due to other factors. Students who perform the same on STEM and non-STEM subjects in relation to their final grade could show that it is not the subject.

Table 5. Average scale scores and standard deviations for grade 8 science, by gender [GENDER] and jurisdiction: 2015.

Year	Jurisdiction	Gender	Average scale score	Standard deviation
2015	National	Male	155	35
		Female	152	33

NOTE: Some apparent differences between estimates may not be statistically significant.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2015 Science Assessment.

Table 5 presents NAEP eighth-grade Science composite average scale scores by gender across the assessment year 2015.

Table 6. Science, grade 8 Difference in average scale scores between variables, for gender National, 2015.

	Male	Female
Male		> Diff = 3 P-value = 0.0000
Female	< Diff = -3 P-value = 0.0000	
LEGEND:		
<	Significantly lower.	
>	Significantly higher.	
x	No significant difference.	

NOTE: Within jurisdiction comparisons on any given year are dependent with an alpha level of 0.05. SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2015 Science Assessment

Table 6 presents the significance test of male and female eighth-grade Science scores. Male scores were statistically significantly (p. <.001) higher than females' scores with an effect size of d=0.08. The effect sizes are interpreted as small across genders [29].

4.4 Research Question #4

Is there a correlation between school setting, gender, and history scores? A student's score could have many factors. Looking at an overall picture of the environment, and the students' gender could shed light on how students of different genders process their environment and if there is any effect on performance.

Table 7. Average scale scores and standard deviations for grade 8 U.S. history, by gender and jurisdiction: 2014.

Year	Jurisdiction	Gender	Average scale score	Standard deviation
2014	National	Male	270	29
		Female	265	28

NOTE: Some apparent differences between estimates may not be statistically significant.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2014 U.S. History Assessment.

Table 7 presents NAEP eighth-grade U.S. history composite average scale scores by gender across the assessment year 2014.

Table 8. U.S. History, grade 8 Difference in average scale scores between variables, for gender National, 2014.

	Male	Female
Male		> Diff = 4 P-value = 0.0030
Female	< Diff = -4 P-value = 0.0030	
LEGEND:		
<	Significantly lower.	
>	Significantly higher.	
x	No significant difference.	

NOTE: Within jurisdiction comparisons on any given year are dependent with an alpha level of 0.05.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2014 U.S. History Assessment

Table 8 presents the significance test of male and female eighth-grade U.S. history scores. Male scores were statistically significantly (p. <.001) higher than females' scores with an effect size of d=0.18. The effect sizes are interpreted as small across genders [29].

Table 9. Average scale scores and standard deviations for grade 8 U.S. history, by percent of new full-time teachers who stayed as full-time and jurisdiction: 2014.

Year	Jurisdiction	Percent of new full-time teachers who stayed as full-time	Average scale score	Standard deviation
2014	National	0-10%	258	30
		11-25%	255	23
		26-50%	265	26
		51-75%	259	30
		76-90%	265	29
		Over 90%	271	28

NOTE: Some apparent differences between estimates may not be statistically significant.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2014 U.S. History Assessment.

Table 9 presents the composite average scale score of new eighth-grade teachers that remained full-time across the assessment year 2014.

Table 10. The difference in average scale scores between variables, for percent of new full-time teachers who stayed as full-time National, 2014.

	0-10%	11-25%	26-50%	51-75%	76-90%	Over 90%
0-10%		x Diff = 3 P-value = 0.5557 Family size = 15	x Diff = -7 P-value = 0.1672 Family size = 15	x Diff = -1 P-value = 0.7745 Family size = 15	x Diff = -7 P-value = 0.1349 Family size = 15	< Diff = -13 P-value = 0.0020 Family size = 15
11-25%	x Diff = -3 P-value = 0.5557 Family size = 15		x Diff = -10 P-value = 0.0660 Family size = 15	x Diff = -5 P-value = 0.3895 Family size = 15	x Diff = -10 P-value = 0.0528 Family size = 15	< Diff = -17 P-value = 0.0034 Family size = 15
26-50%	x	x		x	x	x

