

Utah State University

DigitalCommons@USU

All Graduate Theses and Dissertations

Graduate Studies

5-2019

Empathy in the Middle-School History Classroom: The Effects of Reading Different Historical Texts on Theory of Mind, Empathetic Concern, and Historical Perspective-Taking

Jared P. Collette
Utah State University

Follow this and additional works at: <https://digitalcommons.usu.edu/etd>



Part of the [Education Commons](#)

Recommended Citation

Collette, Jared P., "Empathy in the Middle-School History Classroom: The Effects of Reading Different Historical Texts on Theory of Mind, Empathetic Concern, and Historical Perspective-Taking" (2019). *All Graduate Theses and Dissertations*. 7433.
<https://digitalcommons.usu.edu/etd/7433>

This Dissertation is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Theses and Dissertations by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.



EMPATHY IN THE MIDDLE-SCHOOL HISTORY CLASSROOM: THE EFFECTS
OF READING DIFFERENT HISTORICAL TEXTS ON THEORY OF MIND,
EMPATHIC CONCERN, AND HISTORICAL PERSPECTIVE-TAKING

by

Jared P. Collette

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Education

Approved:

Suzanne H. Jones, Ph.D.
Major Professor

Kathleen Mohr, Ph.D.
Committee Member

Jeff Nokes, Ph.D.
Committee Member

Kerry Jordan, Ph.D.
Committee Member

Nicole Pyle, Ph.D.
Committee Member

Richard S. Inouye, Ph.D.
Vice Provost for Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2019

Copyright © Jared P. Collette 2019

All rights reserved

ABSTRACT

Empathy in the Middle-School History Classroom: The Effects of Reading Different
Historical Texts on Theory of Mind, Empathic Concern, and
Historical Perspective-Taking

by

Jared P. Collette, Doctor of Philosophy

Utah State University, 2019

Major Professor: Suzanne Jones, Ph.D.

Department: School of Teacher Education and Leadership

Theoretical and empirical evidence indicate a possibility that reading certain types of historical texts could improve theory of mind (ToM), empathic concern (EC), and historical perspective-taking (HPT). The objective of this study was to compare the effect of reading a collection of primary documents in comparison to a historical narrative on ToM, HPT, and EC for adolescents in an eighth-grade history class. Students were randomly assigned to read either a historical narrative or a collection of adapted historical documents with approximately the same length, estimated Lexile score, and Flesch-Kincaid grade level. This study controlled for student comprehension scores, ToM scores, estimated amount of reading frequency, gender, and age. Afterwards, students were assessed on ToM, EC, and HPT using age-appropriate and valid measures.

The results demonstrated no statistical difference for individuals assigned to read either text as measured by ToM, EC, and HPT. Individuals with higher standardized

comprehension scores in the historical document group were more likely to read for a longer period of time than individuals with higher comprehension scores in the narrative group. Empathic emotions for the narrative group were significantly correlated with higher HPT. The researcher argues that better ToM assessments need to be developed for adolescents and the relationship of reading historical texts and empathy for adolescents should be a topic of future research.

(208 pages)

PUBLIC ABSTRACT

Empathy in the Middle-School History Classroom: The Effects of Reading Different
Historical Texts on Theory of Mind, Empathic Concern, and
Historical Perspective-Taking

Jared P. Collette

Theoretical and empirical evidence indicate a possibility that reading certain types of historical texts could improve different constructs of empathy that include theory of mind (ToM), empathic concern (EC), and historical perspective-taking (HPT). The objective of this study was to compare the effect of reading a collection of primary documents in comparison to a historical narrative on ToM, HPT, and EC for adolescents in an eighth-grade history class. Students were randomly assigned to read either a historical narrative or a collection of adapted historical documents with approximately the same length, and reading level. This researcher controlled for student comprehension scores, ToM scores, estimated amount of reading frequency, gender, and age. Post reading, students were assessed on ToM, EC, and HPT using age-appropriate and valid measures.

The results demonstrated no statistical difference for individuals assigned to read either text as measured by ToM, EC, and HPT. Individuals with higher comprehension abilities in the historical document group were more likely to read for a longer period of time than individuals with high comprehension abilities in the narrative group. Empathic emotions for the narrative group were significantly correlated with higher HPT. The

researcher argues that better ToM assessments need to be developed for adolescents and the relationship of reading historical texts and empathy for adolescents should be a topic of future research.

ACKNOWLEDGMENTS

I would like to thank Dr. Suzanne Jones for her help, direction, guidance, and support. I would like to thank my children, my mother, my wife's parents for their love and support, my siblings, and my nieces and nephews for their love, support, and prayers. Most of all I would like to thank my wife, who deserves this dissertation as much as I do. Her hard work in a variety tasks made it possible for me to labor solely on this project.

Jared P. Collette

DEDICATION

For Nisha

CONTENTS

	Page
ABSTRACT.....	iii
PUBLIC ABSTRACT	v
ACKNOWLEDGMENTS	vii
DEDICATION	viii
LIST OF TABLES	xi
LIST OF FIGURES	xiii
CHAPTER	
I. INTRODUCTION.....	1
Social and Emotional Development among Adolescents.....	1
Social and Emotional Learning	3
Empathy Defined.....	4
Empathy and Reading	6
Activating Empathy by Reading History	8
Historical Texts: Account and Traces	9
Gaps in the Literature	12
Summary	12
Brief Description of Study	13
II. REVIEW OF THE LITERATURE.....	15
Social and Emotional Learning	16
Adolescence.....	18
Clinical Definitions of Empathy.....	27
How We Engage in Theory of Mind	28
How We Engage in Empathic Concern.....	30
How We Engage in Historical Perspective-Taking.....	35
Reading-Empathy Hypothesis	56
Activation of Theory of Mind through the Reading of Historical Texts.....	70
Gaps in the Research Literature	71
Summary	73
III. METHODOLOGY	75
Study Design	75
Participants	77

LIST OF TABLES

Table	Page
1. Potential Similarities of Literary Fiction and Historical Documents.....	71
2. Consent Forms Received or Denied from Student Participants.....	78
3. Comparison of All Three Texts	82
4. Coh-Metrix Components for Two Texts.....	82
5. Emotions Included in Each Emotion Group	86
6. Correlations and Cronbach's Alpha Scores for Four Emotion Groups	87
7. Exploratory Factor Analysis of Six Empathic Emotions, Upset and Glad	88
8. Examples of Text-Specific Comprehension Questions	92
9. Procedures Table.....	93
10. Procedures for Day One.....	94
11. Procedures for Day Two	95
12. Means, Standard Deviations, and Pearson Correlations Among All Instruments.....	102
13. Coefficient Values for Three Steps of a Regression on the Faces Test	104
14. Coefficient Values for Three Steps of a Regression on Empathic Emotions	110
15. Coefficient Values for Three Steps of a Regression on Historical Perspective-Taking	112
16. Correlations for the Historical Perspective-Taking Components, Word Count, and Empathic Emotions	113
17. Phi Test with the Faces Test and Select Authors	118
18. Phi Test with Empathic Concern and Select Authors.....	119

Table

Page

19. Means and Standard Deviations for Adolescent Males and Females	120
--	-----

LIST OF FIGURES

Figure	Page
1. Study hypothesis.....	14
2. Logic model to justify the need for empathic intervention	26
3. Theoretical model of the reading-empathy hypothesis	61
4. Scatterplot of the Faces Test (y-axis) and standardized time reading times (x-axis).....	105
5. Scatterplot of the Faces Test (y-axis) and HPT scores (x-axis)	106
6. Scatterplot of the Faces Test (y-axis) and SRI scores (x-axis).....	107
7. Scatterplot of the Faces Test (y-axis) and SRI and Time Reading Scores (x-axis).....	108
8. Scatterplot of empathic emotions (y-axis) and SRI and Reading Times (x-axis).....	110
9. Scatterplot of SRI (y-axis) and Time Reading (x-axis) for the Narrative Group.....	115
10. Scatterplot of SRI (y-axis) and Time Reading (x-axis) for the Document Group.....	115

CHAPTER I

INTRODUCTION

Social and Emotional Development among Adolescents

Adolescence is an important period for moral, social, and emotional development (Blakemore, 2008; Merrell & Gimpel, 2014). The prefrontal cortex, a crucial location for moral and social cognition, experiences a period of significant neural development during adolescence (Blakemore, 2008). Prior to adolescence, individuals appear to base moral decisions on hedonism and social pressure (Eisenberg, Carlo, Murphy, & Van Court, 1995). However, with the cognitive developments of adolescence, such as increased logic and abstract thought, individuals demonstrate greater tendency to rely on higher levels of moral reasoning such as universal rules, personal emotions, self-reflection, and taking others' perspectives (Eisenberg et al., 1995). Despite the neurological development of adolescence, moral development is not an automatic result of maturation. Although adolescents have the capacity to develop morally, such moral development needs to be fostered through socialization, modeling, and discussions (American Psychological Association [APA], 2002; Bandura, 1999; Blakemore, 2008; Eisenberg, 2003; Eisenberg & Valiente, 2002; Hoffman, 2000).

Research has shown that adolescents are not being raised with the socialization, modeling, and discussions necessary to provide robust moral development (Putnam, 2000, 2015; Turkle, 2015). For example, many adolescents are being raised today during a time of declining social institutions (e.g., participation in bowling leagues, and political

parties; Putnam, 2000), decreasing religious participation (Smith & Snell, 2009), greater wealth inequality (Putnam, 2015), increasing numbers of single-parent families (Curtin, Ventura, & Martinez, 2014; Pew Research Center, 2015; Putnam, 2015), and increasing numbers of both parents working (Putnam, 2015). Further, researchers have demonstrated that today's adolescents have significantly more exposure to media (Common Sense Media, 2016 Rideout, Foehr & Roberts, 2010; Turkle, 2015, especially violent media (Anderson, Bushman, Donnerstein, Hummer, & Warburton, 2015; Common Sense Media, 2016; Gildemeister, 2016; Parent Television Council, 2013). These factors provide reasonable evidence to be concerned about how the current generation is being raised.

The way that many adolescents are now being raised may be negatively influencing their social and emotional development. Twenge (2006) argues that adolescents now entering adulthood have significant deficits in their social and emotional skills and understanding. In a widespread longitudinal study, Smith, Christoffersen, Davidson, and Herzog (2011) found that individuals entering adulthood in the U.S. have demonstrated high levels of nihilism and moral relativity. In addition, in comparison to previous birth cohorts, many college students and transitioning adults today have lower levels of empathy (Konrath, O'Brien, & Hsing, 2011), higher levels of narcissism (Twenge & Campbell, 2009; Twenge, Konrath, Foster, Campbell, & Bushman, 2008), practice lower levels of volunteerism and civic engagement (Gioia, 2008; Smith et al., 2011; Twenge, 2006, and have greater focus on materialism without the corresponding willingness for hard work (Twenge & Kasser, 2013). Perhaps some of these challenges

can be addressed by the public education system through social and emotional learning programs.

Social and Emotional Learning

This study seeks to apply a Social Emotional Learning (SEL) framework (Collaborative for Academic Social and Emotion Learning [CASEL], 2015). The aim of SEL is to help students obtain the interpersonal, social skills, and moral understanding that would help them to avoid at-risk behaviors, thrive, achieve success, and be happy (CASEL, 2015). CASEL targets five domains for SEL, namely: self-management, self-awareness, social awareness, relationship skills, and responsible decision making. By applying these critical competencies, adolescents should be able to calm themselves when angry, create and maintain friendships, make and keep goals, appropriately resolve conflicts, make ethical and safe choices, and make positive contributions to their communities (CASEL, 2015).

One of the five core competencies of SEL stated above is social awareness, which is defined as “the ability to take the perspective of and empathize with others from diverse backgrounds and cultures” (CASEL, 2015, p. 5). Although only in the description of social awareness does CASEL explicitly include the words ‘empathy’ and ‘perspective-taking,’ research has also connected empathy with other domains of SEL including relationship skills, responsible decision-making, and self-awareness (Batson, 2011; Bloom, 2016; Decety & Jackson, 2004; Hoffman, 2000). Therefore, four out of the five domains for SEL are directly or related to empathy. In summary, SEL provides a

solid framework to promote the moral development of students, and empathy is an essential component of SEL.

Empathy Defined

Empathy is an ambiguous concept with various definitions (Batson, 2011). This study will focus on three constructs related to empathy.

1. Theory of mind (ToM), also called mind reading, cognitive empathy, or mentalizing, which is the ability to understand the thoughts and feelings of another person (Batson, 2011; Decety & Jackson, 2004).
2. Empathic concern (EC), which is compassion felt for another person (Batson, 2011),
3. Historical perspective-taking (HPT), which is the ability to understand the perspective of people in the past (O. L. Davis, Yeager, & Foster, 2001).

In practicing ToM, we have to infer other people's thoughts and feelings through facial, vocal, verbal, situational, or other cues. Humans are deeply social. The ability to understand the thoughts and feelings of others presumably grants humans a Darwinian advantage, allowing us to cooperate, communicate, and raise our young in ways dramatically different from other life forms (Young, 2012). Practicing ToM allows us to outsmart rivals and opponents, work collaboratively with friends, family and peers, and is a crucial part to enjoying literature and history (Bloom, 2016; Young, 2012).

EC is similar to compassion and entails a person's feeling *for* another, rather than feeling *with* another. To induce EC, Batson (2011) had participants engage in perspective-taking by reading a personal account of a struggling individual. In over 30 experiments, Batson found that individuals, when motivated by EC, were more likely to help a victim, even when exiting the situation was easy and when there was not a

discernable personal or social reward.

Historical Perspective Taking is understanding a person's perspective while considering the person's historical context (O. L. Davis et al., 2001; Endacott & Brooks, 2013). Historical context includes the social, political, and cultural influences of the time, as well as the historical events prior and during the event in question. Historical theorists emphasize that people in history had different epistemological, ontological, and existential assumptions (Jenkins, 1991). Perspective-taking is understanding another person's experiences, values, beliefs, and positions and is often a difficult thing to do in the present (Keysar, Barr, Balin, & Brauner, 2000). HPT has the added burden of bridging the chasm of time.

The application of HPT is impossible to perfectly achieve. We cannot completely understand how people in history thought and felt. However, in seeking to understand the times and places of people far different from ourselves, we may obtain a glimpse of radically different life experiences and viewpoints (Endacott, 2014). Such glimpses help students understand the socially constructed nature of society, which, in turn, may build respect and understanding for different perspectives in the present and ultimately allow students to be able to step back and take a critical view of their own perspective (Barton & Levstik, 2004). Historical empathy and perspective-taking are often explicitly stated in state and national standardized skills and expectations (e.g., Common Core State Standards [CCSS], 2010; Utah State Board of Education [USBE], 2016). For example, the USBE standards for social studies requires students to apply "multiple perspectives" (USBE, 2016, p. 5), "various perspectives" (USBE, 2016, p. 11), "historical perspective"

(USBE, 2016, p. 42), and to “develop and demonstrates values...such as...empathy” (USBE, 2016, p. 2). The Common core ELA standards require students to “understand other perspectives and cultures” (CCSS, 2010, p. 7) and practice “multiple perspectives” (CCSS, 2010, p. 49). Whether for Social and Emotional Learning, social studies, or literacy, there is a clear overlap in the requirement to help students develop empathy.

ToM, EC, and HPT may be essential skills and dispositions for moral, social, and civic reasons. As a result, it is important for schools to design lessons that help students frequently apply the skills and dispositions of ToM, EC, and HPT. Having students read, comprehend, and analyze historical texts may provide an excellent opportunity to develop these skills and dispositions.

Empathy and Reading

Theorists and empirical researchers have argued that reading certain types of literature may help increase ToM and EC. For example, some theorists have argued that literature can work as a gateway to help readers understand diverse perspectives and grant them a sense of shared humanity (Mar & Oatley, 2008; Nussbaum, 1997).

Additionally, literature may help us accomplish the famous injunction by Harper Lee (1960/2014) to “climb in [another’s] skin and walk around in it” (p. 35). Pinker (2011) argues that the rise of widely available inexpensive novels during the nineteenth century was a crucial catalyst for heralding an era of normative nonviolence. The philosopher Nussbaum (1995, 1997, 2004) argues that literature helps us understand the perspectives of people dramatically different from ourselves. She explains that gaining others’

perspective through literature humanizes people far different from ourselves and can transform a student into a compassionate *citizen of the world* where a human life of one's own ethnicity, tribe, or nationality is not considered to be of greater value than another (Nussbaum, 1995, 1997, 2004).

Recent studies have begun to grant these claims some empirical credence. For example, several studies have found that individuals who frequently read certain types of literature, such as romance fiction or literary fiction, on average, score higher on ToM assessments (Djikic, Oatley, & Moldoveanu, 2013; Fong, Mullin, & Mar, 2013; Kidd & Castano, 2013, 2016; Mar, Oatley, Hirsh, dela Paz, & Peterson, 2006; Mar, Oatley, & Peterson, 2009; Panero et al. 2016). Four separate studies, employing randomized control designs, found evidence for a causal effect of reading on ToM, whether it was an immediate temporary effect from reading short literary fiction stories (Black & Barnes, 2015a; Kidd & Castano, 2013; Kidd, Ongis, & Castano, 2016), or a long-term effect from reading an entire book of literary fiction over the span of a week (Pino & Mazza, 2016).

Theorists have provided plausible explanations for why literary fiction appears to activate ToM. Kintsch (1998) explains that through the process of text comprehension, our mind constructs a situation model, or a mental representation of the meaning of the text. This is accomplished through a construction phase that activates a wide variety of potential connections. As meaning is constructed, the mind then deactivates irrelevant connections. Mar and Oatley (2008) explicitly connect Kintsch's model of comprehension with their argument that in reading fiction, one constructs social situation models that have application in the real world. They argue that reading fiction works as a

type of practice or simulation for social intelligence (Mar & Oatley, 2008).

The literary theorist, Zunshine (2006), argues that literary fiction often contains significant social complexity, requiring readers to make challenging inferences and engage in multiple layers of nested minds by reading about a character who is imagining a mind of a person who is also imagining a mind. Kidd and Castano (2013, 2016) argue that literary fiction avoids simple stereotypes, such as good guys versus bad guys, and instead often includes ambiguous characters acting in challenging moral situations.

Activating Empathy by Reading History

If ToM is activated through certain types of literature as it appears to be, then perhaps ToM can also be activated through the investigation of historical documents. Similar to literary texts, historical texts often require readers to engage in the mental construction of socially complex situation models of ambiguous characters acting in challenging moral situations (e.g., McCullough, 2001).

Moreover, historical texts may also affect EC. In several empirical experiments, individuals who were instructed to take the perspective of an individual while reading or listening to a first-person account of personal hardships, self-reported high levels of EC (Batson, 2011). The investigation of people in history can humanize them and elicit emotional responses in the present. Barton and Levstik (2004) refer to this as *empathy as care*. This emotional care for others appears to motivate investigation of the past (Endacott & Pelekanos, 2015). Even Pulitzer Prize-winning historians have reported feeling deep care for the people that they study (McCullough, 2003; Stack, 2017). There

is evidence to expect an empathic emotional response from reading certain historical texts.

Reading historical texts may also affect HPT. Historical texts are often written from various perspectives, each with their own positions, perspectives, and goals (Wineburg, 2001). Therefore, in order to fully comprehend a historical text, the reader must engage in the perspective of the author. This is why history education researchers often seek to promote HPT through the analysis of historical documents (Endacott & Brooks, 2013, 2018). Meier (2010) found that students who received high scores in applying historical heuristics of analyzing historical documents also scored high in historical empathy.

Historical Texts: Account and Traces

Seixas and Peck (2004) divide historical texts into *traces* or *accounts*. Traces are fragments of evidence that have to be contextualized, such as an arrow head or a court transcript. Accounts provide a version of the story of what happened with more detailed “narratives and explanations” (Seixas & Peck, 2004, p. 110). Accounts can be a first-hand eyewitness account, such as a memoir, or it can be a secondary account assembled from a wide collection of evidence.

Perhaps one of the most ubiquitous forms of accounts in the history classroom is the history textbook. History textbooks may provide background information about an event that builds on student understanding in a more coherent accumulative manner (Collette, 2012). However, history textbooks have received significant criticism (Paxton,

1999; Ravitch, 2003; Wineburg, 2001). History textbooks typically provide an unambiguous monolithic perspective, free from interpretation and alternative perspectives (Paxton, 1999; Wineburg, 2001). Although it may appear that textbooks are written from an impartial voice, the creation of a single narrative out of such a diversity of perspectives is an act of controversial political power (Paxton, 1999; Wineburg, 2001).

Additionally, history textbooks are typically void of in-text citations, with no discussion about how the invisible authors reached their conclusions. Furthermore, textbooks employ a unique *textbookese* vocabulary and syntax (Paxton, 1999) that is often in passive tense to remove culpability from historical agents (Rockmore, 2015). In addition, they are typically censured and approved by a committee to ensure political correctness (Ravitch, 2003) and often employ incomprehensible generalizable facts and figures, such as the death of 50 million people in World War II.

Such a text may be less likely to promote ToM, EC, and HPT. A singular narrative, void of multiple perspectives, interpretations, and sources, appears to provide little opportunity for students to construct an understanding of authors' or historical individuals' thoughts, feelings, intentions, and perspectives. Furthermore, humans tend to have a much greater capacity to empathize with individuals rather than large abstract groups (Hoffman, 2000). In empathizing for an individual, humans can often employ induction, by generalizing the individual's experience to a large group of people. It is much more challenging for humans to exercise empathy through deduction, by learning about a large abstract group and understanding the individual experience. History textbooks tend to take a deduction approach, while primary sources, through letters,

pictures, and journal entries often allow a more inductive approach.

In addition, the sanitized, politically correct passive language, with unimaginable facts, of a textbook may be less effective in promoting an emotional response or EC for the individuals involved in historical events. Finally, the unambiguous language of textbooks combined with the lack of perspectives, can possibly discourage readers from constructing their own interpretations (Kidd & Castano 2013, 2016). Being acculturated with textbooks, lectures, and multiple-choice assessments, many adolescent students apparently believe that the practice of history is memorization of lots of facts (Nokes, 2013; Wineburg, 1991).

In contrast, many education researchers argue that students should learn to think like historians (Nokes, 2013; Shanahan & Shanahan, 2008; Wineburg, 2001). In order to read like a historian, students need to apply the heuristics or interpretive skills of the historian, including sourcing, corroboration, and contextualization (Nokes, 2013; Wineburg, 2001). Furthermore, students need texts that provide multiple perspectives, including first-hand accounts, that contain sometimes contradictory information, and that require readers to bridge gaps, make more interpretive claims, and construct knowledge.

This study employed two texts.

1. *A narrative text*: A recently written children's storybook, that is similar in nature to a textbook account, in that it shares a single narrative of the Salem witch crisis in unambiguous language without in-text references or alluding to how the author reached her conclusions.
2. *Historical documents*: A collection of primary sources, more fragmentary or trace-like in nature pertaining to the Salem witch crisis. They do not explain the story but contain excerpts of important documents connected to it, such as a sermon, a diary entry, a trial transcript, two letters from prisoners, and a public confession decades later. They are modified and abridged to be at about the same length and reading level as the narrative text.

This study compared the effect of the narrative text versus the collection of historical documents on students' ToM, EC, and HPT.

Gaps in the Literature

The current study focused on addressing four gaps from the empirical literature. First, in empirical investigations of the effect of reading on ToM, participants have almost always been adults recruited through college campuses, *The New York Times*, and Amazon Turk (Djikic et al., 2013; Fong et al., 2013; Kidd & Castano, 2013, 2016; Mar et al., 2009; Panero et al., 2016). Adolescence is a crucial time for social development and schools are important sources of socialization. This study will be implemented with adolescents in the context of a public middle school located in the Western U.S.

Second, although researchers have investigated the effect of several different reading genres and even different television shows on ToM (Djikic et al., 2013; Fong et al., 2013; Kidd & Castano, 2013, 2016; Mar et al., 2009; Panero et al., 2016), no studies have measured the effect of reading historical documents on ToM. Third, previous studies have not measured the effect of historical documents on EC. Fourth, previous research has not investigated a causal relationship of reading historical documents on HPT.

Summary

In summary, young adults today demonstrate declining levels of empathy and growing levels of narcissism (Konrath et al., 2011; Twenge & Campbell, 2009; Twenge

et al., 2008). Adolescents would likely benefit from more instruction that engages in social and emotional learning (SEL). SEL can be delivered in tandem with academic instruction especially in areas that overlap. For example, both SEL and history academic standards require empathy and perspective-taking. Four empirical studies (Black & Barnes, 2015a; Kidd & Castano, 2013, 2016; Pino & Mazza, 2016) provide evidence that the reading of certain types of literature increases ToM. One correlation study shows a connection with individuals analyzing historical texts and historical empathy (Meier, 2010) while one other study provides correlational evidence of historical empathy and social perspective-taking (Gehlbach, 2004). Researchers have not included adolescent participants nor used primary documents in assessing for the effect of reading on ToM. We do not know if a historical narrative or a collection of historical documents would produce a greater effect in ToM, EC, or HPT.

Brief Description of Study

The purpose of this study was to investigate the effect of reading different historical texts on ToM, EC, HPT for adolescents. The research question for the study was: To what extent do different types of historical texts influence adolescents' ToM, HPT, and EC? The hypothesis of this study was that the reading of historical documents would activate theory of mind, empathic concern, and historical perspective-taking in comparison to a historical narrative. The hypothesis is illustrated in **Error! Reference source not found.**

Students enrolled in an eighth-grade history class were invited to participate in

this study. Participants were randomly assigned to either read a single historical narrative

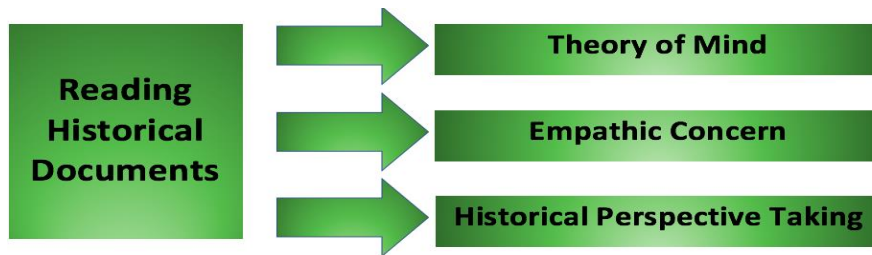


Figure 1. Study hypothesis.

text or a collection of historical documents, which will just be referred to as historical documents. Both texts pertain to the Salem witch crisis. After reading the texts, students were assessed in ToM through a non-self-reported assessment, EC through a self-reported survey, and HPT through a written historical paragraph scored by blind judges with master's degrees in history using a rubric.

The data analysis controlled for several potentially confounding factors including estimated reading comprehension ability, estimated ToM ability, estimated reading exposure, self-reported emotions, birth sex, and age. The analysis compared readers of the different texts in regard to their scores for ToM, EC, and HPT.

CHAPTER II

REVIEW OF THE LITERATURE

The topic of this study intersects with various disciplines including psychology, sociology, neurology, literacy, education, and history. The multidisciplinary nature of this study prohibits the literature review from being comprehensive. It would be impossible to summarize all the information from each of the domains related to empathy, history, and reading. As a result, the researcher used EBSCO and Google Scholar to attempt to provide a comprehensive review on all published empirical studies that show the effect of reading on ToM, EC, and on HPT. Other information is included to provide context, a theoretical framework, and circumstantial evidence.

The author begins this literature review with a discussion of the importance of social and emotional learning. He argues that adolescents today face unique challenges and experience empathic deficits that make an intervention focusing on empathy especially important. He then defines three constructs of empathy: theory of mind, empathic concern, and historical perspective taking and summarizes relevant research and theory pertaining to the practice of each. Next, the author discusses the theoretical literature and empirical evidence pertaining to the reading-empathy hypothesis. He then explains that, based on the theoretical literature pertaining to the reading-empathy hypothesis, there is a possibility that historical texts could increase empathy. Finally, the author closes with a discussion of the research gaps pertaining to the use of adolescent participants, certain historical texts, and a public-school context.

Social and Emotional Learning

This study relies on a SEL framework. CASEL defines SEL as:

the process through which children and adults acquire and effectively apply the knowledge, attitudes, and skills necessary to understand and manage emotions, set and achieve positive goals, feel and show empathy for others, establish and maintain positive relationships, and make responsible decisions (CASEL, 2015, p. 5).

SEL targets a set of personal characteristics, skills, and dispositions that help students develop into successful contributing members of society. SEL programs seek to prevent negative behaviors such as drug abuse, violence, and dropping out, and increase positive behaviors such as establishing healthy relationships, succeeding academically, and making and keeping goals (CASEL, 2015). CASEL seeks to accomplish this ambitious agenda by targeting five personal, interpersonal, and social domains for SEL, namely: self-management, self-awareness, social awareness, relationship skills, and responsible decision making. By instilling these domains in children, SEL ultimately seeks to help children progress from being morally dependent on external incentives to becoming morally autonomous individuals who act according to internalized values, beliefs, and compassion (Bear & Watkins, 2006).

Schools should make the development of social and emotional skills a greater priority for public education because social and emotional skills can be successfully taught, and they are predictive of a wide variety of important student outcomes (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011). For example, social and emotional skills have a greater impact on relationships, personal thriving, and academic and career success, than does I.Q. alone (Goleman, 2005). Social and emotional competencies are

crucial to help children understand and relate to each other (Hartup, 1983), achieve interpersonal success and acceptance from peers (Asher & Taylor, 1981), obtain and maintain a job (Berrueta-Clement, 1984), avoid criminal behavior (Berrueta-Clement, 1984), and complete a college degree (Savitz-Romer & Bouffard, 2012).

Advocates of SEL point out that a singular focus on academic achievement does not address the needs of much of the student population, who may be struggling with social and emotional challenges (CASEL, 2003; Durlak et al., 2011; Payton et al., 2008). Educational achievement not only requires intelligence but several other personal characteristics such as self-discipline, motivation, and interpersonal skills. Schools can be an avenue to prepare students with practical skills for success including the ability to obtain and hold a job, to participate civically, and to have healthy relationships.

Social and emotional learning programs have been associated with increased learning outcomes by an average of 11% (Durlak et al., 2011). Additionally, well-designed and effectively implemented SEL programs can significantly improve mental health and positive youth development, as well as decrease the likelihood of antisocial and aggressive behavior, depressive symptoms, drug use, and problem behaviors (Catalano, Berglund, Ryan, Lonczak, & Hawkins, 2002; Greenberg, Domitrovich, & Bumbarger, 2001; Horowitz & Garber, 2007; Tobler et al., 2000; Wilson & Lipsey, 2007; Zins, Weissberg, Wang, & Walberg, 2004). In a cost-benefit analysis, Belfield et al. (2015) found that SEL can decrease public expense and increase public revenue, which could potentially provide an additional million dollars to government budgets per 100 students over the span of a lifetime. Multiple meta-analyses of the effect of SEL within

the past decade have found positive effects on a variety of outcomes for emotionally disturbed populations (Cook et al., 2008), general K-8 populations (Payton et al., 2008), and general secondary student populations (Durlak, et al., 2011).

However, studies indicate that schools are perhaps not sufficiently addressing the importance of SEL. For example, Benson (2006) surveyed 150,000 6th to 12th graders nationally, and found that only 29-45% reported practicing empathy, decision making, and conflict resolution. Additionally, only one-third of the students reported thinking that their school provided a caring and encouraging environment. In a separate study, Benson, Scales, Leffert, and Roehlkepartain (1999) found that a high percentage of students had insufficient social-emotional competence, perceive their teachers as being indifferent to their welfare, and engage in behavior that undermines the education of classmates.

In summary, the above argues that adolescents need SEL, SEL programs have demonstrated effectiveness that can potentially provide dramatic benefits to society, and schools are doing a less-than-adequate job at providing SEL.

Adolescence

Adolescent Development

Adolescence is a period after childhood and before adulthood where individuals become physically and sexually mature, increase in sexual libido, and acquire more skills preparatory for adulthood (APA, 2002; Eccles et al., 1993; Feldman & Elliot, 1990).

Adolescents become increasingly autonomous, more fully develop primary and secondary sex characteristics, and experience a significant shift in their relationships

(Eccles et al., 1993; Feldman & Elliot, 1990). Adolescence comes with many stereotypical challenges. Adolescent hormonal changes are associated with increased aggression, sexuality, and mood swings (Buchanan, Eccles, & Becker, 1992). Adolescents experience the highest arrest rate out of all age groups, and many adolescents engage in risky behavior such as substance abuse (Eccles et al., 1993). Early adolescence sometimes marks a downward spiral towards dropping out of school (Eccles et al., 1993). Furthermore, parent-adolescent relationships often undergo challenges as adolescents seek autonomy away from parents and towards greater conformity and acceptance among peer groups (Eccles et al., 1993).

In addition, adolescents often experience more demands with sports, school, and work (APA, 2002). Secondary education usually includes an increasingly difficult curriculum and attendance at a larger school with multiple teachers, which often comes with less personalization, more anonymity, and increased risks (APA, 2002; Eccles et al., 1993). SEL may help schools retain struggling students and help students perform at higher levels.

Adolescent Brain Development

Adolescence is a crucial period for neurological development, especially in regard to empathy. Adolescence brings benefits and advantages that include an increased cognitive ability for abstract thought, hypothetical situations, applying multiple dimensions of a problem, a deeper understanding of self, and using more sophisticated information processing strategies (Keating, 1990). This greater capacity for thought as well as a greater focus on peers often results in a greater interest and capacity in

understanding the internal processes of others (Eccles et al., 1993; Hoffman, 2000).

Neurological studies have shown that the prefrontal cortex, which is a key location for perspective-taking, undergoes dramatic development during adolescence (Blakemore, 2008). No other age experiences more dramatic changes to the prefrontal cortex as adolescence except for infancy (Blakemore, 2008). The onset of puberty is accompanied by significant growth of neural connections, also termed as grey matter, in the prefrontal cortex. These connections allow for a much wider variety of neural networks, and greater plasticity, but this increase in grey matter also slows down cognition (Blakemore, 2008). Blakemore explains that adolescents experience increasing gray matter in their prefrontal cortex, which hinders their ability for perspective-taking.

For example, Dumontheil, Apperly, and Blakemore (2010) found that in a perspective-taking task, children improved significantly from ages 7.3 to 11.5. However, children did not statistically improve between the ages of 9.8–13.9, although a significant improvement was again found from ages 14 to 17.7. This indicates that the dramatic growth in grey matter in the prefrontal cortex during adolescence may slightly hinder adolescent perspective-taking cognition.

In the years following adolescence, the growth of grey matter is followed by a period of growth of white matter. White matter is myelin, which is formed by supporting glial cells that massively speed up and insulate neural transmissions. Moreover, the connections that are not used or rarely used get pruned during this period. By the end of adolescence, the prefrontal cortex has fewer connections and more myelination, which decreases its plasticity and speeds up cognition (Blakemore, 2008). The prefrontal cortex

is purportedly the center of decision making, inhibition, planning, and judging consequences, but in regard to this study, the prefrontal cortex is also a crucial place for perspective-taking and ToM (Blakemore & Choudhury, 2006). This dramatic development of the prefrontal cortex in adolescence may explain why adolescents often have lower abilities at perspective-taking and facial recognition than adults (Carey, Diamond, & Woods, 1980; Dumontheil et al., 2010).

The relative malleable adolescent prefrontal cortex makes adolescence a crucial period for moral, social and emotional development (Blakemore, 2008). The experiences that are provided during adolescence may engender certain habits of mind and capacities that become embodied and more difficult to change in the brain circuitry. This means that adolescence poses a brief window of greater plasticity that becomes much more immutable after adolescence. For this reason, it is important to provide an abundance of high-quality experiences for adolescents that can help develop their ability for perspective-taking and exercising ToM (Blakemore, 2008).

Challenges of Adolescents Today

The dramatic developmental changes during adolescence necessitate an intentional focus on the moral, social and emotional education of this age group (APA, 2002; Bandura, 1999; Blakemore, 2008; Eisenberg, 2003; Eisenberg & Valiente, 2002; Hoffman, 2000). Adolescents are being raised during an especially challenging period in history that includes declining social capital (Putnam, 2000, 2016), fracturing families (Bramlett & Mosher, 2002), and the ubiquity of media (Rideout et al., 2010), especially violent media (Anderson et al., 2015). These factors may be undermining adolescents'

development for empathy.

Today's youth are being raised at a time in America where there is significant evidence of deteriorating social capital (Putnam, 2000; Wilcox, Cherlin, Uecker, & Messel, 2012). Social capital is one's network of social connections that can often translate directly into economic opportunities. Americans attend about half as many public meetings as they did in the 1960s (Putnam, 2000). Families eat dinner together less often and self-reported trust in strangers has declined rapidly from above 50% to less than a third today (Putnam, 2000). Since 1970, church attendance has fallen from almost 60% to below 40% (Levine & Kawashima-Ginsberg, 2017). Americans today demonstrate significantly less civic and community participation than in the past (Gioia, 2008; Putnam, 2000, 2015). It is likely that this drastic decline in quality social experiences would lead to less socially active and less empathetic youth.

Americans are raising their children within fractured families. Out of wedlock birth rates have increased from below 5% in 1960, to more than 40% in the present (Bramlett & Mosher 2002; McLanahan & Percheski, 2008). Single parents typically have fewer financial resources and less time to invest in their children (McLanahan & Percheski, 2008). Children growing up in single-parent homes are at greater risk than children raised by their married biological parents for a host of negative outcomes including: dropping out of high school, not attending college, experiencing behavioral and psychological challenges, and becoming sexually active and become pregnant at an early age (McLanahan & Percheski, 2008).

One of the most unique aspects of today's youth is that they are being raised with

the universal presence of media. It is estimated that children, 8-18 years-old, spend an average of 7.5 hours each day engaged in media (Rideout et al., 2010). This is equal to about 30 hours of movies and television, 17 hours of music, 10 hours on a computer and 8 hours of video games each week (Rideout, 2015; Rideout et al., 2010). About a third of the time of their media engagement was with more than one form of media (Rideout et al., 2010). This is likely more time than devoted to any other task besides sleeping (Rideout et al., 2010). Some adolescents report feeling anxiety when they are separated from their smartphone for a short time (Common Sense Media, 2016). Some opponents argue that this easy access to information and constant stimulation may transform the rising generation into more 'shallow' thinkers (Carr, 2011). Through countless hours of qualitative interviews Turkle (2011, 2015) explains that the youth's constant engagement with video screens and earphones often results in superficial conversations, lower social maturity, greater discomfort with being alone, and less authentic friendships.

Not only are American youth constantly engaged in media, but a growing amount of media includes violent content. Gildemeister (2016) found that violence in television programming had increased dramatically in quantity and intensity and was increasingly linked to sexual content. Top selling video games are primarily violent in nature and each year increase in graphics and appearance of reality, which also makes the violence more impactful on the game player (Anderson et al., 2015).