	0-10%	11-25%	26-50%	51-75%	76-90%	Over 90%
	Diff = 7 P-value = 0.1672 Family size = 15	Diff = 10 P-value = 0.0660 Family size = 15		Diff = 5 P-value = 0.2722 Family size = 15	Diff = 0 P-value = 0.9189 Family size = 15	Diff = -7 P-value = 0.0541 Family size = 15
51-75%	x Diff = 1 P-value = 0.7745 Family size = 15	x Diff = 5 P-value = 0.3895 Family size = 15	x Diff = -5 P-value = 0.2722 Family size = 15		x Diff = -6 P-value = 0.2274 Family size = 15	< Diff = -12 P-value = 0.0046 Family size = 15
76-90%	x Diff = 7 P-value = 0.1349 Family size = 15	x Diff = 10 P-value = 0.0528 Family size = 15	x Diff = 0 P-value = 0.9189 Family size = 15	x Diff = 6 P-value = 0.2274 Family size = 15		x Diff = -6 P-value = 0.0514 Family size = 15
Over 90%	> Diff = 13 P-value = 0.0020 Family size = 15	> Diff = 17 P-value = 0.0034 Family size = 15	x Diff = 7 P-value = 0.0541 Family size = 15	> Diff = 12 P-value = 0.0046 Family size = 15	x Diff = 6 P-value = 0.0514 Family size = 15	
LEGEND:						
<	Significantly lower.					
>	Significantly higher.					
x	No significant difference.					

NOTE: Within jurisdiction comparisons on any given year are dependent with an alpha level of 0.05.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2014 U.S. History Assessment

Table 10 presents the significance test of the percent of new teachers who remained as full-time. Over ninety percent of new full-time teachers who remained full-time were statistically significantly (p. <.001) higher than 0-10% of new full-time teachers that remained full-time with an effect size of d=0.44. Over ninety percent of new full-time teachers who remained full-time were statistically significantly (p. <001) higher than 11-25% of new full-time teachers that remained full-time with an effect size of d=0.62. Ninety percent of new full-time teachers who remained full-time were statistically significantly (p. <.001) higher than 51-75% of new full-time teachers that remained full-time with an effect size of d=0.41. The effect sizes were interpreted to range from small to medium across the percentage of new full-time teachers who remained full-time [29].

Table 11. Average scale scores and standard deviations for grade 8 U.S. history, by school location, 12 categories, and jurisdiction: 2014.

Year	Jurisdiction	School location, 12 categories	Average scale score	Standard deviation
2014	National	City, midsize	260	30
		Suburb, large	270	29
		Rural, fringe	273	28

‡ Reporting standards not met.
NOTE: Some apparent differences between estimates may not be statistically significant.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2014 U.S. History Assessment.

Table 11 presents the average scale score between the variable of jurisdiction across the assessment year 2014

Table 12. The difference in average scale scores between variables, for school location, 12 categories National, 2014.

	City, midsize	Suburb, large	Rural, fringe
City, midsize		< Diff = -11 P-value = 0.0001 Family size = 3	< Diff = -13 P-value = 0.0002 Family size = 3
Suburb, large	> Diff = 11 P-value = 0.0001 Family size = 3		x Diff = -2 P-value = 0.3625 Family size = 3
Rural, fringe	> Diff = 13 P-value = 0.0002 Family size = 3	x Diff = 2 P-value = 0.3625 Family size = 3	
LEGEND:			
<	Significantly lower.		
>	Significantly higher.		
x	No significant difference.		

NOTE: Within jurisdiction comparisons on any given year are dependent with an alpha level of 0.05. SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2014 U.S. History Assessment

Table 12 presents the significance test of school location. Scale scores for Suburban (large) were statistically significantly (p. <.001) higher than scale scores for City (midsize) with an effect of d=0.33. Scales scores for Rural (fringe) were statistically significantly (p.<.001) higher than City (midsize) with an effect of d=0.44. The effect sizes are interpreted as small across the jurisdictions [29].