In several meta-analyses, involving hundreds of thousands of participants, researchers have shown that violent media is causally linked to violence (Anderson et al., 2015). Even a small exposure of media violence results in an immediate increase in

aggressive cognition and a decline in empathy and actions that helps others. Longitudinal studies have shown that the cumulative effect of violent media exposure causes individuals to become more violent as measured by physical and verbal violence (Anderson et al., 2015). Studies have linked media violence to assault, intimate partner violence, robbery, and gang fighting (Anderson et al., 2015).

There is evidence that these factors of declining social capital, fracturing family structure, increased media use, and increasing violent media, are adversely affecting the social and emotional competencies of adolescents. For example, Konrath et al. (2011) performed a meta-analysis on a frequently administered empathy survey, the Interpersonal Reactivity Index (M. H. Davis, 1983). They found 72 published administrations of the survey given primarily to incoming college freshmen ($N = 13,737$) between 1979 and 2009 and found that EC had decreased by 48% and perspective-taking had decreased by 34%. This is likely not just a phenomenon happening to college students but appears to be widespread for all adolescents and emerging adults. In a separate study, O'Brien, Konrath, Grühn, and Hagen (2013) analyzed three widespread surveys, two of which were representative samples of Americans between ages 18 and 90 years old. They found that empathy is lowest among the youngest adults and increases gradually until about age 60 when it begins to slightly decline, forming an inverted u-shape pattern. The researchers give this as evidence that the age cohort reaching adulthood, sometimes termed as transitioning adults or emerging adults, are significantly lower in empathy than previous birth cohorts.

In addition, Twenge et al., (2008) performed a longitudinal meta-analysis of the

40 self-reported questions from the Narcissistic Personality Inventory from 1979 to 2006 given to college students ($N = 16,475$) in several different studies and found that narcissism had increased by about 30%. Individuals who are high in narcissism tend to lash out at others when offended and seek to use others for their own purposes (Twenge & Campbell, 2009). Research provides evidence of an inverse relationship with narcissism and empathy (Watson, Biderman, & Sawrie, 1994; Watson & Morris, 1991).

Not only are young adults demonstrating changing levels of empathy and narcissism, they also appear to be attracted to moral relativism and nihilism. Through hundreds of in-depth interviews of a nationally representative sample of transitioning adults ages 18-25, Smith, Christoffersen, Davidson, and Herzog (2011) found that this population sample demonstrated poor moral reasoning. When asked simple questions about what is right and wrong, transitioning adults often could not provide clear answers. Some equated morality with law or defined something as moral if it made one feel good. Some argued that a behavior was morally permissible as long as it did not directly harm another individual (therefore stealing from a government or a corporation was acceptable) and insisted that morality was a personal preference. Many respondents openly admitted their lack of consideration of morality. Smith et al. (2011) pointed out that the participants in their study demonstrated very immature moral development according to Kohlberg's theory of moral development (Kohlberg, 1971). The participants tended to be civically disengaged and tended to be much more interested in materialism, individualism, sex, and intoxication (Smith et al., 2011).

The combination of these various studies tells a persuasive story: Adolescence is a challenging transitional period of social, neurological, hormonal, emotional, and physical development. Adolescents should be provided with high-quality instruction, modeling, and practice of SEL (CASEL, 2015). Instead, today's adolescents are facing much greater challenges with unstable families for many, declining social capital, and the ubiquity of increasingly violent media. In consequence of these challenges, emerging adults demonstrate lower levels of empathy (Konrath et al., 2011), higher levels of narcissism (Twenge & Campbell, 2009) materialism, declining religious participation, and a growing allegiance towards relativism and nihilism (Smith et al., 2011). These are key reasons why public schools should try to counteract these influences through academic lessons that are compatible with SEL. Most especially, lessons should focus on increasing perspective taking and empathy (CASEL, 2015) to build moral, social, and emotional awareness. The above argument is demonstrated graphically in Figure 2.

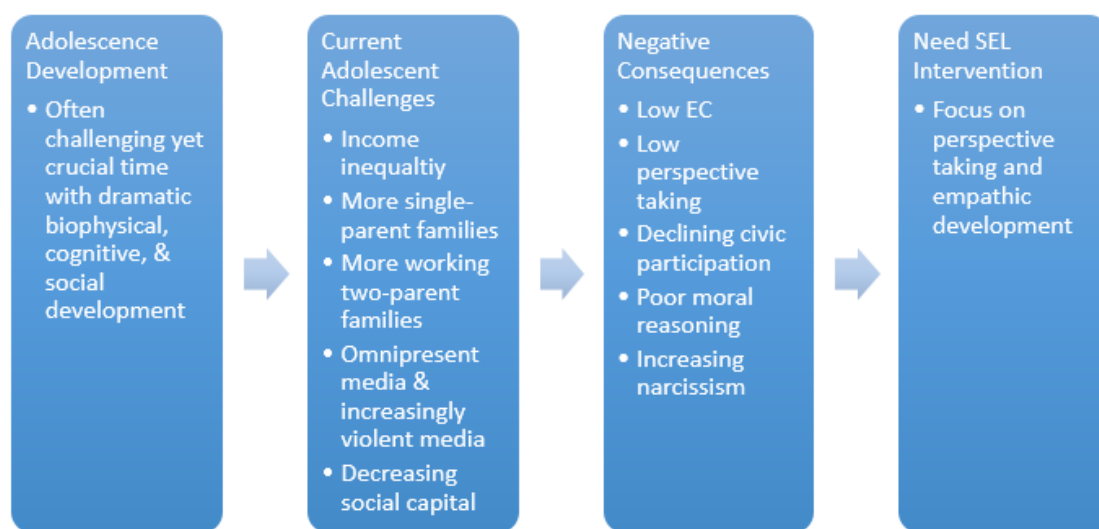


Figure 2. Logic model to justify the need for empathic intervention.

Clinical Definitions of Empathy

Empathy has been called “notoriously ambiguous” (Zahavi & Overgaard, 2012, p. 3) with almost as many definitions as there are researchers (Batson, 2011; Bloom, 2016). The original word for empathy, *Einfurling*, was a German invention at the turn of the Twentieth Century that meant projection of oneself into something, usually an inanimate object, and was used for artists to imagine what it would be like to be the rock, fence, or blade of grass they were painting (Zahavi & Overgaard, 2012). Over time, the English equivalent of empathy has evolved significantly from its German origin to become a topic of intense social scientific interest. However, researchers face the challenge of creating a uniform construct of the term. This study focused on three constructs related to empathy:

1. Theory of mind (ToM): The ability to perceive the thoughts and feelings of another (Batson, 2011; Decety & Jackson, 2004; Frith & Frith, 2005; Zahavi & Overgaard, 2012). It overlaps with a myriad of terms that includes perspective-taking, social information processing, cognitive empathy, social intelligence, mind reading, and mentalizing (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001; Donahue, 2014; Hoffman, 2000 Mar & Oatley, 2008). ToM is the ability to infer other people’s thoughts and feelings because we do not have direct access to other people’s minds (Decety & Jackson, 2004).
2. Empathic concern (EC): A feeling of compassion or care for another person, but not necessarily feeling the same emotion as the other person (Batson, 2011; Hoffman, 2000)
3. Historical perspective-taking (HPT): The ability to take a historical person’s perspective, especially within the historical context (Endacott & Brooks, 2013; 2018; VanSledright, 2004; Wineburg, 2001).

These next three sections will discuss the research for how ToM, EC, and HPT are practiced.

How We Engage in Theory of Mind

In the last few decades, new technological tools such as functional magnetic resonance imaging (fMRI) have enabled researchers to understand much better the brain's inner workings in regard to ToM. While interpreting others' thoughts and feelings it appears that we use our own minds and bodies as reference tools (Decety & Jackson, 2004). In order to perceive what a person making a certain facial expression is feeling, the perceiver may need to activate a particular neural network by employing the neurons associated with the same facial expression. This phenomenon, often referred to as motor mimicry, may involve the use of mirror neurons (Decety & Jackson, 2004). Advocates of mirror neurons argue that mirror neurons are automatically activated when one perceives another performing an action. The same neural networks that are required for the action of the one performing the action, are activated in the perceiver (Decety & Jackson, 2004). By making the same facial expressions of another, or at least simulating the neural networks of the same facial expressions, our bodies create an emotional feedback response and we begin to feel what the perceived person feels (Decety & Jackson, 2004; Hoffman, 2000; Preston & de Waal, 2002). The extent of these claims of motor neurons are controversial and not universally accepted (Bloom, 2016).

Previous research demonstrates that an essential connection exists between action and perceiving (Decety & Jackson, 2004). We act to perceive and we perceive to act, and we do both within our bodies. This means that we often cannot fully perceive another person without activating neural networks that are associated with the action. For example, when we perceive a person grasp something with a hand, the neural networks

associated with our hand are activated (Decety & Jackson, 2004). Importantly, however, we do not lose the self-other distinction, meaning that we do not mistake the other person's action as our own (Decety & Jackson, 2004).

Neurological studies using fMRI have demonstrated significant neurological overlap in personally experiencing pain and perceiving another person experience pain (Bloom, 2016; Decety & Jackson, 2004). These studies typically investigate the neural reaction of a person being hurt by being pricked, shocked, or burned, and then they observe the neural reaction of a person perceiving another person being pricked, shocked or burned. Such neurological studies demonstrate that there is significant overlap in the neurological processes of a person actually experiencing pain and a person perceiving another experiencing pain (Bloom, 2016; Decety & Jackson, 2004). The perceiver's mind, to a degree, imitates the neural networks as if it was directly experiencing the pain (Decety & Jackson, 2004). There appears to be some limited sense in which perceiving pain in another is experiencing the pain (Bloom, 2016). Advocates of empathy argue that this recognition of others' pain can help prevent the perpetration of harm and motivate people to alleviate the suffering of others (Bandura, 1999; Batson, 2011; Hoffman, 2000).

Using our own bodies and minds as references to understand others poses some challenges. Flavell (1977) argued that in a sense, our own emotions, thoughts, and beliefs are constantly ringing in our ears so loudly that it is difficult to suppress them to understand the perspective of another. Research has demonstrated that people think they are judged, noticed, and remembered much more than they actually are (Decety & Jackson, 2004). They often assume that if they know something or believe something,

then others know and believe the same (Decety & Jackson, 2004). For example, in one study, Van Boven and Loewenstein (2003) asked participants how thirsty they would be if they had been lost in the woods for days without food and water. Participants who were asked right after vigorous exercise without water rated the amount of thirst as much higher than those who had not recently exercised.

In other words, engaging in empathy by using our own mind and body is a problematic activity. Often the experiences of others are different from what we would imagine if we were in their place. As a result, transcending our own understanding to truly understand another's perspective may be one of the greatest challenges in practicing ToM, whether in the present, through literature, or in the study of history across the chasms of time and place.

How We Engage in Empathic Concern

Religious and Philosophical Traditions of Empathic Concern

Many ancient philosophers, religious leaders, and current day psychologists argue that empathy—defined as EC or compassion—is crucial for morality (Armstrong, 2006; Batson, 2011; Hoffman, 2000; Nussbaum, 1997). Empathy, compassion, and love are central tenets to various ancient faith and philosophical traditions. For example, Confucius sought to promote submission to the traditional rites and values that placed loyalty, altruism, and empathy at the center of morality (Armstrong, 2006) to replace the egoism and selfish ambition that was tearing China apart. In ancient Athens, a major purpose of the theater was to achieve catharsis, vicariously experiencing the suffering of

the protagonist, who may be a woman, a slave, or even an enemy (Armstrong, 2006; Nussbaum, 1997). The Hindu master Mahavira promoted universal compassion for all creatures, and Buddha abandoned a life of pleasure and dedicated his life to share his message, which he believed would alleviate suffering (Armstrong, 2006). The Jewish scriptural canon is full of empathy-inducing accounts that often focus on the less privileged brother, the lower-in-status polygamous wife, the stranger, the orphan, or the widow (Sacks, 2015). According to the Christian *New Testament*, Jesus intended Christian love or compassion as the defining characteristic for his followers: “By this all men will know you are my disciples, if you love one another” (John 13:35, New International Version). According to Armstrong (2006), most prominent world faith traditions, including Islam, Christianity, Judaism, Buddhism, and Hinduism, were built on this tradition of love or compassion.

This tradition appears to have also influenced European enlightenment philosophers including David Hume, and Adam Smith (Watson, 2005). Adam Smith (1759/1817) developed a theory of morality that was centered on the concept of sympathy, which is very similar to the contemporary use of EC. He argued that as people perceive the situation of others they use their imagination to understand what they are feeling, and explained that people experience joy and sorrow through the joy and sorrow of others. He believed that people are motivated not just by self-love, but by love of others to engage in actions that help others (Roberts, 2014). These faith and philosophical traditions continue to impact society’s views of healthy moral and social development in the present.

Empathic Concern Motivates Prosocial Behavior

Empathy is a crucial moral attribute, because empathy motivates prosocial behavior. Over the last century, many theorists (e.g., Rand, 1964; Skinner, 1938; Spencer, 1897), have made the claim that any prosocial behavior could always be traced back to an egoistic motive such as to obtain a positive social reward or to remove personal distress at seeing someone in suffering (Batson, 2011). However, Batson demonstrated, in over 30 different experiments over the span of his career, that when people feel EC they seek to help the victim regardless of positive social recognition or of personal distress. Moreover, when feeling EC, they are not satisfied unless they perceive that the victim is relieved (Batson, 2011). Batson defined this causal relationship as the empathy-altruism hypothesis.

From a completely different theoretical perspective, Bandura (1999) also provides evidence supporting the empathy-altruism hypothesis. Bandura explains that moral agency has to be disengaged for people to commit inhumanities. There are several mechanisms that disengage moral agency, one of which is dehumanization. Dehumanization is a process of defining an ethnic or religious group as less than human, or removing all empathic attachment (Bandura, 1999). Historians have noted the process of dehumanization that often precedes inhumane cruelty (Walker, Turley, & Leonard, 2008). Whether in the Holocaust, the Rwandan genocide, or the genocide of the Khmer Rouge in Cambodia of urbanites, mass murder was preceded by intense propaganda that labeled entire groups as *untermensch*, cockroaches, saboteurs, feudal, or parasites (e.g. Levi, 1986).

Similarly, Bandura, Underwood, and Fromson (1975) found that when participants were administering what they believed to be punishments to individuals, they were more likely to use more punitive punishments to dehumanized individuals. They concluded that dehumanization appears to make individuals treat others with greater severity (Bandura et al., 1975). In contrast, engaging in perspective-taking of individuals may be a powerful antidote against dehumanization. Batson, Chang, and Orr (2002) found that a participant taking the perspective of an individual in a stigmatized group led to feeling EC for the stigmatized individual, which undermined the participant's negative perception of the entire stigmatized group. This change in attitude increased the participant's willingness to help the stigmatized group.

How Empathic Concern is Activated

Hoffman (2000) explains that it is very difficult for individuals to feel compassion for large abstract groups of people. For example, a large statistic like two million refugees fleeing Syria, appeared to be ineffectual in promoting charitable donations, but a single photo of a three-year-old child's lifeless body, with his face in the sand, prompted an outpouring of donations and interest in the Syrian Refugee Crisis (Slovic, Västfjäll, Erlandsson, & Gregory, 2017). People typically feel compassion for an individual, especially one that they perceive as being blameless (Hoffman, 2000). EC is often evoked through personal first-hand accounts, but then it can be generalized to a large group of people (Batson, 2011; Hoffman, 2000). In other words, EC does not seem to be activated very well through deduction, by hearing large-scale statistics, such as the death of 50 million in World War II, but rather by induction, by understanding someone's personal

experience and generalizing to a similar group of people. Perhaps this is why taking the perspective of an individual from a stigmatized group can improve attitudes and helpful behavior toward the entire group (Batson et al., 2002). Lamm, Batson, and Decety (2007), applied behavioral and fMRI measurements with adult participants (ages 18-31) and found that taking the perspective of a victim while listening to the victim's personal account was associated with the activation of regions of the brain associated with EC.

In various experiments, Batson (2011) activated empathy in individuals by instructing the participants to take the perspective of the victim while listening, watching, or reading a first-person account. However, the condition of taking the perspective of a victim only explained some of the variance. Batson explains that personal values may have a more effectual impact on why people do or do not feel EC. It appears that focusing on the condition of others is ultimately a personal choice that is based on what people consider to be most important. In other words, just because one has the capacity to practice empathy does not mean that they will always employ that ability.

Perspective-taking through first-person narratives may work as what Hoffman defines as "inductions" (2000, 2008). The use of inductions is a useful strategy that parents often use to engender empathy in their children by directing the children's attention towards the hurt feelings of another (Hoffman, 2000, 2008). When a child harms another person, parents often attempt to help the child realize how his/her behavior affected the victim. Certain types of texts often work like inductions. Certain types of texts place attention on the victims and allow the reader to view the detailed inner, emotional, and psychological consequences of hurtful actions.

How We Engage in Historical Perspective-Taking

History is a Foreign Country, as one book title quips (Lowenthal, 1985). As a result, when studying history, students encounter bewildering human practices that defy present-day morals and logic, such as the obese beauty standards of the stone-age fertility statuette Venus of Willendorf, the Aztec practice of human sacrifice, the Spartan performance of infanticide, and the medieval execution of neighbors as witches. Students often subscribe to two ultra-simplified explanations for such behavior: people in the past were (1) not as intelligent and (2) not as moral as we are today (Barton, 2008; Barton & Levstik, 2004; Endacott, 2010; Lee & Ashby, 2001). Barton (2008) reports from one study that demonstrated that American elementary students employed a deficit paradigm to explain everything about the past. According to his elementary participants, everything in the present is better, not just technology, but present-day morals, fashion, even people's names (Barton, 2008). Similarly, Lee and Ashby performed a study that showed that British students, ages 7-14 explained harsh Roman punishment and Anglo-Saxon trials primarily as "stupid, ignorant, or morally defective" (p. 44). This way of thinking appears to be our automatic default way of understanding other viewpoints in history. According to Wineburg (2010), "'Presentism' – the act of viewing the past through the lens of the present – is not some bad habit we've fallen into, but is instead our psychological condition at rest, a way of thinking that requires little effort and comes naturally" (p. 91). Escaping our narrow presentist mindsets poses a difficult challenge.

Nevertheless, as one further investigates the strange practices of the past, one discovers that practices of the past were typically influenced by different, but strict, sets

of moral values and different, but often rigorous, modes of reasoning (Watson, 2005).

Indeed, most people from history would likely view many practices in the *present* as bizarre, illogical, and even morally reprehensible. The key to studying history is that one should seek to understand people on their own terms and then to see how their experiences can relate to the present (Barton & Levstik, 2004). To accomplish such a herculean task requires the application of historical empathy.

Professional Historians Are Motivated by Historical Empathy

For many historians, historical empathy appears to be the motivating force to their research and writing. For example, McCullough (2001), in his Pulitzer Prize-winning biography *John Adams*, depicts a complex character who defies idolization or demonization but captures heroic attributes within a flawed, often-insecure human. In a lecture, McCullough argued that in writing about historical figures...

[E]mpathy is essential.... Who were those people? What was it like to have been alive then, in their shoes, in their skins? Of what were they afraid? What didn't they know? ... [I]t is not possible to understand what happened in that tumultuous, protean time [of the American Revolution] without knowing and understanding [the people that experienced it]" (McCullough, 2003).

After 8-years of intense research on polygamist Mormon women, another Pulitzer Prize winning historian, Ulrich expressed feeling "so much empathy for the people who did really remarkable things in terribly difficult circumstances with the power of faith" (Stack, 2017). McCullough and Ulrich are two Pulitzer-prize winning historians whose works provide examples that demonstrate how some types of history can be motivated by and powerful depictions of empathy.

Historical Empathy as a Pedagogical Practice

Despite being an important part of the practice of history, historical empathy has not become a pedagogical concept until fairly recently, beginning in Britain in the 1970s and 1980s (Endacott & Brooks, 2018). Endacott and Brooks explain that in the period after World War II, many believed that history education in Britain was experiencing a pedagogical crisis. British history education was perceived as extremely boring and irrelevant as students were expected to memorize and recite massive blankets of information of Anglo-centric history. These concerns led to the Schools History Project of 1972 (SHP), which sought to transition the curriculum to a greater focus on the skills and dispositions of historians, such as interpretation and logical reasoning in an attempt to contribute to students as well-rounded individuals. One significant aspect of this new national curriculum included the idea that students should practice historical empathy in order to make them more humanely educated (Endacott & Brooks, 2018). As a result, historical empathy became a major focus of research and debate pertaining to British history education (Lee & Ashby 2001).

Two Types of Historical Empathy

Similar to clinical empathy, historical empathy is a highly ambiguous and contested term (Brooks & Endacott, 2013; O. L. Davis et al., 2001). However, many history education researchers distinguish two main types of empathy: an affective or emotional historical empathy, often termed as empathy as care, and a cognitive historical empathy, often termed as HPT (Barton & Levstik, 2004; Brooks & Endacott, 2013).

Historical empathy as care. In engendering historical empathy as care, history

teachers tend to focus on the plight of oppressed groups such as women, workers, children, or racial, ethnic, religious, or sexual minorities (Brooks, 2009). Empathy as care often entails the student feeling compassion and perhaps indignation at the treatment of a persecuted individual or group. Such lesson plans can include images, such as Lewis Hines' photographs of children working in factories and first-person accounts such as the *Narrative of Frederick Douglass* (Douglass, 1845). Often, lessons that focus on empathy as care have students try to identify with the historical characters. Some teachers use activities where students write about the historical characters in the first person (e.g., Endacott, 2010). Advocates of empathy as care seek to affirm the humanity of all people (Barton & Levstik, 2004; Endacott & Brooks, 2013).

Historical perspective-taking. In contrast to empathy as care, many emphasize a more rational approach with greater emotional distance and no self-identification with the historical person, termed as historical perspective-taking (HPT; O. L. Davis, 2001). HPT is different from historical empathy in that it is primarily “rational, intellectual, and concerned with explaining actions, attitudes, and concepts which are alien to our own” (Boddington, 1980, p. 18). Foster (2001) argues that practicing HPT does not entail creative imagination, identification, or sympathetic emotions. Rather, HPT seeks “to understand historical characters’ frame of reference without trying to identify or sympathize with his or her feelings” (Yeager & Foster, 2001, p. 15).

The primary purpose in applying HPT is to understand why people behaved the ways that they did in the past (Yeager & Foster, 2001). This is different from history as care, because we do not want students to permanently adopt the potentially racist or

violent views of historical peoples. Rather, we encourage them to briefly entertain historical agents' goals, intentions, beliefs, and values and knowledge of the world, in order to better understand why they behaved the way they did. HPT does not seek to rationalize people's behavior but to understand people's behavior on their own terms. For example, Doppen (2000) had students engage in an analysis regarding President Truman's attitude toward dropping the atomic bomb on Japan at the end of World War II. This lesson encouraged students to not identify with Truman or adopt Truman's negative attitudes towards the Japanese, but rather to understand his perspective, reflect on it, evaluate the evidence, and decide for themselves whether Truman made the right decision regarding the dropping of atomic bomb.

Four Components of HPT

Based on the literature of leading experts (O. L. Davis et al., 2001; Endacott & Brooks, 2013; 2018), there are four main components of HPT: (a) perspective taking, (b) historical context, (c) inference, and (d) evidence.

Perspective-taking. Perspective-taking entails the attempt to understand other view-points, including an attempt to understand moral values, ways of reasoning, knowledge of the world, goals, and beliefs (O. L. Davis et al., 2001). Again, the primary goal for perspective-taking is to seek to understand why people behaved the way that they did without resorting to simplistic labels such as racist, evil, or stupid.

Historical context. In order to practice HPT, one needs a thorough, rigorous understanding of the historical context (O. L. Davis, 2001; Endacott & Brooks, 2018; Foster, 2001; Huijgen, van de Grift, Van Boxtel, & Holthuis, 2017; Wineburg, 1998).

The historical context includes the spatio-temporal context, related to the events before, during and after, the geo-spatial context, which relates to location, and the social context, which includes the social, cultural, and political milieu of the time (Huijgen et al., 2017; Nokes, 2013; Wineburg, 1991). Wineburg (1998) also included the linguistic, rhetorical, biographical, and historiographic as part of the context. In regard to Truman's decision for dropping the atomic bomb, Yeager and Foster (2001) asked "What were the sociopolitical pressures on him? Did public opinion play a role? What were the customs, values, and conventions of the era?" (p. 14).

Understanding the historical context is advantageous because "students who acquire relevant contextual understandings are better placed to understand why people in the past acted as they did" (Foster, 2001, p. 172). Similarly, O. L. Davis (2001) explains that "students become increasingly able to engage empathy better as they have more historical knowledge" (p. 5). Without understanding the context, then trying to understand historical agents' perspectives amounts to inventing fiction.

Inference. It is impossible to completely fill in the historical context with all the complexities, subtleties, and detail that perfect understanding would require. Usually there is insufficient historical evidence and when there is bounteous evidence, a person simply lacks the time and cognitive resources to absorb it all. Inevitably, historians have to resort to the application of HPT to make inferences and to "fill in the gaps." This component can be labeled 'inference' because several researchers explicitly discuss the importance of empathy in allowing people in the present to infer the thoughts and feelings of people in the past.

According to O. L. Davis (2001), empathy is applying “imagination constrained by evidence” (p. 4). As explained by Yeager and Foster (2001), historical empathy “helps the historian bridge the gaps in what is known; [through] some ability to *infer* [emphasis added] from given knowledge an explanation of certain actions” (Yeager & Foster, 2001, p. 14-15). Ashby and Lee (1987) argue that HPT rests on reasoned evidential reconstruction that is also “broadly *inferential* [emphasis added]” (p. 63). Endacott and Brooks (2018) also explain that “Historical empathy enables the *inferential* [emphasis added] thinking and imaginative reconstruction needed to explain why people believed and acted as they did in the past” (p. 211). These gaps can be inferred from other documents, from our own lives, and from a shared sense of humanity.

Evidence is a fourth component of HPT. All work in history requires evidence. This is the link that connects history with the reality of the past and makes it distinct from fiction. These four components of HPT form the foundation of the HPT rubric, which is explained and portrayed in Chapter Three.

Historical empathy as affect and cognition. Although, empathy as care and perspective-taking are often distinguished, some researchers advocate that they should not be separate activities and that students should do both simultaneously (Barton & Levstik, 2004; Endacott, 2010). For example, Endacott had participants write essays in the first person as if they were a historical figure making a difficult decision that contradicted personal values, such as when Jefferson had to decide whether to purchase the Louisiana Territory, even though Jefferson did not believe that it was strictly constitutional. Endacott coded the essays for discussion of principles, opportunity, fear of

failure, relationships, and responsibilities and demonstrated that students could achieve emotional and cognitive empathy through the use of reading primary and secondary texts. Crucially, however, his historical empathy lesson plan used scaffolding with small-groups discussions, with interaction from the teacher, and questions that accompanied the historical documents.

Teaching Historical Empathy Through Historical Texts

Researchers recommend that historical empathy lesson plans be “embedded in the historical method” (Yeager & Foster, 2001, p. 14). Crucial components of a historical empathy lesson plan include: (a) background information, (b) a variety of historical sources, (c) central debatable investigative questions, and (c) the opportunity for students to analyze, discuss, and produce their own conclusions regarding the sources and class discussions (Doppen, 2000; Endacott & Brooks, 2013; Kohlmeier, 2006; Nokes, 2013).

Wineburg (1991) points out that a large ‘breach’ exists between how history is practiced and how it has traditionally been taught. Nokes (2013) argues that, in science, students engage in science labs, in physical education they play sports, in mathematics they solve real mathematical problems, but in history, students typically listen to teachers lecture, read the textbook, and then answer multiple-choice factual questions. In contrast, real historians analyze primary and secondary sources using key heuristics (Nokes, 2013; Stahl & Shanahan, 2004; Wineburg, 1991).

A primary source is an original source and can include a first-hand narrative through a letter or diary, a photograph, a census record, or some other historical artifact.

A secondary source is a source (about a historical event by a person who was not there) that is ostensibly based on primary sources. Primary sources, as the original sources of history are often considered to be more reliable sources of evidence. On the other hand, a meticulously researched and well-written secondary source that is based on a preponderance of primary sources is often more historically reliable than any single primary source (Barton, 2005).

Wineburg (1991) identified three key heuristics or skills that historians apply when analyzing historical sources: sourcing, contextualization, and corroboration. Historians almost always look to the source of the document first. The source includes the author, the date, and the type. Sourcing includes taking-into-account the “why” of the source, recognizing that many historical sources are created for intentional reasons (Stahl & Shanahan, 2004). For example, Neumann (2010) encourages his students to see texts as ‘speech acts’ and specifically asks students to think about what the text is *doing*. In addition, historians attempt to understand historical sources by placing them within their historical contexts. Contextualization, includes understanding the cultural, social, economic, and linguistic differences of a time period (Nokes, 2013; Wineburg, 1991), which is the same skill discussed above for HPT.

Finally, historians practice the skill of corroboration (Wineburg, 1991), which entails comparing documents with other documents, including what is similar, uniquely contradictory, and uniquely left out in a particular document compared to other documents. This is similar to the skill of intertextuality. For example, the Common Core standards require students to be able to make “an increasing number of

connections...between texts, considering a wider range of textual evidence” (CCSS, 2010, p. 8). With historical texts, one has to read very closely to keep track of who claims what. With historical heuristics, facts are never out in the world to be discovered, rather, every truth claim is tied to a specific source or sources (Wineburg, 1991).

Wineburg (1991) has documented how professional historians used these three heuristics to read across a collection of disparate accounts related to the Battle of Lexington. They almost always looked to the source first, even completely disregarding some accounts entirely based on the sources (such as a historical fictional account), they corroborated evidence, and they contextualized each of the accounts.

However, even after reading carefully and applying the heuristics, one cannot construct meaning solely by the information in the text. The text can never fully explain everything. There is not enough space on the page or enough time. Authors have to make assumptions about what the reader already understands and therefore authors leave out a significant amount of detail. For example, the primary accounts of the Battle of Lexington do not explain what a common is (an empty field in the center of town used for grazing) because they assume that their intended audience would already know. However, contemporary eighth-grade students living in the west do not usually know what a common is, for a common is now very uncommon. This is one example demonstrating greater cultural gaps that results in a greater distance from the time and place of historical documents. Every historical document contains a significant number of gaps that must be filled by inferences in order to construct comprehension. Students can infer from other documents, they can infer from their own historical background

understanding, and finally they can infer from their own common humanity.

Various researchers have employed historical document analyses for lesson plans targeting historical empathy (e.g., Doppen, 2000; Endacott, 2010; Kohlmeier, 2006), but as far as this researcher is aware, there is only one study that has quantitatively linked the practice of analyzing historical documents to stimulate historical empathy. In a study comparing historical heuristics with historical empathy, Meier (2010) developed a rubric to assess high school student essays on topics that included the Boston massacre, the Salem witch trials, Sherman's march to the sea for each of the three historical heuristics (sourcing, contextualization, and corroboration) and a rubric for historical empathy and compared the scores of students who had studied from traditional historical textbooks with students who applied the heuristics to analyze historical sources. Meier found that as students improved and learned these heuristic skills of sourcing, corroboration, and contextualization, they also increased in historical empathy. He also learned that students who learned from traditional textbooks, who were not taught the heuristics of history, actually declined in demonstrating historical empathy across the length of the study.

Meier points out that although we cannot be certain that applying heuristics to analyze historical texts causes an increase in historical empathy, "what cannot be denied is the importance of the use of primary source historical documents in conjunction with these heuristics involved in historical reasoning did have a positive impact upon historical empathy" (Meier, 2010, p. 69). Meier's study provides evidence that there is likely a link between practicing historical heuristics and exercising historical empathy.

In contrast to learning history through the analysis of historical documents, over

the last several decades many history curriculums have relied on a textbook to provide a grand historical narrative that is typically meant to be memorized and parroted back (Wineburg, 2001). As discussed in Chapter One, historical narratives in the classroom typically fail to provide alternative explanations, or perspectives (Paxton, 1999; Wineburg, 2001). They often explain the story of history without references, citations, or discussion about interpretation (Paxton, 1999; Wineburg, 2001).

The researcher hypothesized that a singular narrative, void of multiple perspectives, interpretations, and sources, would provide little opportunity for students to construct an understanding of the author's or historical individuals' thoughts, feelings, intentions, and perspectives. In addition, the sanitized, politically correct passive language may be less effective in promoting an emotional response or EC for the individuals involved in historical events, nor most effectively assist students to take the perspective of individuals in history.

The Paradox of Historical Empathy

Seeking to understand people in the past poses a paradox (Endacott, 2014; VanSledright, 2004). On one hand, the perceiver needs to recognize the common humanity of individuals in history. As fellow humans, we presumably share a range of similar emotions and thoughts. On the other hand, aspects of the past can be radically strange in comparison to the present (Endacott & Brooks, 2013; Wineburg, 2001). Many of the desires, beliefs, and perspectives of people in the past are shockingly different. This mix of similarity and difference poses a unique challenge. To fully engage in historical empathy, would require a complete understanding of the political, social,

cultural, and linguistic context, which, of course, is not possible. There are not enough historical sources. Moreover, even if there were, there would not be enough time to absorb them all. As a result, to understand historical events, we have to substitute to one degree or another, our own thoughts and feelings. For example, we can seek to understand people in the past by comparing our own recent personal experiences. However, the moment we apply our present perspective, thoughts, and feelings to understand the perspective of someone in the past, we automatically warp or soil their experiences to a degree.

Even if we had limitless historical sources and limitless time to study them, we would *still* likely misinterpret the sources at least to some degree. This is because our own goals, beliefs, and ways of looking at the world will interfere with the meaning of the historical evidence. VanSledright (2001) discusses the challenges of positionality, which includes who we are and with what we identify. Our positionality is influenced by “sociocultural, racial, ethnic, class, and gender components,” which will affect our “ontological (what’s my worldview), existential (who am I), and epistemological (how do I know) assumptions” (VanSledright, 2001 p. 57). This positionality likely effects the topics historians seek to research, the topics that history teachers choose to emphasize, and the topics that students choose to pay attention to, discuss, and study on their own.

Historians ostensibly bracket their positionality and ontological, existential, and epistemological assumptions, to ignore their own biases and to understand the perspective of the individuals that they are studying. This is a monumental task “because it means holding in mind whole structures of ideas that are not one’s own, and with which one

may profoundly disagree” (Ashby & Lee, 1987, p. 63). However, many argue that we cannot comprehend words, historical or present, without reference to our own sociocultural understandings (Jenkins, 1991; Lowenthal, 1985; VanSledright, 2001; 2004). The very means that enables us to comprehend the words on the page, distort the original meaning. Practicing historical empathy to make inferences, to fill in the gaps in the historical record, is inevitably based on our own views, beliefs, and experiences. VanSledright argues that “no matter how much we attempt contextualization...we can only approach the past from the standpoint and deportment of where we are now, from the inescapably historicized positions we presently hold” (p. 63).

Even ostensibly empirical evidence is still subject to interpretation, which is motivated by the present context (VanSledright, 2001). Even after historians meticulously analyze multiple pieces of evidence, apply the strictest heuristic standards, and try their best to avoid overt bias, we can never know for certain that their conclusions are correct because we cannot directly access the past and compare our conclusions to it (Jenkins, 1991).

Furthermore, literary theorists have emphasized that all texts are dialogic, or written in response or reaction to previous texts (Allen, 2011). Therefore, the exact meanings of texts are embedded in a sophisticated sociocultural context (Allen, 2011). Similarly, historical documents were created in reaction to a vast array of events, oral conversation, and other writings. Perusing a historical document, in a sense, is listening to a snippet of a complex conversation, without knowing exactly what was said right before or right after. The conversational exchange can be partially inferred, but can never

be fully recovered.

Notwithstanding these challenges, various history education researchers argue that practicing historical empathy is still a worthy ambition (Barton & Levstik, 2004; Brooks, 2009; Endacott, 2010; Gehlbach, 2004; Kohlmeier, 2006; Lee & Ashby, 2001; VanSledright, 2001). Despite launching a harsh critique about the challenges of historical empathy, VanSledright argues that teaching historical empathy is still “compelling” because it grants us self-knowledge by making us aware of our own positionalities (p. 66). Wineburg (2001) argues that practicing perspective-taking in history with people in distant times and places has the potential to facilitate understanding and communication in the present (Wineburg, 2001). If one can see the legitimacy in radically different perspectives of the past, perhaps they will be more adept at seeing the legitimacy of present positions on the other side of the political aisle. Ultimately, gaining an understanding of the historical context for people’s perspectives may help students realize that their own perspectives are influenced by the present context (Barton & Levstik, 2004; Endacott & Brook, 2013). Understanding the influence of the current historical context on our own viewpoints may empower us to transcend moral shortcomings of the present.

Extreme attacks on the practice of history are the result of oversimplification. They portray the practice of history “as an all-or-nothing endeavor. That is, historians either can absolutely embrace all the mores, values, perspectives, and ideas of a bygone age to fully appreciate and understand actions in the past, or they cannot” (Foster, 2001, p. 171). There is a wide spectrum of examples of HPT, from works that are blatantly

biased and stripped of their historical context in order to serve a present socio-political agenda (Wineburg, 2013), to works that are meticulously researched over periods of years that provide immense understanding about certain eras or events (Wineburg, 2010).

In addition, everything that has happened is now part of history. Therefore, even the words of the skeptics are part of history. As a result, there is a degree of hypocrisy by cynics who argue that it is impossible for others to be able to understand history. For the very act of publishing their critique demonstrates their assumption that they expect to be understood by a wide diverse audience even though their own words are now part of history.

The biggest assumption that HPT is based on, is the idea that despite radically different life experiences, sociocultural context, and positionalities, there must be some universal human element that transcends space and time, which makes people living millennia ago in various continents, comprehensible. Seixas and Peck (2004) call this “an historically transcendent human commonality” (p. 113). Empathy as care perhaps may only require an acceptance of all other humans as sentient beings, capable of experiencing pain. However, HPT requires much more in its attempt to understand the thoughts, beliefs, and motivations of other people. This study is based on the theoretical assumption that although language is imperfect and insufficient to fully communicate lived experiences, it is still possible to a degree, to transmute the historical perspectives of the past, into the present.