5. Discussion

This study aimed to verify if non-STEM courses had the same gender gaps reported in STEM courses. Much of the research presented speaks to several reasons for the gender gap, both intrinsic and extrinsic factors are thought to be contributing. None of the research looked at NAEP scale scores to compare male and female outcomes for non-STEM subjects, specifically in this study, eighth grade U.S. history. History was chosen because it does not fit the characteristics of other non-STEM subjects, which are human focused, and feelings based. Much like STEM, history is a subject that is grounded in facts that are indisputable.

5.1. Gender Influences on Scores in 8th Grade U.S. History

The NAEP results did show that there is a gap between male and female scores. While the scores of males were statistically significantly higher than females, there was a 5-point gap. This study focuses primarily on 2014, to understand how the gap came about, a look at the previous scores from the years 1994- 2014 was conducted. In 1994, male and female scores were the same [32]. Over time both scores increased, but between 1994 and the 2001 analysis that did not include accommodations, males saw a more considerable increase from 259 to 264 [32]. While the female scores only increased from 259 to 261 [32], which began this trend of males having higher scores. As a note, the 2001 scores with accommodations were 261 for males and 260 for females. With accommodation, the male population did not score better, but the female population did. The NAEP does not explain what accommodations are given, but in preceding years accommodations are available, and the male scores continued to trend higher than female scores.

This is important to note because intrinsic factors that were presented in both by Su, Rounds, and Armstrong [33] and Wang, Eccles, and Kenny [16] postulated that the reasons for these gender gaps were intrinsic and interest based. As humans we do not evolve in a matter of a few years in a way that would change the very nature of who were are, so with the results showing the decline of female scores it is possible that the influence on scores is not biological, intrinsic, or interest-based as presented, but possibly due to the way in which female and male students were being taught, lending this to be an extrinsic issue.

In the Legewie and DiPrete [17] study the authors surmised that gender gaps existed because of societal norms and conditioning. That the school community, interactions with other students, administrators, and teachers was a driving factor in scores more so than gender. Due to the scores from NAEP, this seems to be the most probable cause of the gender gaps seen in these scores. Nationally teachers are overwhelmingly female, except in the field of history where males outpace female teachers 64.1% to 35.9% [5]. Males are also shown to do better when having a same-gendered teacher, while the gender of teacher does not affect female students the same.

Unfortunately trying to find a correlation between the percentage of male and female teachers between the years of 1994 and 2014 is unavailable. NAEP does not provide the gender information for the responding teachers past 1994. It could be surmised that the incidence of higher male scores is a direct result of more male history teachers, allowing males to respond favorably and begin to increase the gap with their female counterparts.

5.2. Correlation Between STEM Scores and Non-STEM Scores in Relation to Gender

As previously discussed there has been discussion regarding the gender gap in STEM scores but is there a correlation between a student's non-STEM score and their STEM scores. This study reviewed the STEM score for male and female students through the NAEP Data Explorer and found that the 2015 Science NAEP scores showed males scored 155 on the composite scale score compared with the female scores of 152 [32]. Science teachers like history teachers tend to have a higher percentage of male versus female teachers. In one study conducted by Bottia, Stearn, Mickelson, Moller, and Valentino [34], the researchers looked at 270 schools in North Carolina. What they found was that while 56% of the science teachers were women when compared to women teachers in all subjects 80%, there was a proportionally lower percentage of female teachers in Science versus other fields such as English.

Bottia, Stearn, Mickelson, Moller, and Valentino [34] surmised that the issue with gender gaps in STEM is not a lack of interest or ability on the part of female students, but instead a lack of role models, they contribute this to the expectancy of males to pursue STEM careers and thus leaves the gap with female teachers. With the research at hand, the common thread is that the gender of the teacher has impacts on the scores of the students. The missing link in all these studies is how the gender disparage began, while there is a correlation between STEM scores and non-STEM scores such as history due to the overwhelming underrepresentation of female students in those two subjects the history of how these gender gaps began in the teaching fields is not understood or defined in any of the preceding literature.

The scores of males and females in subjects such as music, and writing showed an overwhelmingly higher scale score for females (140 males, 155 females Music, 140 males, 160 females Writing) [32] these scores are in line with the idea that female students do better in emotion and communication-based subjects. With the diverging and competing evidence, this information is not convincing that these scores are only because of the gender of the student, but there is a strong indication that societal norms and gender of instructor also has an impact and, in many cases, much larger than the intrinsic feelings of the students. This is concretely shown in the scores of history students where this non-STEM subject which is male teacher dominated tends to trend towards male students having higher scores. With the evidence presented it can be argued that gender of the teacher has an impact on both male and female students both positively and negatively.