Historical Empathy and Agency

Central to the practice of historical empathy is the focus on action (Endacott &

Brooks, 2018; Yeager & Foster, 2001), specifically why people behaved the way they did. Such a focus requires an appropriate understanding of historical agency. Individuals were not predetermined by historical circumstance to behave in some inevitable fashion. Nor were individuals granted limitless agency to act in any way that may seem possible to a modern mind. An empathetic approach to history recognizes that people in the past were imperfect individuals with agency working within real limitations (Barton & Levstik, 2004). Historical writings have often been plagued by the perpetual challenge between seeing individual agency as the causal force in historical change or as seeing social, political, or economic structures as causal forces in historical change. For example, Carlyle (1841/1897) argued for a theory of history based on ‘great men’ with limitless agency, who forced their vision onto the world through tremendous will and effort. He explained, “The history of the world is but the biography of great men” (1841/1897, p. 13), and argued that “the Great Man was always as lightning out of Heaven; the rest of men waited for him like fuel, and then they too would flame” (p. 125). Carlyle believed that individuals could transcend their historical context to radically alter the direction of history.

On the other hand, many historians and philosophers of history, such as Hegel (Watson, 2005), resort to the extreme opposite of Carlyle’s (1841/1897) great man theory, citing a deterministic version of history, where social, cultural economic, or other structures take on indomitable force that destine human behavior to an inevitable end. An example of this is Marx and Engel’s (1867/2007) *Das Capital*, which argues that human society was inexorably pressing forward through different stages of development from

feudalism, to capitalism, to socialism, and ultimately to communism.

Either extreme—Carlyle’s man of limitless agency or Marxism’s historical inevitability with no human agency—fails to engage historical empathy. One sociological perspective that provides a more moderate balance between the extremes of determinism and human agency is Giddens’s (1984) structuration theory, which argues that all of society is constituted by structures, such as family, democracy, justice, and gender. All people inherit these structures, but they are not permanent inevitable structures. They both constrain and allow human action. These structures are top-down in the sense that people inherit them and they frame our actions. However, they are bottom-up in the sense that people are required to exercise their agency to perpetuate or alter such structures. This balance of inheritance and perpetuation, of structure and agency, allows what some have called “wiggle room” (Erickson, 2004; Parker 2011). Wiggle room is agency in the face of structure. It is an apt analogy of constrained yet possible human action. Historical empathy seeks to capture both the limitations and the agency found within real historical figures. Teaching students to practice historical empathy may be a means of granting students more agency in the face of social structures.

Historical Empathy: Quantitative versus Qualitative Methodology

This current study applies a quantitative approach to the investigation of ToM, EC, and HPT. However, leading proponents of historical empathy (Endacott & Brooks, 2018), criticize quantifiable measurements of historical empathy as “problematic” (p. 218) and “shortsighted” (p. 219). They argue that historical empathy does not follow a

specific developmental path and they critique past quantifiable assessments (e.g., Ashby & Lee, 1987) as Piagetian, which requires a series of developmental stages to fully realize the highest level of historical empathy. They also challenge the validity of simplistic self-report assessments of historical empathy (e.g., Greene, Kisida, & Bowen, 2014).

Furthermore, Endacott and Brooks (2018) argue that quantifiable assessments are problematic because HPT is influenced by shifting personal and social preferences, “contexts, perspectives, beliefs, predilections, emotions, and positionalities” (p. 218), as well as a student’s ability to analyze documents. Perhaps a quantifiable measurement is likely to not capture when, how, and why specific students engage in historical empathy. Finally, they argue that measuring historical empathy entirely misses the mark because HPT is not an achievement that can be scored and completed, but rather a life-long process that should influence a person’s permanent disposition and behavior in order to motivate “pluralistic civic action” (Endacott & Brooks, 2018, p. 219).

Despite Endacott and Brooks (2018) harsh critique of quantifying historical empathy, they equate historical empathy directly to clinical empathy, even applying definitions and research from leading empirical psychology researchers (e.g., Decety & Jackson, 2006; Eisenberg, 2000; Hoffman, 1985). They explain that in their model of historical empathy “the cognitive understanding of historical context and perspective shapes the affective connection made with historical figures *in a similar fashion to contemporaneous empathy* [emphasis added]” (p. 209).

Endacott and Brook’s (2018) position on historical empathy is inconsistent

because they critique efforts to quantify historical empathy and then equate historical empathy to clinical empathy in the psychology domain, and cite social psychologists who engage in rigorous quantitative methodology (Decety & Jackson, 2006; Eisenberg, 2000; Hoffman, 1985). Contemporaneous empathy is quantified frequently through various social psychology instruments, and Endacott and Brooks (2013, 2018) do not demonstrate any apprehension quoting from and citing literature reviews based on quantitative social and neuropsychological studies. If quantifiable methodology is inappropriate for historical empathy, why is it okay for clinical empathy? Why do Endacott and Brooks (2013, 2018) quote from and cite rigorous quantitative empiricists (e.g., Decety & Jackson, 2006; Eisenberg, 2000; Hoffman, 1985) from one domain of human science (psychology) and then reject potentially rigorous quantitative empiricists from another domain (education)? If historical empathy is the same or at least similar to clinical empathy, then it should be subject to similar research techniques.

This study argues for the use of rigorous quantitative methods to measure various constructs of empathy. Although, Endacott and Brooks (2013, 2018) have made important *qualitative* contributions to the research, rigorous quantitative designs can potentially further illuminate the relationship of historical empathy with clinical empathy and help researchers understand the specific activities that can most effectively induce and develop empathy for what kinds of students. Moreover, clinical psychology shares very similar challenges in regard to shifting preferences, perspectives, beliefs, and positionalities, yet it is still quantitative in orientation.

Finally, Endacott and Brooks (2018) seem to assume that historical empathy is

virtually the same construct as clinical empathy except that historical empathy is for people in the past. However, historical empathy and clinical empathy have different literature reviews, researchers, traditions, and theorists. If they are the same construct then it is incumbent on researchers to quantitatively prove it, which underscores the need for rigorous quantitative research regarding historical empathy.

At this point, the researcher is only aware of evidence of a weak correlation between historical empathy and clinical empathy. Gehlbach (2004) assessed students in their social perspective-taking ability through a video and a written task, and in their performance in accomplishing historical empathy through a written task. Gehlbach found a correlational relationship with historical empathy and social perspective-taking on the written task, $r(226) = .24, p < .01$; and the video task, $r(226) = .15, p < .05$. Gehlbach explains that it is “plausible that students who are adept at understanding how a situation might impact their friends (i.e., social perspective-taking) might also be adept at understanding how a situation might impact historical figures or events (i.e., historical empathy)” (p. 97). This suggests that a carryover may exist in understanding people and situations in history and understanding people and situations in the present. The fact that there is so little evidence connecting historical empathy with clinical empathy demonstrates that there is a need for further empirical research.

If research demonstrates a positive relationship between clinical and historical empathy, then perhaps historical empathy lesson plans could be used as an intervention to increase clinical empathy. This study will not be able to fully answer the question regarding the relationship of clinical empathy and historical empathy, but the research

design, discussed below, will at least enable an analysis of a potential correlational relationship.

Reading-Empathy Hypothesis

For centuries, philosophers, authors, and literary theorists have postulated that reading certain types of literature enhances one's proclivity to understand others (Mar & Oatley, 2008; Nussbaum, 1997). This hypothesis can be termed the reading-empathy hypothesis, which argues that the reading of certain types of texts can induce and increase empathy. One of the leading philosophic advocates of reading literature for building empathy is Nussbaum (1995, 1997, 2004) who argues that literature helps develop better lawyers, judges, and policy makers. For these professionals often rely solely on quantitative data that are stripped of their full human context so that crucial administrative and judicial decisions become a cold, utilitarian, cost-benefit analysis. In contrast, Nussbaum (1995, 1997, 2004) argues that a study of the humanities, especially through literature, does a much better job of providing the crucial context of the diverse experiences of humanity, enabling much more informed policy and judicial decisions.

Nussbaum (1997) argues that literature helps acculturate the rising generation to extend its empathy beyond just the immediate family, tribe, or nation, to become what she terms "a citizen of the world" (p. 8). She argues that, as citizens of the world, we should not engage in a purely nationalistic agenda at the expense of other nations. Rather, we should value other people of other nations and ethnicities as much as we value our own.

A crucial component in learning to be a citizen of the world is learning to practice the “narrative imagination” (Nussbaum, 1997, p. 85). Researchers have argued that we understand people as stories (Bruner, 1991; Nussbaum, 1997). Therefore, the basic childhood stories shared at the bedside open for the child not only imagination about mythical worlds, but also imagination into the thoughts and feelings of others. Thoughts and feelings can be less visible in the real world, but literature has a way of opening the heart and mind of characters, whether explicitly or inferentially, to understand the inner world of another. Nussbaum (1997) argues that development of the narrative imagination is “essential preparation for moral interaction [and] inspires intense concern with the fate of characters and defines those characters as containing a rich inner life” (p. 90). Nussbaum argues that learning other people’s stories is essential for understanding their ways of life, which is essential for building compassion. Nussbaum’s theory provides an explanation for why the engagement of literature may help develop empathy for others. Next, the researcher addresses how an engagement with certain types of texts may bring about greater empathy.

Reading-Empathy Hypothesis: Theory

Over the last few decades, researchers have made significant progress in explaining how readers comprehend text. Kintsch (1998) explains that comprehension is accomplished by “building a mental model” (p. 93), also known as a mental representation or a situation model of the text. He proposes the construction-integration (CI) model as a process through which the mind perceives text and transforms it into a situation model in two phases: (1) the construction phase where textual input activates a

wide range of potential connections even if it later proves incorrect, and (2) the integration phase where activation is constrained and irrelevant connections are deactivated to provide for a coherent situation model (Kintsch, 1998). This construction of meaning is accomplished primarily through bottom-up automatic processing similar to other forms of perception (Kintsch, 1998).

According to Kintsch (1998), mental models are constructed from two sources: from the textbase, or the semantic meaning of the propositions and from the reader's background understanding. Kintsch (1998) explains that the textbase is divided into the microstructure, which includes each individual word organized sentence-by-sentence, and the macrostructure, which is the total sum of the propositions from the microstructure organized hierarchically. The reader uses the microstructure and the macrostructure to construct the textbase, which is a mental representation of the propositions of the text (Kintsch, 1998).

Kintsch (1998) describes how background information is retrieved in the mind through what he calls a *knowledge net* or a network of nodes, and he defines nodes as propositions, schemas, frames, scripts, and production rules. Nodes are activated probabilistically by the strength of their relationship with other nodes (Kintsch, 1998). The potential meaning of a word is given by the concentric circles of other words that are closely connected to it, but the actual word meaning is only the specific associations that are activated in working memory on a specific occasion. The textbase and the reader's background knowledge combine to create a situation model.

Kintsch (1998) explains that situation models are crucial component of how

organisms function in the world. Situation models include concrete experiences closely tied to the environment and memory, as well as abstract verbal situation models (Kintsch, 1998). He argues that mental representations are crucial for comprehension because “the environment rarely provides all the information” (Kintsch, 1998, p. 15). As a result, organisms must draw on background knowledge to “fill in the gaps” or infer from fragmentary information (Kintsch, 1998). Kintsch explains that the cognitive system combines fragmentary sensory input with background knowledge to generate mental models that are compatible with the environmental and which allow the organism to interact with the environment.

Although, Kintsch (1998) does not specifically address empathy, his theory is compatible with the empathy-literature hypothesis. For example, he explains that frequent retrieval of information and increased background understanding will improve the corresponding reading comprehension within that specific domain (Kintsch, 1998). For example, Ericsson and Kintsch (1995) have postulated a long-term working memory of information that is used so often it is retrieved as quickly as if it were in the short-term working memory.

According to Pearson’s (2014) ‘Kintschian-derived model,’ readers draw from their background knowledge and the text to create a situation model. Pearson (2014) claims that a constructed situation model contributes to our background knowledge, making further construction of situation models easier in a positive reciprocal cycle. Pearson’s claim is similar to Stanovich’s (2000) claim that comprehension of a certain topic increases knowledge on that topic and makes further comprehension easier in a

positive reciprocal cycle. Kintsch's work is not domain specific. His focus is on general comprehension of narrative or expository text. However, his theory of comprehension likely applies to the comprehension of empathy.

For example, Zwaan and Radvansky (1998) extended the theory of situation models to include particular types of situation models such as situation models of *intentions* and *protagonists*. They explain that readers build situation models related to intentions, or goal-directed behavior of human agents, and that readers tend to retain details primarily as they relate to key protagonists. In other words, details are retained in memory inasmuch as they are related to a particular character. Mar and Oatley (2008) directly connect their claim that comprehension of fiction is a simulation of social experience with Kintsch's theory of situation models.

Multiple researchers argue that if situation models are dependent on the text meaning, then a text that is rich in social complexity would challenge readers to construct socially complex situation models (Kidd & Castano, 2013; 2016; Mar & Oatley, 2008; Zunshine, 2006). Similar to Pearson's (2014) Kintschian-derived Model, as one engages in the comprehension of socially complex texts, then the reader is challenged to construct socially complex situation models, which contributes to the reader's background knowledge in a positive reciprocal cycle. Not only does this improve comprehension of future texts, advocates argue that it improves empathy for the real-world (Mar & Oatley, 2008). This model is represented in

Literary theorist, Zunshine (2012), explains that literary fiction often has readers construct situation models of characters' minds. She calls the imagining of another

Figure 3.

Literary theorist, Zunshine (2012), explains that literary fiction often has readers construct situation models of characters' minds. She calls the imagining of another

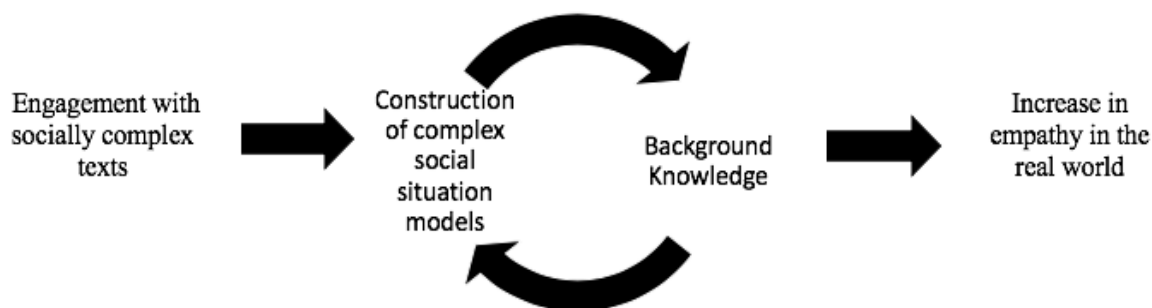


Figure 3. Theoretical model of the reading-empathy hypothesis.

person's mind as “nestings” and argues that literature often takes nested states to multiple levels. In other words, like a Russian matryoshka doll: a nested state, within a nested state, within another nested state. For example, a statement by a mother, such as, “I hope my son remembers tomorrow how he feels today,” requires the reader to imagine the maternal longings of the mother (first nested state), the son's feelings today (second nested state), and the future recollections of the son (third nested state). Zunshine calls such multiple layers of nested minds “sociocognitive complexity.” She explains further that literary fiction frequency entails nested minds to the third level and sometimes even to the fifth and sixth level.

Kidd and Castano (2013, 2016) argue that popular fiction employs simplistic stereotypical characters that are easy to identify with or demonize, so that comprehension of such texts fails to challenge readers or activate significant social cognition. In contrast, they argue that literary fiction, which they define as fiction that has won a literary award such as the O. Henry Prize, avoids stereotypical plots and cliché characters and contains

much greater ambiguity. They label literary fiction as ‘writerly’ because it requires the reader to infer or fill in the gaps. They explain that literary fiction employs complicated characters acting in morally ambiguous situations (Kidd & Castano, 2013, 2016).

Reading-Empathy Hypothesis: Empirical Research

Although authors have claimed for centuries that reading may improve compassion and human understanding of others, it is only recently that researchers have begun to test these claims empirically.

Correlational studies. A relationship between reading certain genres of literature and ToM ability was first demonstrated empirically by Mar et al. (2006), who recruited a large number of college undergraduates and measured their proclivity for reading as well as whether they tended to read fiction or nonfiction. These researchers divided the participants into two types of readers: “nerds,” those who read nonfiction, and ‘bookworms,’ those who read fiction. After assessing the participants’ genre reading exposure, Mar et al. assessed the participants’ ability to practice ToM. The bookworms, or readers of fiction, scored significantly higher on a ToM measure than did readers of nonfiction. This correlation did not establish that fiction causes an increase in empathy but it demonstrates a positive relationship with individuals that read certain types of literature and higher ToM score (Djikic et al., 2013; Fong et al., 2013; Kidd & Castano, 2013, 2016; Mar et al., 2006, 2009; Panero et al., 2016).

Causal studies through random assignment. To more closely examine a potential causal link of reading fiction on ToM, Kidd and Castano (2013), in five different experiments, randomly assigned approximately 800 total participants recruited

through Amazon Mechanical Turk. Amazon Mechanical Turk is an Internet marketplace that allows third parties to provide payment for the completion of various tasks (including psychology surveys; Buhrmester, Kwang, & Gosling, 2011). The participants were randomly assigned to read a popular fictional text, a nonfictional text, a literary text, or nothing at all, and then were assessed on ToM (Kidd & Castano, 2013). Each text was 2,000-3,500 words in length. Results showed that participants who read literary fiction scored significantly higher on a ToM task compared to participants who read nonfiction, popular fiction, or nothing at all. Kidd and Castano defined these results as a priming affect, meaning that a recent use of a neurological network prepares it for greater ease of access moments later. This study suggests that reading specific types of literary fiction may provide at least a short-term activation of ToM.

Kidd and Castano's (2013) study has been followed with a few attempts at replication. Black and Barnes (2015a) used a pretest-posttest design of fiction on ToM and found statistically significant higher posttest scores of participants who were randomly assigned to read fictional texts than those assigned to read nonfiction. In another replication study, Kidd, Ongis and Castano (2016) found a statistical difference with participants assigned to read literary fiction versus popular fiction. These three studies (Black & Barnes, 2015a; Kidd & Castano, 2013; Kidd et al., 2016) demonstrated an immediate activation of ToM by engaging in the reading of literary fiction. Two other studies, of similar designs, demonstrate an immediate effect of ToM from other means including watching Television (Black & Barnes, 2015b) or playing a video game (Bormann & Greitemeyer, 2015). These two addition studies result in a total of five

studies that provide evidence of an immediate short-term priming effect of ToM by engaging in certain tasks.

A key aspect of priming is that it is of short duration. Researchers define priming as a “*temporary activation...of an individual’s mental representations....* [P]riming produces *for a short time* a level of activation and accessibility in a representation (Bargh & Chartrand, 2000, p. 258, emphasis added). This short-term effect means that if the study is designed in such a way that there is a delay between the stimulus and the assessment, then the priming effect may decay and may not be significant by the time it is measured.

There is also evidence of a long-term enduring effect of reading on ToM. Pino and Mazza (2016) randomly assigned college students to read entire books from three different genres: autobiographical nonfiction, science fiction, and literary fiction. A week later they measured each student’s ToM and found that only those randomly assigned to the literary fiction genre scored statistically higher than autobiographical fiction and science fiction.

In addition, at least three different studies that found a positive relationship with participants’ ToM ability and higher exposure to fiction controlled for various personality attributes (Djikic et al., 2013; Fong et al., 2013; Mar et al., 2009). These researchers attempted to hold constant personality traits such as being an extrovert or introvert in order to isolate causation. These three correlational studies (Djikic et al., 2013; Fong et al., 2013; Mar et al., 2009) when placed in the context of Pino and Mazza (2016) seems to indicate the potential for an enduring and cumulative effect of reading certain types of

literature on ToM.