5.3. Correlation Between School Setting, Gender, and History Scores

The previous discussion brought up that the school community could impact scores. The correlation between school community and score does not just apply to history, but school community has an overall impact on student scores. An NEA Policy Brief by Van Roekel [35] found that when parents, schools, communities, and families work together to provide support to a student, the student tends to earn higher scores. This support also correlates to regular attendance, retention, and enrollment in higher level programs [36].

The NAEP scores found that suburban (large) and rural (fringe) schools trended towards having higher scores than city (midsize) schools [32]. Some factors that could be contributed to this difference is the community involvement. These community factors are further stressed in that urban schools often have fewer resources to work with which has an impact on the student outcomes [39]. The evidence points to extrinsic factors; school, community, and parental support as having the highest correlation between history outcomes more so than the gender of the student. This is also due to the attitudes that this subject is seen as a masculine subject and therefore female students are not encouraged to pursue as they are not often encouraged to seek STEM-based subjects.

Another correlation between history scores and school setting is the length of time a teacher has been employed. The NAEP Data Explorer showed that first-year full-time teachers who remained full-time teachers over 90% of the time had the highest scale score outcomes for eighth-grade U.S. history [32]. While the lowest scores were recorded in the 11-25% range of teachers who remained full time, this could be due to the lack of experience being gained by no longer being a full-time teacher and does not have any correlation to a student’s gender, but more so is tied to the school, and quality of instructor.

6. Conclusion

The cause of gender gaps in 8th Grade U.S. History cannot be simply thought of as intrinsic or of biological consequence as previous research has concluded. Looking at the gender gaps within scores of non-STEM subjects such as history is important because there has been an occurrence these gender gaps carrying over into professional fields.

6.1. Gender Gaps Can Be Attributed to Extrinsic Factors

Gender gaps as seen in 8th grade U.S. history can be attributed not just to intrinsic factors as described by Su, Rounds, and Armstrong [33] and instead, extrinsic contributors do exist. These contributing factors include the school community, instructors, and home life. There was more convincing evidence that the focus on the gender gaps should not be the difference between male and female students, but the difference in the way male and female students are taught, treated, and the difference in the support each gender receives both inside and outside of the school community.

6.2. Teacher Gender Does Not Just Affect Male Students

Thomas Dee [18, 19] presented the idea that male students were more negatively impacted by female teachers than female students were impacted by male teachers. This implication could be incorrect. While overwhelmingly females outnumber males as teachers nationally, in the field of history males outnumber female teachers 64.1% to 35.9%. Within-subjects where female teachers outnumber males, the female students also outscore their male counterparts. This is not the case in history, where males both outscore female students, but male teachers outnumber female teachers. With these factors, it cannot be said that a teacher with a different gender impacts only male students. This is also another factor in showing that gender gaps do not exist primarily due to internal differences in males and females.

6.3. Teacher Experience Can Have Both Negative and Positive Impacts on Scores

The NAEP scores showed that the percentage of teachers that remained full-time after their first year had students with higher scores, while those that did not remain full-time had lower scores [32]. These percentages further show that factors contributing to scores and gender gap are extrinsic. The intrinsic nature that was reported in Dee [18, 19] is in correlation with societal norms and expectations. Experience of teachers was shown in the scores to have impacts on scores of students. Research provides information that teachers effectiveness continues to improve throughout their careers [37].

This study was conducted only using data available through the NAEP Data Explorer. Limitations present was the inability to match the students U.S. history scale scores with their STEM scores to see a side by side comparison of the student's performance in STEM and non-STEM subjects. Further limitations included not being able to match the student scores with their school community, and teachers experience to glean if those factors had any concrete correlations with the scores.