Different genres and mediums that are related to empathy. Although several studies have only found an effect on ToM through the reading of literary fiction, there may be various other genres (other than literary fiction) and mediums (other than reading) that activate ToM. For example, Black and Barnes (2015b) measured ToM among participants after they watched an award-winning television drama and a documentary and found an immediate effect on ToM for those who watched the drama, rather than the television documentary. Bormann and Greitemeyer (2015) found an immediate positive effect on a ToM assessment in playing a narrative video game in comparison to a non-narrative video game.

In a correlational study of ToM and four different fiction genres: domestic fiction, romance, science-fiction/fantasy, and suspense/thriller, Fong et al. (2013), found that only the romance and suspense/thriller genres demonstrated a statistical relationship on a ToM assessment, after controlling for personality, age, gender, English fluency, and exposure to nonfiction. If award-winning television, narrative video games, literary fiction, and romance fiction, all demonstrate a relationship with ToM, perhaps it is because they each include significant social complexity that requires individuals to construct complex social situation models. If this were the case, then various other mediums that demand complex social cognition could also affect ToM.

Empirical studies that did not find an effect on theory of mind from reading. Not every study has found an effect for reading fiction on ToM. Djikic et al. (2013) randomly assigned individuals to read a short narrative fictional text or a short

informational text of similar length and difficulty then measured ToM. They found no significant difference on ToM between participants who were assigned the narrative or the informational text. As stated above, according to Zunshine (2012), literary fiction often has readers engage in sociocognitive complexity. Perhaps the fictional narrative text they provided in this text did not include significant social complexity that required sufficient social cognition.

Panero et al. (2016) used the same texts as Kidd and Castano (2013) and attempted to use the same experimental design and participant population (Amazon Mechanical Turk). However, Panero et al. (2016) found no statistical difference between the readers assigned to read popular texts, literary fiction, or nothing at all.

Kidd and Castano (2017) challenged Panero et al.'s (2016) findings on two grounds. First, they argued that Panero et al. (2016) did not actually ensure that participants read the texts. Many of the reading times that participants spent on the texts were far too short, which suggests that participants did not read them. Second, the participants did not appear to be truly randomly assigned because the assigned groups were not even remotely equally distributed.

Kidd and Castano (2017) performed a re-analysis on the Panero et al. (2016) data by excluding participants who read for extremely short periods of time and found slightly higher but statistically significant scores for participants who were assigned to read literary fiction in comparison to other genres. Unexplainably however, their reanalysis found that participants assigned to read nonfiction scored similarly as participants that were assigned to read literary fiction. Perhaps, some of the issues that Panero et al. (2016)

experienced with random assignment (as disclosed in Panero et al., 2017) can explain these anomalies. Perhaps the nonfictional group was not comparable to the other participant groups.

Another attempt to replicate Kidd and Castano (2013) published (Samur, Tops, & Koole, 2018) used the same texts and population sample (Amazon Turk), except with about twice as many participants as Kidd and Castano. In one of the four experiments they excluded participants that had previously participated in similar studies and whose first spoken language was other than English. However, as with Kidd and Castano, there were no controls besides reading time to ensure that individuals actually read the text. They found a similar positive relationship between high scores on the ToM assessment and exposure to fictional texts, but they found no statistical difference for individuals randomly assigned to read different texts, whether popular fiction, literary fiction, nonfiction, or nothing at all.

In one of the four experiments, they achieved a marginal, but statistically significant effect in favor of the nonfiction condition group scoring higher than the literary fiction group (the exact opposite of their hypothesis). They argue that using the Reading the Mind in the Eye Task (Baron-Cohen et al. 2001) to measure ToM, is likely highly affected by motivational factors and that the different scores achieved by Kidd and Castano (2013) are perhaps explained by potential motivational differences. In other words, they argue that perhaps the literary fiction text was simply more interesting, which motivated participants to complete the ToM task with greater effort in the Kidd and Castano study.

These mixed findings of using literature to foster ToM (Kidd & Castano, 2013; Panero et al., 2016; Samur et al., 2018) warrant further research in this area. More replication studies with different genres, methods, contexts, and participants may shed greater light and clarity on if, when, and how certain texts influence ToM.

Empirical instrumentation. As discussed above one of the most consistent findings has been a correlation between greater fictional reading and higher ToM scores (Djikic et al., 2013; Fong et al., 2013; Kidd & Castano, 2013, 2016; Mar et al., 2006, 2009; Panero et al., 2016). Asking participants how frequently they read can be problematic because social desirability may motivate participants to exaggerate the amount of reading they do (Stanovich & West, 1989). One way to measure reading behaviors is to provide participants a list of real authors mixed with a number of distractors, which inhibits readers' ability to falsely report on their own reading frequency. Stanovich and West found a high correlation between reading frequency and author recognition. Therefore, the number of recognized authors can serve as a proxy or an estimation for amount of reading exposure. Stanovich and West named this instrument the Author Recognition Test (ART).

Virtually all of the studies reviewed above employed an instrument called the Reading the Mind in the Eye Task (RMET) developed by Baron-Cohen et al. (2001) to measure ToM (Djikic et al., 2013; Fong et al., 2013; Kidd & Castano, 2013, 2016; Mar, et al., 2009; Panero, et al., 2016). The RMET provides photographs of the zoomed-in area around the eyes of actors and has participants select the best multiple-choice response about what the actor is thinking or feeling. However, this instrument demonstrated low

levels of internal consistency with adolescent participants in a published study (Müller & Gmünder, 2014; $\alpha = .53$), and in a pilot study conducted with adolescent participants in preparation for this proposed study. As a result, the investigator of this study decided to use a different instrument to measure ToM.

In contrast, Pino and Mazza (2016) also employed the Faces Test (Baron-Cohen, Wheelwright, & Jolliffe, 1997) and the False Belief Test (Rowe, Bullock, Polkey, & Morris, 2001). Pino and Mazza (2016) found that college students randomly assigned to read literary fiction scored higher on the Faces Test and the False Belief Test than those randomly assigned to read a popular fiction text or a historical memoir.

Summary of reading-empathy hypothesis empirical evidence. In summary, multiple studies have found a correlation between individuals who frequently engage in literary fiction and higher scores on ToM tasks (Djikic et al., 2013; Fong et al., 2013; Kidd & Castano, 2013, Kidd & Castano, 2016; Mar et al., 2009; Panero et al., 2016). Four of six randomized studies have found a causal effect on ToM from reading literary fiction, whether immediate short-term activation (Black & Barnes, 2015a, Kidd & Castano, 2013; Kidd et al., 2016), or enduring over the span of a week (Pino & Mazza, 2016). These studies have employed a variety of instruments for assessing ToM including the RMET (Baron-Cohen et al, 2001), the Faces Test (Baron-Cohen et al., 1997) and the False Belief Test (Rowe et al., 2001) and have still found statistically higher scores for individuals assigned to read literary fiction in comparison to individuals randomly assigned to read other types of texts (Djikic et al., 2013; Fong et al., 2013; Kidd & Castano, 2013, 2016; Mar et al., 2009; Panero et al., 2016; Pino & Mazz, 2016). In

addition, research has demonstrated an effect for the reading of romance fiction (Fong et al., 2013), for watching award-winning television (Black & Barnes, 2015b), and even for playing a narrative video game (Bormann & Greitemeyer, 2015). Two studies did not find an effect for engaging literary fiction, one of which appears to have had methodological issues (Kidd & Castano, 2017; Panero et al., 2016).

Activation of Theory of Mind through the Reading of Historical Texts

There is good reason to believe that the reading of historical texts can activate ToM. Historical thinking and engaging in HPT has much in common with the arguments made on behalf of literary texts. Literary fiction tends to involve round characters who avoid simple stereotypical categorization, in socially complex situations with lots of ambiguity that leaves room for readers to infer motivations and reasons for their behavior (Kidd & Castano, 2013, 2016; Mar & Oatley, 2008; Zunshine, 2006). Literary fiction often requires readers to make challenging social inferences through the construction of sophisticated situation models (Mar & Oatley, 2008; Zunshine, 2006). Crucially, most theorists have not emphasized the attribute of ‘narrative’ in texts that activate ToM (Kidd & Castano, 2013, 2016; Mar & Oatley, 2008; Zunshine, 2006). In fact, one study (Djikic et al., 2013), where the narrative aspect of the experimental text was the most salient component of the experimental text, did not find a statistical difference in ToM in comparison to a control group.

Real-life historical figures are no less complex, acting in challenging situations that often have no risk-free alternatives. Historical sources tend to be ambiguous,

allowing a degree of interpretation especially pertaining to motives and intentions (Doppen, 2000). If the comprehension of literary fiction activates ToM, then it is plausible that the comprehension of certain historical texts could activate ToM as well. These certain historical texts may include diary entries, letters, biographical accounts, or any other historical text that facilitates the construction of socially complex situation models. This rationale is demonstrated below in

Gaps in the Research Literature

This study seeks to address three gaps in the research literature pertaining to reading and empathy: Namely, the use of adolescent participants, the implementation of an educational context, and the application of historical documents for ToM, EC, and HPT.

Participants in empirical studies investigating the effect of literature on ToM have been almost exclusively adults who were recruited through college campuses, Amazon-Turk, or readers of the NY Times (Black & Barnes, 2015a; Kidd & Castano, 2013; Kidd et al., 2016; Pino & Mazza, 2016). A search on EBSCO using “adolescence,” and “theory of mind” and variations of “fiction,” “literacy,” or “reading” as key words did not find any empirical studies that measured the effect of reading fiction on adolescent ToM.

Table 1.

Gaps in the Research Literature

This study seeks to address three gaps in the research literature pertaining to reading and empathy: Namely, the use of adolescent participants, the implementation of an educational context, and the application of historical documents for ToM, EC, and HPT.

Participants in empirical studies investigating the effect of literature on ToM have been almost exclusively adults who were recruited through college campuses, Amazon-Turk, or readers of the NY Times (Black & Barnes, 2015a; Kidd & Castano, 2013; Kidd et al., 2016; Pino & Mazza, 2016). A search on EBSCO using “adolescence,” and “theory of mind” and variations of “fiction,” “literacy,” or “reading” as key words did not find any empirical studies that measured the effect of reading fiction on adolescent ToM.

Table 1

Potential Similarities of Literary Fiction and Historical Documents

-
- Involve social complex situations, which require complex social situation models
 - Writerly: text is ambiguous; readers have opportunities to fill in the gaps and make references
 - Round characters that avoid simple stereotypes
 - Involve challenging social inferences to construct meaning
 - Involve nested minds
-

Hence, it appears that a gap in the research literature exists in relation to using adolescent participants to measure the effect of literature on ToM.

In addition, very few studies have targeted ToM assessment from the educational research perspective or the literacy research perspective. An EBSCO search using “reading,” “classroom,” and “theory of mind,” produced no relevant results. A search of

“empathy” in *Reading Research Quarterly*, often considered the flagship of literacy research, found five articles that are associated with the term empathy. None of the five articles measured empathy quantitatively or sought to ascertain the effect of reading on empathy. No published studies have empirically investigated the role of ToM within the context of a classroom. As discussed above, empathy is a crucial component of SEL learning (CASEL, 2005). It is also required in some state history standards (e.g., USBE, 2016) and the Common Core State Standards (CCSS, 2010). Because empathy is a required part of the curriculum, it is vital for literacy and education researchers to contribute to this area of research.

The effect of reading historical documents or a historical narrative on ToM has not been investigated directly. Studies have investigated the reading of literary fiction, historical memoirs, nonfictional texts, popular fiction, and of watching certain shows, but they have not investigated a potential causal effect of reading historical documents or a historical narrative on ToM (Djikic et al., 2013; Fong et al., 2013; Kidd & Castano, 2013, 2016; Mar et al., 2006, 2009; Panero, et al. 2016).

There appears to be insufficient research regarding the relationship of different historical texts on EC and HPT. An EBSCO search of “reading,” “empathic concern,” and “history” produced zero results. An EBSCO search of “historical empathy” and “empathy” in the subject search location found 98 studies, only one of which provided empirical evidence for a correlational relationship between practicing historical empathy and *social perspective taking* (Gehlbach, 2004). Further research needs to determine if there is a causal relationship in reading historical texts on ToM, EC, and HPT for

adolescents. Evidence for such a relationship could potentially lead to the employment of certain historical texts as empathic interventions.

Summary

This research review has argued that adolescence is a crucial period for social and emotional development (Blakemore, 2008; Eccles et al., 1993), yet adolescents are currently exposed to significant challenges that may adversely affect their social and emotional development (e.g., Anderson et al., 2015; Konrath et al., 2011; Putnam, 2000; 2015; Rideout et al., 2010). As a result, the education system has an opportunity to take a greater part in providing SEL. One primary means of providing more SEL is to apply SEL in conjunction with academic lessons. A crucial part of SEL is developing social awareness, which is primarily defined by CASEL (2005) in terms of empathy and perspective-taking. Empathy and perspective-taking are explicitly stated as goals for SEL, while historical empathy and HPT are often explicitly required in state and national standards (e.g., Utah Social Studies Standards; USBE, 2016). This provides a clear overlap of teaching empathy for SEL and historical academic skills.

This study addresses three constructs of empathy: ToM, EC, and HPT. ToM can include a mental simulation of the action or emotion being perceived. EC has been promoted by several religious and philosophical traditions for centuries (Armstrong, 2006), and motivates altruistic behavior (Batson, 2011). History teachers and researchers have recently engaged in teaching HPT by having students engage in the analysis of historical sources (Endacott & Brooks, 2018).

There is growing theoretical and empirical evidence that demonstrates a correlational and perhaps an even causal relationship between engaging in tasks that require social cognition, (whether through reading, award winning television, or narrative video games), and higher scores for ToM (Black & Barnes 2015a, 2015b; Djikic et al., 2013; Fong et al., 2013; Kidd & Castano, 2013, 2016; Kidd et al., 2016; Mar et al., 2006, 2009; Panero et al., 2016; Pino & Mazza, 2016; Samur et al., 2018). Empirical research has not investigated a causal effect of analyzing historical sources on ToM, EC, or HPT, for adolescents within a school context.

CHAPTER III

METHODOLOGY

In this study, adolescent students were randomly assigned to read one of two historical texts pertaining to the Salem witch crisis and then completed assessments of theory of mind (ToM), empathic concern (EC), and historical perspective-taking (HPT). The three purposes of this study were as follows.

1. Investigate if historical texts can activate ToM among adolescents.
2. Investigate if historical texts can activate EC for adolescents
3. Investigate if historical texts can activate HPT for adolescents

The research questions that addresses these three purposes are:

1. To what extent do different types of historical texts influence adolescents' ToM?
2. To what extent do different types of historical texts influence adolescents' EC?
3. To what extent do different types of historical texts influence adolescents' HPT?

This chapter begins with a description of the study design, which is followed by the participants and materials. Then the researcher addresses the procedures.

Study Design

This study employed an empirical quantitative design based on a post-positivist theoretical view (Popper, 1959). Post-positivism assumes that reality exists and that we can access it to a degree, and that researchers can make limited truth claims based on evidence through the scientific method (Popper, 1959). This study employed a

randomized design to provide empirical evidence for a causal relationship of reading one of two types of historical texts on ToM, EC, and HPT.

This study design had three advantages: random assignment, a shared procedural history for the participants, and the use of control variables. First, in the application of random assignment to a large group of participants, it is unlikely to have systematic differences between the two randomly assigned groups before the treatment (Christensen, Johnson, & Turner, 2014).

Second, this study design used two separate groups that shared the same history in regards to the procedures. Having separate groups that complete all of the same tasks in the same way allows researchers to be more confident that the treatment is the single cause of any difference. The study was administered through a computerized online platform. This provided a uniformity in the administration of the study to all of the participants, which helps further isolate causation (Christensen et al., 2014).

Third, this study design controlled for several confounding variables including: reading exposure, gender, age, estimated comprehension ability, and estimated facial emotion reading ability. These extra sources of data enabled the researcher to control for potential background differences using multiple regression analysis (Cohen, Cohen, West, & Aiken, 2003). The advantage of applying these three processes (randomization, uniform treatment, and the collection of control variables), is that they provide several ways to isolate possible causation.

Participants

The selected middle school was located in an upper middle-class, predominantly white neighborhood in the Western U.S. Less than 1% of the students were listed as English Language Learners, 14% as ethnic minorities, 7% participated in special education, and about 10% of the student population were eligible for free or reduced lunch.

The researcher received permission from the university IRB committee (see Appendix A), the school district (see Appendix B), the school principal, the participating teachers, parents whose students were participant, and the students themselves. About 525 eighth-grade student participants were introduced to the study by the major professor on the advisory committee. All of the participants were enrolled in eighth-grade US history at the same school. Each of the participants had one of three US history teachers. The researcher was one of the teachers. Two hundred and thirty-six students provided consent forms signed by their parents with assent forms signed by themselves. Each student who provided a consent form also provided an assent form. An additional 30 students turned in consent forms that denied consent. The remainder of the students did not turn in any form and did not participate in the study. The data for consent received and denied for each class is included in Table 2.

The actual number of students who participated on both days of the study was $N = 227$. Seven percent of the students who participated were excluded from the study because they scored greater or less than three *SDs* on key assessments discussed below in the findings. The final total number of participants included in the data was $n = 212$. The

Table 2

Consent Forms Received or Denied from Student Participants

	Students of the researcher-teacher		Students of the participating teacher 1		Students of the participating teacher 2		Total	
Total or percent	Consent	Denied	Consent	Denied	Consent	Denied	Consent	Denied
Total <i>N</i>	110	8	63	8	63	14	237	30
Percentage of total forms (<i>N</i> = 267)	41	3	23	3	23	5	88	11

participants included $n = 121$ females and $n = 91$ males. Almost all of the participants were ages 13 or 14 with 97% of the participants born within the same 12-month period. All the participants were born within a 19-month period.

Materials

Texts

This study employed three different historical texts.

1. **Introductory text.** Modified from Stanford's *Reading Like a Historian*, the introductory text explains some of the basic vocabulary and concepts behind the early American witchcraft trials (e.g., "The Puritans believed that witches were people that made a covenant or pact with the Devil where he gave them special magical powers in exchange for their soul"). The introductory text (Appendix C) is 347 words in length and a Coh-metrix analysis (Graesser, McNamara, Louwerse, & Cai, 2004) places this text at an eighth-grade Flesch-Kincaid level.

The introductory text scores high on syntactic simplicity, word concreteness, and

deep cohesion, and is low in referential cohesion and narrativity. The introductory text has an estimated Lexile® score between a 1000-1100 (Stenner, Burdick, Sanford, & Burdick, 2006). According to the Common Core State Standards, a Lexile of 1000-1100 is between a sixth and a tenth-grade reading level (CCSS, 2010), making this text an appropriate level for eighth-grade students. Included at the end of the introductory text are five reading comprehension questions for students to provide written responses (e.g., “According to the Puritans, what was a witch?”).

2. **Historical documents.** A collection of abridged and modified historical documents about the Salem witch crisis (see Appendix D), these historical documents are primary sources and would likely be classified as ‘traces’ because they are fragmentary pieces of evidence that do not tell a complete story (Seixas & Peck, 2004). The collection includes a sermon, a journal account, two court examinations, two petitions from prisoners, and a public confession. All of the sources have been abridged and modified by the researcher to make them of comparable length and difficulty to the narrative text. Phrasing and vocabulary were modernized similar to published primary documents modified for precollegiate history courses, as is common practice in secondary history classrooms (e.g., Wineburg, Martin & Monte-Sano, 2013) with the purpose of making the text more accessible. To comprehend these texts, readers may have to make challenging social inferences, build complex social situation models, and engage in frequent second and third-level mind reading (Mar & Oatley, 2008; Zunshine, 2012).

For example, in a court examination when Abigail Hobbs is told she is accused of witchcraft and asked if she is guilty, she gives the ambiguous answer “I have seen sights

and been scared. I have been very wicked.” A reader might infer that she is trying to avoid punishment by walking a tightrope of not denying nor confirming directly that she is a witch. However, the judge will not relent. He continues to interrogate her and she appears to be trying to tell him what she thinks he wants to hear without confessing to actually doing anything serious. She admits the Devil has come to her. She admits that he wants to make her a witch. The judge continues to ask leading questions that generate the answers that he expects from his terrified victim. Because the judge believes she is telling the truth, he may become more certain that witchcraft is real and has a firm hold in Salem. His determination appears to create a self-fulfilling prophecy, where his original conclusion is reinforced by the confessions induced by his interrogations.

The historical documents are a total of 1,141 words, 43 words longer than the historical narrative and at a sixth-grade Flesch-Kincaid level. According to a Coh-metrix analysis (Graesser et al., 2004) the historical documents score higher on narrativity in comparison to the historical narrative. This means that the historical documents use more personal pronouns, verbs, and references to intentional action than the historical narrative. They are much lower than the narrative text on syntactic simplicity and word concreteness. This means the historical documents tend to include longer sentences and more abstract words. A Lexile® analysis places the primary documents between an 800-900 Lexile® level (Stenner et al., 2006). According to the Common Core State Standards that is between a fifth and an eighth-grade reading level (CCSS, 2010).

3. **Historical narrative.** The historical narrative, *The True Story of the Salem Witch Hunts* (Zumbusch, 2009; see also Appendix E) uses one short quotation from

primary source material; otherwise, it is a secondary ‘account’ of the historical event (Seixas & Peck, 2004). The historical narrative is written to tell the reader what to know and think about the events rather than allow the reader to decide, and often narrates in an omniscient unambiguous manner (e.g., “historians now know that these people were not witches,” “let’s take a look at the true story,” “These [innocent people] were killed because they would not confess to something of which they were not guilty,” “Many family members were scared and they lied to save themselves,” “we must make sure witch-hunts like these never happen again,” etc.). This matter-of-fact, unambiguous language, may not activate ToM because it does not have the reader build complicated social situation models (Mar & Oatley, 2008), make difficult inferences, or engage in second or third-level mind reading (Zunshine, 2012).

The historical narrative is 1,097 words in length. A Coh-metrix analysis (Graesser et al., 2004) places this text on a sixth-grade Flesch-Kincaid level. It scores high on syntactic simplicity, word concreteness, and deep cohesion, which means it uses numerous short sentences with less abstract vocabulary and refers to similar underlying concepts throughout the text. A Lexile® analysis places the text at an estimated Lexile® score between 800-900 (Stenner et al., 2006). According to the Common Core State Standards this is between a fifth and an eighth-grade reading level (CCSS, 2010). The word length, Flesch-Kincaid Grade level, and Lexile scores for all three texts are portrayed in

Table 3. The separate components of the Coh-metrix analysis for the historical documents and the historical narrative are portrayed in

Table 4.

Table 3

Comparison of All Three Texts

Text	Word length	Flesch-Kinkaid grade level	Lexile
Introductory text	347	eighth	1000-1100
Historical documents	1,141	sixth	800-900
Historical narrative	1,097	sixth	800-900

Table 4

Coh-Metrix Components for Two Texts

Coh-metrix component	Historical documents	Historical narrative
Narrativity	90	53
Syntactic simplicity	56	96
Word concreteness	41	80
Referential cohesion	52	44
Deep cohesion	75	92

Instructions and Guiding Question

All students were provided instructions and a guiding question. The instructions are located in Appendix F. The exact wording of the guiding question was: “Why did people behave the way they did during the Salem witch crisis?” Students were then given the following instructions for developing their response to the guiding question:

You may focus on all the participants or select participants. You may answer why some people accused, some confessed, and others refused to confess. In particular, make sure that you explain their perspectives (thoughts, feelings, and beliefs), and the context (how their culture and society is different from ours). It should be at least a paragraph (250 words), but may be longer.

Your paragraph will be graded according to the school rubric on your use of a

thesis, evidence, elaboration, and conventions. The citations are provided. The texts are not provided, but do your best to paraphrase the evidence.

Author Recognition Test

The Author Recognition Test (ART; Appendix G) provides a simple proxy to estimate reading exposure (Stanovich & West, 1989). The researcher adapted the ART for adolescents by obtaining best-selling juvenile authors promoted on Amazon. Then the researcher used Google's search engine to rank each of the top bestselling juvenile authors according to the greatest number of found results. The researcher kept the top 65 and randomly added 65 names of the Reading Hall of Fame inductees as distractors. In a pilot study there were only 31 real authors chosen substantially above the distractors. The current study used these 31 real authors mixed with 31 Reading Hall of Fame researchers as distractors for a total of 62 items. An example of an actual famous juvenile author is C. S. Lewis. An example of a distractor is Keith Stanovich.

In this study, this version of the ART received a Kuder-Richardson 20 (KR-20) reliability score of .815. The ART demonstrated positive skewness (>1), but it was still within acceptable limits (West, Finch, & Curran, 1995). Kurtosis was also within appropriate limits.

Demographics Survey

The demographics survey consisted of seven items: the participant's first name, last name, teacher, period, birth sex, birth month, and birth year.

The Diagnostic Analysis of Non-Verbal Accuracy 2 Child Faces (DANVA)

The Diagnostic Analysis of Non-Verbal Accuracy (DANVA) 2 Child Faces (Nowicki & Duke, 1994) consists of 24 images of children making four different facial expressions (happy, sad, angry, fearful). Participants are shown each image for two seconds, after which they select the most appropriate emotion from a list of the four emotions provided immediately following the presentation of the image. The DANVA has a high Cronbach's alpha coefficient (.88), test-retest reliability (.84; Nowicki & Duke, 1994) and high factor saturation ($\omega = 0.93$; Olderbak, et al., 2015). An example of an item from the DANVA 2 Child Faces is located below in Appendix H.

For this study, the researcher randomly combined all 24 items on the DANVA into four groups because entering all 24 items individually can artificially inflate a Cronbach's alpha score, and received a Cronbach's alpha of .586.

Faces Test

The Faces Test (Baron-Cohen et al., 1997) contains 20 pictures, 10 of which represent basic emotions (e.g., happy, sad, angry, etc.) and 10 that represent complex emotions (e.g., admiration, interest, thoughtfulness, etc.). The Faces Test reliably distinguishes between normal adults and adults with high functioning Autism (Baron-Cohen et al., 1997). For each question on the Faces Test there is one distractor so that participants have a 50% chance of getting the correct answer. Two example items are provided in Appendix I.

For this study, the researcher analyzed the reliability of the Faces Test. In order to

perform a reliability analysis on the Faces Test, the researcher randomly combined all 20 items into five groups because instruments that have more than five or six items can artificially inflate Cronbach's alpha scores. The reliability analysis produced a Cronbach's alpha of .178 for reliability.

Emotions Survey

The Topic Emotions Survey (TES; Broughton, Sinatra, & Nussbaum, 2013) was developed to measure emotions sparked by a specific topic—in this case, students' emotions about the Salem witch crisis (see Appendix J). The TES was adapted from the Class-Related Emotions Scales (CRES; Pekrun, Goetz, & Perry, 2005). The CRES was designed to measure general classroom emotions such as emotions related to studying for a test, emotions related to classroom instruction, and emotions related to the teacher.

The TES (Broughton et al., 2013) measures 18 topic-related emotions including enjoyment, curiosity, anxiety, boredom, and interest. Students are asked to rate their emotions on a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree). Cronbach's alpha for the TES ranges from 0.77 to 0.94. The TES allows participants to report a range of emotions in order to potentially more fully understand the participants' emotions in relation to EC.

This study combined items from the Topic Emotions Survey (Broughton, et al., 2013) with the Emotional Response Survey (ERS; Batson, Eklund, Chermok, Hoyt, & Ortiz, 2007). The ERS targets six empathic emotions that have been used several times in assessing EC: sympathetic, softhearted, warm, compassionate, tender, and moved (Batson et al., 2007). These six empathic emotions were combined with 14 emotions of the TES

(Broughton et al., 2013) to create a total of 20 emotion items. Having an instrument that targets empathic emotions and having an instrument that assessed for a range of emotions were both considered important to understanding the empathic response of students. All 20 emotions were randomized when combining the TES and ERS into one survey for the current study. This instrument asks students to rate their emotions on a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree). A sample item is “When I read about the Salem witch crisis, I felt glad.”

Using an exploratory factor analysis (EFA) and Pearson correlation analysis, three emotion groups were determined: Positive Emotions, Negative Emotions, and Empathic Emotions. The emotions included in each emotion group are portrayed in Table 5.

These emotion groups are displayed in Table 6 with their respective Cronbach’s alpha scores and their Pearson correlation scores. All emotion groups were well within acceptable limits for kurtosis and skewness.

The Emotional Response Survey (ERS) included six items that were designed to measure for Empathic Emotions including softhearted, warm, tender, moved, sympathetic, and compassionate. Warm did not correlate as strongly with the other

Table 5

Emotions Included in Each Emotion Group

Emotion Groups	Emotions in each group
Empathic emotions	Sympathetic, compassionate, tender, moved, and softhearted (warm removed)
Negative emotions	Mad, disappointed, scared, upset, nervous, annoyed, uneasy, irritated, and worried
Positive emotions	Glad, excited, happy, surprised, warm

Table 6

Correlations and Cronbach's Alpha Scores for Four Emotion Groups

Emotions	Cronbach's α	1	2	3	4
1 empathic emotions	.692		.262**	.176*	-.150*
2 negative emotions	.815			-.007	-.256**
3 positive emotions	.590				.612**

Empathic Emotions. Warm is also correlated with positive emotions such as glad, $r(212) = .306, p < .001$; an indication of unempathetic feelings, considering that the individuals in the texts experienced extreme hardship. In an EFA, warm was a better fit with glad than with the other Empathetic Emotions as demonstrated in Table 7, while the other Empathic Emotions more strongly loaded onto upset. For these reasons, warm was removed from the Empathic Emotions group and this study only included the five Empathic Emotions softhearted, tender, moved, sympathetic, and compassionate as a measure of EC.

Scholastic Reading Inventory

The Scholastic Reading Inventory (SRI) is an online adaptive reading assessment created by Scholastic that matches elementary or secondary students to a Lexile® score (Scholastic, 2014). The assessment has each participant read multiple authentic informational and literature texts then poses multiple-choice questions that target understanding of the main idea, causality, inference, making conclusions, and generalizations (Scholastic, 2014). The Lexile® score is a developmental reading score that ranges from 0 for beginning readers to 1725 for advanced readers (Scholastic, 2014).

Table 7

Exploratory Factor Analysis of Six Empathic Emotions, Upset and Glad

Emotions	Component	
	1	2
Glad	.059	.737
Warm	.482	.560
Softhearted	.712	.072
Moved	.552	.286
Tender	.718	.105
Sympathetic	.609	-.492
Compassionate	.688	-.083
Upset	.371	-.649

The SRI generates 25 questions from a bank of over 5,000 questions that seek to match student ability (Scholastic, 2014), typically administered during the span of a single class period. When a student answers a question correctly, the subsequent question is more difficult. When a student answers incorrectly, the subsequent question is easier. The SRI is designed to hone in on the exact level of a reader's comprehension ability. Ideally, the SRI is administered between 3-5 times a year to assess student reading growth (Scholastic, 2014).

The SRI online version is correlated with the paper version ($r = .83$), and multiple other reading comprehension assessments, including the North Carolina End-of-Grade Tests ($r = .73$ for third grade, $r = .67$ for fourth grade), the Pinellas Instructional Assessment Program ($r = .62$ for third grade), and the Comprehensive Test of Basic Skills ($r = .74$ for fifth and $r = .56$ for seventh; Scholastic, 2014). It also achieved high internal

consistency/reliability (Cronbach's $\alpha = .85$), high test-retest reliability ($r = 0.89$) and sufficient criterion validity ($r = .70$ to $.83$) (Scholastic, 2014). Although the SRI results include a variety of norm or criterion referenced scores, this study will only use the Lexile® Score.

Historical Perspective-Taking Rubric

Student essays were scored according to a HPT rubric (See Appendix K).

According to the research literature, HPT requires the use of four different criteria.

1. Evidence (O. L. Davis et al., 2001; VanSledright, 2004)
2. Imagination and inferences (O. L. Davis et al., 2001; Endacott & Brooks, 2013; 2018; VanSledright, 2004)
3. Contextualization (O. L. Davis et al., 2001; Endacott, & Brooks, 2013, 2018, Wineburg, 1998)
4. Perspective-taking (O. L. Davis et al., 2001; Endacott & Brooks, 2013, 2018)

Each of these four criteria has a potential score of 0-3, for a possible total score of 12. A score of zero signifies no discernable application of HPT. A score of 12 signifies that extremely high HPT is evidenced in the student essay. Scores are assigned at half-point intervals.

In the Evidence category, low scores were represented by a zero for a person who provides no citations or textual references, a one for one citation, a two for multiple citations, and a three if all claims were backed by a specific source. Evidence is crucial for HPT to ensure that the perspective-taking is accurate. According to O. L. Davis (2001), HPT is “imagination restrained by evidence” (p. 4).

However, historical evidence always contains significant gaps that enable

opportunities to imagine, infer, and ‘fill in’ the gaps (VanSledright, 2004). In a historical paragraph, the primary opportunity to do this may be in the elaboration, after the use of evidence. A high score in the imagination and inferences category signified that students explained and expanded the text and provide key insights.

A key aspect that makes HPT so challenging is the unique historical context. History is radically different from the present (Wineburg, 2001). A high score in the Contextualization category signifies that the student attended to the relevant historical, cultural, social and/or political context of the time (Wineburg, 1998). A low score has no or minimal recognition of the context. Low scores include what is called presentism (Endacott & Brooks, 2013; Wineburg, 2001), or an application of the present context onto the text.

Finally, HPT requires perspective-taking, which focuses attention on the thoughts, feelings, desires, and motivations of the people involved (Endacott & Brooks, 2013). Perspective-taking means that students do not just understand how they would feel in the others situation, but they seek to understand how the historical figure may have felt in the situation. This is similar to what the social psychologist Batson (2010) refers to as ‘perspective-taking as other’. A high score in the Perspective-taking category signifies that the historical paragraph includes a discussion of the thoughts and feelings of others, and how they may be different from what people in the present may think or feel in a similar situation. A low score signifies no recognition of the internal state of the people involved or may include what Endacott (2010) calls ‘egoistic drift,’ an insertion of self with present values.

Text-specific Reading Comprehension Questions

The text-specific comprehension questions included eight comprehension questions for each text for a total of 16 different questions (Appendix L). Four of the questions for each text measured literal comprehension and four questions measured inferential comprehension. Literal comprehension questions target what is explicitly stated (Day & Park, 2005). If students read and comprehend the surface level of the text, they should be able to answer the literal comprehension questions adequately. Inferential questions may be more difficult to correctly answer because the answers are not explicitly stated in the text (Day & Park, 2005). Inferential questions often require students to apply their own background understanding with the information in the text (Day & Park, 2005). Examples of literal and inferential comprehension questions for each text are provided in Table 8. The eight questions for each text are in a multiple-choice format, providing automatic instant scoring. The 16 questions were created by the researcher and then reviewed by two literacy experts with doctorate degrees in education.

The text-specific comprehension questions demonstrated adequate internal consistency with Cronbach's alpha scores of .959 for the narrative account and .908 for the historical documents. Participants who read the narrative account scored more than a point higher on average (narrative text comprehension $M = 6.72$, $SD = .22$; historical document comprehension $M = 5.52$, $SD = 1.55$). The comprehension questions are also moderately correlated with SRI scores (historical documents: $r(119) = .387$, $p < .001$; narrative text: $r(115) = .480$, $p < .001$).

Table 8

Examples of Text-Specific Comprehension Questions

Question type	Question	Answers
Literal narrative question	According to the text, what were Puritans like?	<ul style="list-style-type: none"> • Strict and hardworking • Relaxed and lazy • Hateful and mean • Kind and patient
Inferential narrative Question	According to the author, why did the accused often confess to being witches?	<ul style="list-style-type: none"> • To protect themselves • Because they were witches • Just as a joke • Because they believed they were witches
Literal documents Question	According to Cotton Mather's sermon in Boston, what did he believe about witches?	<ul style="list-style-type: none"> • They were real • They were not real • He did not know either way • He did not care either way
Inferential document question	According to the diary of Reverend Deodat Lawson, how may Abigail William's behavior be characterized?	<ul style="list-style-type: none"> • Strange • Polite • Friendly • Shy

Procedures

The researcher and the participating history teachers implemented the study as part of the regular eighth-grade U.S. history curriculum. The researcher and the two participating teachers used an online platform (Qualtrics) to administer the study in a uniform manner. The study was administered over two days. The first day primarily entailed collecting data on control variables. The second day primarily entailed the participants reading the appropriate texts and then completing the post-assessments. A complete procedures table with each day, instrument, instrument purpose, estimated time, and the type of variable is located in Table 9.

Table 9

Procedures Table

Day	Part of the study	Instruments	Purpose	Time in min	Variable
Day 1	Pretest	DANVA	Pretest of ToM	5-7	Control
		ART	Reading Exposure	2-3	Control
		Demographic survey	Control variables	1	Control
		Intro. Text	Provide background	3-7	Control
Day 2	Treatment	Instructions	Ensure Quality	3	
		Historical texts	Effect EC, ToM & HPT	5-10	Experimental
	Posttests	Emotions survey	Measure emotions	3-5	Dependent
		Faces test	Measure ToM	2-3	Dependent
		Text-specific comprehension questions	Obtain evidence of comprehension	5	
		Historical paragraph Writing	Measure HPT	30	Dependent

The researcher also collected two additional pieces of information that were used as control variables: the SRI score and the duration times per online page used as a measure of reading speed and speed of completing the instruments. Qualtrics, the online platform used for this study (Snow, 2011), has the ability to keep track of student time spent on each browser page. The study placed each text and each instrument on different browser pages and recorded the amount of time each student spent on the respective browser pages.

On Day One of the study, students logged into Qualtrics and completed the DANVA 2 Child Faces, the ART, and the demographics survey, which included first name, last name, teacher, period, birth sex, birth month and birth year. Next, all students

read the introductory text that provided background information on the Salem witch crisis. Students then answered the five comprehension questions related to the introductory text, which helped ensure a minimal background understanding of the historical context. Table 10 shows the procedures for Day One.

On Day Two of the study, students again logged into Qualtrics and read the instructions and the guiding question. Next, half of the students were randomly assigned to read either a narrative text or a collection of historical documents. Immediately following the readings, students were asked to complete the Faces Test and the Emotions Survey. These two instruments were counterbalanced to control for order effects (Russell & Purcell, 2009). Specifically, half of the participants randomly received the Faces Test first then completed the Emotions Survey second while the other participants received the Emotions Survey first and then completed the Faces Test second. After completing the Faces Test and the Emotions Survey, students completed a comprehension assessment for their assigned text. With the remainder of the class time, students wrote a prompted paragraph. These procedures are provided in Table 11.