The implications of the evidence showing that the gender gaps exist due to extrinsic factors should lead educators to study these factors and put into place means to overcome or at least lessen the negative impacts that these factors have. This is not only important to female students who are shown to lag in this study but for male students who also lag in some subjects such as music and writing. As educators, the ability to learn and succeed should be equal and only student ability should be a factor in deciding scores, not factors as shown in this report.

How to alleviate the external factors that lead to gender gaps in education should be studied. Researchers should not only look at the internal differences between male and females for explaining why and how gender gaps exist in education, but future research should look to explain how external are contributing factors. One such factor that should be studied is how male and female students are taught differently. If there are any prejudices in teaching based on gender, these issues should be identified, called out, and educational institutions should work to eliminate such bias and prejudice.

References

[1] Wang, M. and Degol, J. "Gender gap in science, technology, engineering, and mathematics (STEM): Current knowledge, implications for practice, policy, and future directions." *Educational Psychology Review,* 29. 119-140. 2017.

[2] Ganley, C. M., George, C. E., Cimpian, J. R. and Makowski, M B. "Gender equity in college majors: Looking beyond the STEM/Non-STEM dichotomy for answers regarding female participation." *American Educational Research Journal,* 56. 453-487. 2018.

[3] Ackerman, P. L., Kanfer, R. and Calderwood, C. "High school advanced placement and student performance in college: STEM majors, non-STEM majors, and gender differences." *Teachers College Record*, 115. 1-43. 2013.

[4] Bian, L., Leslie, S. J. and Cimpian, A. "Gender stereotypes about intellectual ability emerge early and influence children's interests." *Science*, 355. 389-391. 2017.

[5] Yakimowski, M. "*Demographics Characteristics and Career Paths for School Psychologists: A Review of Literature*." Neag School of Education at the University of Connecticut. http://assessment.education.uconn.edu/assessment/assets/File/Revised%20Sch%20Psych%20ASEPS%20final%20draft.pdf. Accessed 16 Jun. 2018.

[6] National Research Council (NRC). *National science education standards*. National Press, Washington DC, 1996.

[7] American Association for the Advancement of Science (AAAS). *Benchmarks for science literacy*. Oxford University Press, New York, 1994.

[8] International Society for Technology in Education (ISTE). *ISTE standards: Teachers*, 2014, www.iste.org/docs/pdfs/20-14_ISTE_Standards-T_PDF.pdf. Accessed 16 June 2018.

[9] NGSS Lead States. *Next generation science standards: For states, by states*. National Academies Press, Washington DC, 2013.

[10] Lederman, N G., Antink, A. and Bartos, S. "Nature of science, scientific inquiry, and socio-scientific issues arising from genetics: A pathway to developing a scientifically literate citizenry." *Science & Education,* 23. 2014.

[11] Yao, Y., Zhong, N. and Zhao, Y. "A conceptual framework of data mining. In Data Mining: Foundations and Practice". Springer Berlin Heidelberg, 2008.

[12] Lederman, J. S., Lederman, N. G., Bartos, S. A., Bartels, S. L., Meyer, A. A. and Schwartz, R. S. "Meaningful assessment of learners' understandings about scientific inquiry—The views about scientific inquiry (VASI) questionnaire." *Journal of Research in Science Teaching,* 5165-83. 2014.

[13] Piatetsky-Shapiro, G. *Advances in knowledge discovery and data mining* (Vol. 21). U. M. Fayyad, P. Smyth and R. Uthurusamy (Eds.). Menlo Park: AAAI Press, 1996.

[14] Bond, J. and Zhang, M. "The impact of conversations on fourth grade reading performance - What NAEP Data Explorer tells? *European Journal of Educational Research*, 6 (4). 407-417. 2017.

[15] Su, R., Rounds, J. and Armstrong, P. I. "Men and things, women and people: A meta-analysis of sex differences in interests." *Psychological Bulletin*, 135859-884. 2009.

[16] Wang, M. T., Eccles, J. S. and Kenny, S. Not lack of ability but more choice: Individual and gender differences in choice of careers in science, technology, engineering, and mathematics. *Psychological Science,* 24 (50). 770-775. 2013.

[17] Legewie, J. and DiPrete, T. A. "The high school environment and the gender gap in science and engineering." *Sociology in Education,* 87. 259-280. 2014.

[18] Dee, T. S. "How a teacher's gender affects boy and girls." *Education Next*, 69-75. 2006a.

[19] Dee, T. S. "Teachers and the gender gaps in student achievement." *Journal of Human Resources,* 42 (3). 528-554. 2006b.

[20] Escardibul, J. O. and Mora, T. "Teacher gender and student performance in mathematics -Evidence from Catalonia (Spain)." *Journal of Education and Training Studies,* 1. 39-46. 2013.

[21] Singh, R. and Mukherjee, P. 'Whatever she may study, she can't escape from washing dishes: Gender inequity in secondary education – evidence from a longitudinal study in India." *Compare: A Journal of Comparative and International Education*, 48 (2). 262-280. 2017.

[22] Hiller, V. "Gender inequality, endogenous cultural norms and economic development." *The Scandinavian Journal of Economics,* 116. 455-481. 2014.

[23] Heddleston, K. *Social norms and gendered expectations*. 26 March 2015. https://kateheddleston.com/blog/social-norms-and-gendered-expectations. Accessed 16 June 2018.

[24] National Center for Education Statistics (NCES). *Assessment*. 2018a. https://nces.ed.gov/surveys/. Accessed 16 June 2018.

[25] National Center for Education Statistics (NCES). *U.S. History Assessment*. 2018c. https://nces.ed.gov/nationsreportcard/ushistory/. Accessed 16 June 2018.

[26] National Center for Education Statistics (NCES). *NAEP Data Explorer*. 2018b. https://www.nationsreportcard.gov/data_tools.aspx. Accessed 16 June 2018.

[27] Klecker, B. M. "*NAEP fourth-, eighth-, and twelfth-grade reading scores by gender: 2005, 2007, 2009, 2011, 2013*." Knoxville: Mid-South Educational Research Association. 2014.

[28] Klecker, B. M. and Klecker, R. L. "*Impact of student calculator use on the 2013 NAEP twelfth-grade mathematics assessment.*" Knoxville: Mid-South Educational Research Association. 2014.

[29] Cohen, J. "Statistical power analysis for the behavioral sciences." Hillsdale, Lawrence Erlbaum Associates, 1988.

[30] Becker, L. A. *Effect Size Calculators*. 2001a. www.uccs.edu/~lbecker/. Accessed 16 June 2018.

[31] Becker, L. A. *Effect Size (ES)*. 2000b. www.bwgriffin.com/gsu/courses/edur9131/content/EffectSizeBecker.pdf. Accessed 16 June 2018.

[32] National Center for Education Statistics (NCES). *The Nation's report card*. www.nationsreportcard.gov/about.aspx. Accessed 16 June 2018.

[33] Su, R., Rounds, J. and Armstrong, P. I. "Men and things, women and people: A meta-analysis of sex differences in interests." *Psychological Bulletin*, 135. 859-884. 2009.

[34] Bottia, M. C., Stearns, E., Mickelson, R. A., Moller, S. and Valentino, L. "Growing the roots of STEM majors: Female math and science high school faculty and the participation of students in STEM." *Economics of Education Review,* 14-27. 2015.

[35] Van Roekel, D. "*Parent, family, community involvement*." 2008. Washington DC, National Education Association. www.nea.org/assets/docs/PB11_ParentInvolvement08.pdf. Accessed 16 June 2018.

[36] NEA Education Policy and Practice Department. 2008. "*Parent, family, community involvement in education*." District of Columbia: National Education Association. www.nea.org/neapolicybriefs. Accessed 16 June 2018.

[37] Kini, T. and Podolsky, A. "*Does Teaching experience increase teacher effectiveness? A review of the research.*" Palo Alto, Learning Policy Institute. 2016.

[38] Sharobeam, M. M. "The variation in spatial visualization abilities of college male and female students in STEM fields versus non-STEM Fields." *Journal of College Science Teaching,* 46. 93-99. 2016.

[39] Hudley, C. "*Education and urban schools*." American Psychological Association. www.apa.org/pi/ses/resources/indicator/2013/05/urban-schools.aspx. Accessed 16 June 2018.