The Faces Test and the Emotions Survey needed to be completed immediately after students read their assigned text because the effect of reading on students' emotional states is likely of short duration (Broughton et al., 2013). Previous studies (Kidd &

Table 10

Procedures for Day One

Step 1	Step 2	Step 3	Step 4	Step 5
DANVA 2 Child Faces	ART	Demographics survey	Introductory text questions	Comprehension questions

Table 11

Procedures for Day Two

Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7
Instructions	Guiding question	Randomly assigned to read narrative text or historical documents	Faces test counterbalanced with emotions survey	Emotions survey Counterbalanced with faces test	Text-specific comprehension questions	Historical paragraph writing

Castano, 2013; Kidd et al., 2016) have found an immediate small effect after randomly assigning individuals to read different texts. Defined as a priming effect (Kidd & Castano, 2013), it indicates that a recent use of a neurological network prepares it for greater ease of access moments later.

Most of the control variables were collected on Day One in order to maximize the time available for reading and completing the post-assessments on Day Two. The only control variable collected on Day Two was the text-specific comprehension assessment because it had to be administered after students had read the texts. The comprehension assessment was placed after the Faces Test and Emotions Survey and before the students wrote their paragraphs so that students were able to focus all of the remaining class time on writing the paragraph.

Using a written task has been a common means for assessing HPT (Endacott, 2014; Endacott & Brooks, 2013). The student-generated paragraph was not for the purpose of activating HPT, ToM, EC, or improving comprehension, although it could potentially be used for that purposes in other studies. Rather, in relation to this study, the student-generated paragraph is for the purpose of providing evidence of HPT or lack of HPT among participants.

The student participants were provided about 30 minutes to complete the written paragraph. Typically, students are provided the documents while engaged in historical writing. However, allowing later exposure of the experimental texts would undermine the validity of the study. As a result, students wrote their historical paragraphs based on their initial comprehension of their assigned reading.

CHAPTER IV

RESULTS

This chapter describes the results for the study. It begins with a preliminary analysis of the exclusions applied in this study, an analysis of the potential effects of the three different teachers, an analysis of the scores for the HPT paragraphs, and the potential effect of the order of the Faces Test and the Emotions Survey. After the preliminary analysis, the researcher addresses the three research questions. The three Research Questions were as follows.

1. To what extent do different types of historical texts influence adolescents' theory of mind (ToM)?
2. To what extent do different types of historical texts influence adolescents' empathic concern (EC)?
3. To what extent do different types of historical texts influence adolescents' historical perspective-taking (HPT)?

The first analysis investigated the effect of the different texts on ToM, the second analysis investigated the effect of the different texts on HPT and the third analyses investigated the effect of the different texts on EC. Finally, the final section includes three post-hoc analyses: one related to reading times, another to specific authors, and finally an analysis pertaining to birth sex.

Preliminary Analysis

Exclusions

Social science research often excludes data that are greater than three standard deviations from the mean in key measures (e.g., Cordova, Sinatra, Jones, Taasoobshirazi,

& Lombardi, 2014; Hatzenbuehler, Corbin, & Fromme, 2008; Sargent, Dopkins, Philbeck, & Chichka, 2010). This analysis applied this same standard for excluding outliers to insure the quality of the data.

Students completed the Faces Test and the Emotions Survey immediately following their reading of either the historical narrative or historical documents text. The potential effect of reading was expected to be of short duration, whether it was a priming effect for ToM (Bargh & Chartrand, 2000; Kidd & Castano, 2013), or an emotional response (Rosenberg, 1998). As a result, individuals were excluded who did not complete the instruments within three SDs above the mean (longer than 289 seconds for the Faces Test and longer than 306 seconds for the Emotions Survey) to help ensure that the analysis was able to capture the potential short-term effect of the reading.

In past studies that assessed author recognition with adults, participants selected relatively few false authors (e.g., Mar et al., 2006). The present study with juvenile participants had a higher rate of students selecting false authors. Recall that the number of real authors that participants select serves as an estimate for the participants' amount of reading. A selection of a greater number of real authors is correlated with greater reading exposure, while very few real authors is likely an indication of less reading exposure (Stanovich & West, 1989). Additionally, the selection of an inordinate number of false authors may be an indication of a participant not being faithful to the study. As a result, individuals who selected false authors more than six times, or three standard deviations above the mean, were excluded from the analysis.

Individuals who scored more than three *SDs* below the mean on the assessments

of ToM, which consisted of the DANVA (14 or lower out of 24), and the Faces Test (15 or lower out of 20) were excluded from the analyses. Extremely low scores on the ToM assessments may be an indication of low cognitive ToM ability or a lack of engagement.

There were 227 participants who completed both days of the study. Six individuals who selected more than six false authors on the ART were excluded ($N = 221$). An additional five individuals with times more than three SDs above the means on the Emotions Survey and the Faces were also excluded from the data ($N = 216$). An additional four participants were excluded who scored three SDs below the means on the Faces Test or the DANVA test. These exclusions resulted in a total of 212 participants or 93% of the original participant pool.

Participating Teachers

The researcher performed a series of analyses to examine possible teacher effects on the experimental variables. This study employs multiple regression analysis (Cohen et al., 2003) in each of the analyses of the dependent variables by applying control variables to step one and the experimental variable to step two. In order to determine if the participating teacher had an additional effect on student performance for the dependent variable, the teacher variable was added to step three of the regression (as in Wolters, 2004). Teacher assignment did not account for a significant amount of additional variance in step three of the regression on the Faces Test ($F(1, 206) = 1.180, p = .279, \Delta r^2 = .005$), for the regression on EC ($F(1, 207) = 1.143, p = .286, \Delta r^2 = .005$), and for the regression on the HPT scores ($F(1, 151) = .310, p = .578, \Delta r^2 = .001$). These three regressions demonstrate that the differences among the three teachers did not

significantly affect the dependent variables of the study.

Scoring Historical Perspective Taking Paragraphs

The HPT paragraphs were scored by two history experts with Master's degrees in history. Five percent of the HPT paragraphs were scored together; then, another 20% of the essays were scored separately. The graders then compared scores on the shared portion and resolved any differences to ensure that all scores were within a .5 point. The graders then divided and scored the remaining essays. Inter-rater agreement was at 79.68% on the 20% of items that both history experts scored. An independent samples t test on the items that the two scorers graded separately were not significantly different, demonstrating that neither grader systematically scored significantly higher or lower than the other, $t(91) = -.629, p = .531$. The HPT scores demonstrated acceptable limits of skewness and kurtosis.

Order Effects

The researcher checked for order effects on the Emotions Survey and the Faces Test. These instruments were counterbalanced and provided to students immediately following the text passage reading. The researcher applied the variable "order effect" to a Pearson Correlation analysis and could find no significant relationship with order effects and the Faces Test, $r(212) = -.010, p = .882$; or the Emotions Survey, $r(212) = -.061, p = .377$. As a result, the order effects variable was left out of subsequent analyses.

Correlation of the Instruments

The correlations, means, and standard deviations for each instrument are provided in Table 12. The Faces and the DANVA are the two ToM measurements. Empathic Concern (EC) was a subcomponent the Emotions Survey. Historical Perspective Taking (HPT) is the score that individuals received on the HPT written task. The Author recognition test (ART) is a proxy for reading exposure. The SRI is a standardized comprehension assessment. Document and narrative comprehension are the comprehension scores individuals received for reading their assigned text. Positive and Negative Emotions are two additional emotion subgroups of the Emotions Survey.

Although the Faces Test and the DANVA both entail the discerning of emotions from facial expressions, they appear to have a significant amount of variance that is not shared. The Faces test and the DANVA test are weakly correlated, $r(212) = .142, p < .05$. They are both negatively correlated with positive emotions (Faces Test & Positive Emotions $r(212) = -.158, p < .05$; DANVA & Positive Emotions $r(212) = -.192, p < .01$). This means that individuals with higher ToM ability tended to report lower positive emotions after reading the account of the Salem witch crisis. But only the DANVA demonstrates a negative relationship with the EC (DANVA & EC $r(212) = -.191, p < .01$), while the Faces Test is weakly positively correlated with the EC ($r(212) = .036, p = .602$). These statistics demonstrate that although these two instruments appeared to assess the same skill, namely discerning emotions from facial expressions, their subtle differences in timing, emotions, distractors, and actors may make it so that they actually do not measure the exact same cognitive ability.

Table 12

Means, Standard Deviations, and Pearson Correlations Among All Instruments

Instrument	N	M	SD	1	2	3	4	5	6	7	8	9	10
1 Faces Test	212	18.04	1.39		.142*	.036	-.039	.086	.033	-.029	.116	.007	-.158*
2 DANVA	212	21.11	1.86			-.191	.007	.129	.029	-.037	.140	.043	-.192**
3 Empathetic concern	212	14.60	3.95				.166**	.102	.052	-.008	-.022	.262**	.176*
4 Historical perspective taking	162	5.02	2.06					.307**	.478**	.157	.271*	-.002	.052
5 Author recognition test	212	8.54	4.73						.447**	.249**	.283**	.011	-.171*
6 Scholastic reading inventory	208	1224	142							.546***	.367**	-.060	-.132
7 Document comprehension	105	5.59	1.47									-.137	-.014
8 Narrative comprehension	107	6.83	1.10									.084	-.327**
9 Negative Emotions	212	29.15	6.54										-.256**
10 Positive Emotions	212	6.04	2.01										

* = Correlation is significant at the 0.05 level (2-tail).

** = Correlation is significant at the 0.01 level (2-tailed).

The correlational analysis demonstrated a positive relationship between EC and HPT, $r(162) = .166, p = .034$. Individuals in the narrative group demonstrated a stronger relationship of EC and HPT, $r(79) = .276, p = .014$, than individuals in the document group, $r(83) = .071, p = .522$. In other words, individuals who read the historical narrative demonstrated a stronger correlation with EC and HPT than individuals who read the historical documents.

The HPT, ART, SRI, document comprehension, and narrative comprehension variables all demonstrate significant correlations. This is likely because they are each connected to literacy ability.

In this sample, Empathic Emotions were correlated with both Negative Emotions such as “angry,” $r(212) = .262, p < .01$, and Positive Emotions such as “happy,” $r(212) = .176, p < .05$. It was surprising that Positive Emotions were correlated with Empathic Emotions. Upon further investigation the researcher realized that much of the correlation is due to particular items related to heightened emotions such as “excited”. In other words, as a person experienced Empathic Emotions for the people in the Salem witch crisis, they were also more likely to feel more excitement. Negative and Positive Emotions are negatively correlated, $r(212) = -.256, p < .01$.

Influence of Historical Texts on Theory of Mind

The first research question was: To what extent do different types of historical texts influence adolescents’ ToM? Readers of the historical narrative scored slightly higher than readers of the historical documents as measured by the Faces Test (narrative,

$N = 107$, $M = 18.12$, $SD = 1.40$; historical documents $N = 105$, $M = 17.96$, $SD = 1.37$).

The researcher created a regression to investigate if this difference was statistically significant. The ART, the DANVA, and gender were used as control variables in step one of the regression. Step one was statistically significant, $F(3, 208) = 3.891$, $p = .010$. Step two included the experimental variable, the text assignment. Text assignment did not explain a significant amount of additional variance, $F(1, 207) = 1.460$, $p = .228$, $\Delta r^2 = .007$. This indicates that there was not a statistical difference between the two text groups on the Faces Test. Finally, in step three, the teacher variable was added to the regression. The teacher variable did not account for a significant amount of additional variance, $F(1, 206) = 1.180$, $p = .279$, $\Delta r^2 = .005$, suggesting that the potential influence that the different teachers may have had on students' Faces Test scores, was not a statistically significant amount.

Evidence for a Potential Interaction

In three different graphs discussed below, there is evidence that the narrative text

Table 13 provides the coefficients for each step.

Evidence for a Potential Interaction

In three different graphs discussed below, there is evidence that the narrative text

Table 13

Coefficient Values for Three Steps of a Regression on the Faces Test

Variable	β step 1	β step 2	β step 3
ART	.088	.092	.082
Gender	.160*	.155*	.165*
DANVA	.087	.104	.103
Text		-.083	-.081
Teacher			.074
	$R^2 = .053^*$	$R^2 = .060$	$R^2 = .065$
	$\Delta R = .053^*$	$\Delta R = .007$	$\Delta R = .005$

group scored higher than the historical document group on the Faces Test at all levels

except for individuals who read longer, had high comprehension scores, and who had the

highest scores on HPT.

In a scatterplot on the Faces Test (y-axis) and reading times (x-axis) one can observe that at most times, the narrative readers scored higher on the Faces Test than the historical document readers. The primary exception is for those readers with times about a half of a *SD* above the mean as demonstrated in Figure 4.

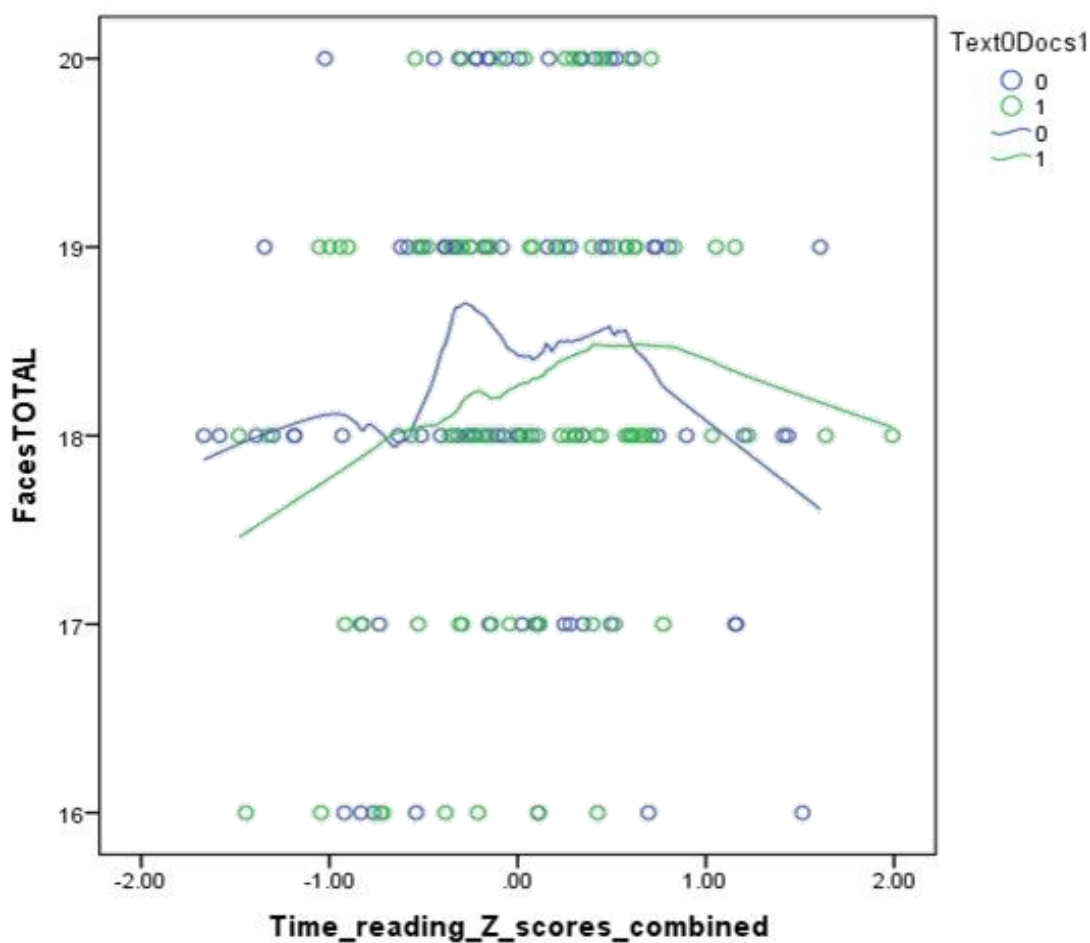


Figure 4. Scatterplot of the Faces Test (y-axis) and standardized time reading times (x-axis).

Furthermore, in a scatterplot on the Faces Test (y-axis) and HPT (x-axis) scores, Loess lines for each text group demonstrates that on all levels of the HPT scores, the

narrative group outscored the document group, except for the highest scores, of 8 and above (out of 9 possible). This is demonstrated in Figure 5.

In a separate scatterplot (see Figure 6) on the Faces Test scores (y-axis) and SRI scores (x-axis), it appears that the narrative group scored equal to or higher on the Faces Test for all levels of SRI scores.

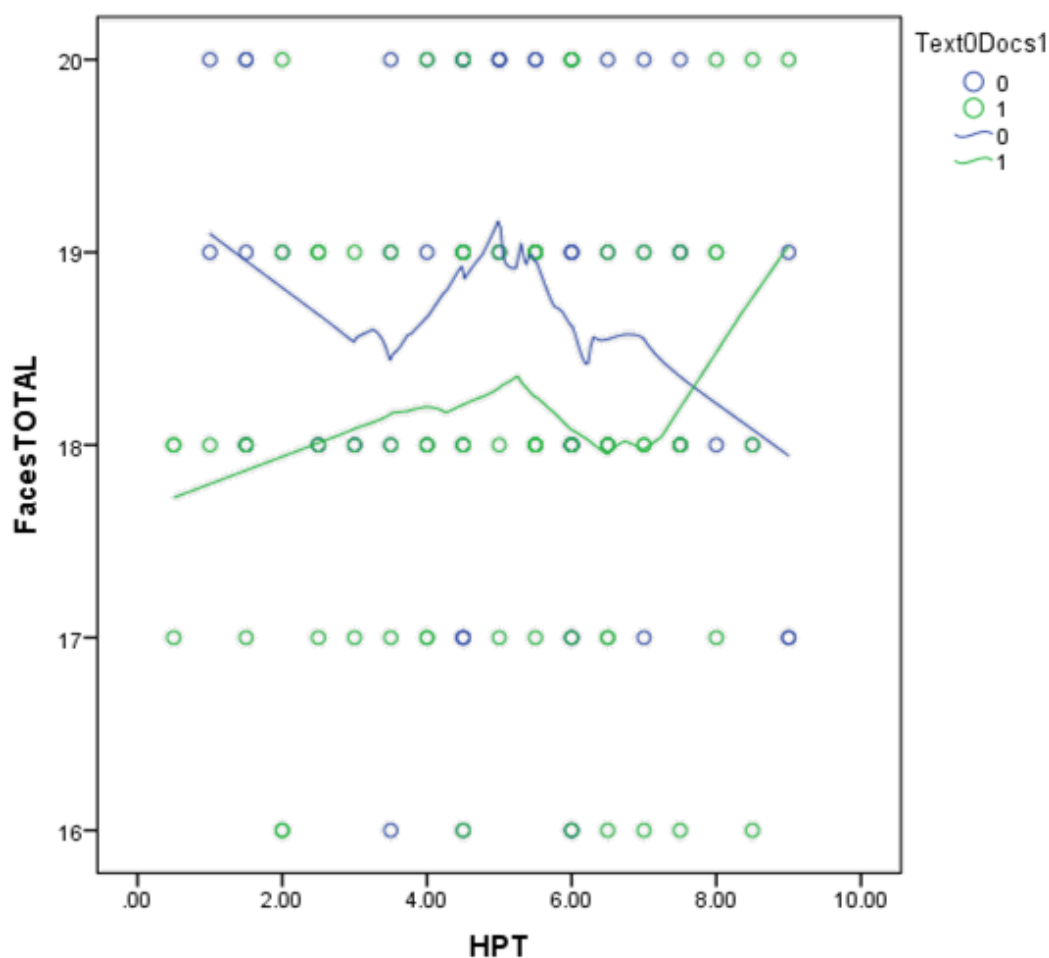


Figure 5. Scatterplot of the Faces Test (y-axis) and HPT scores (x-axis).

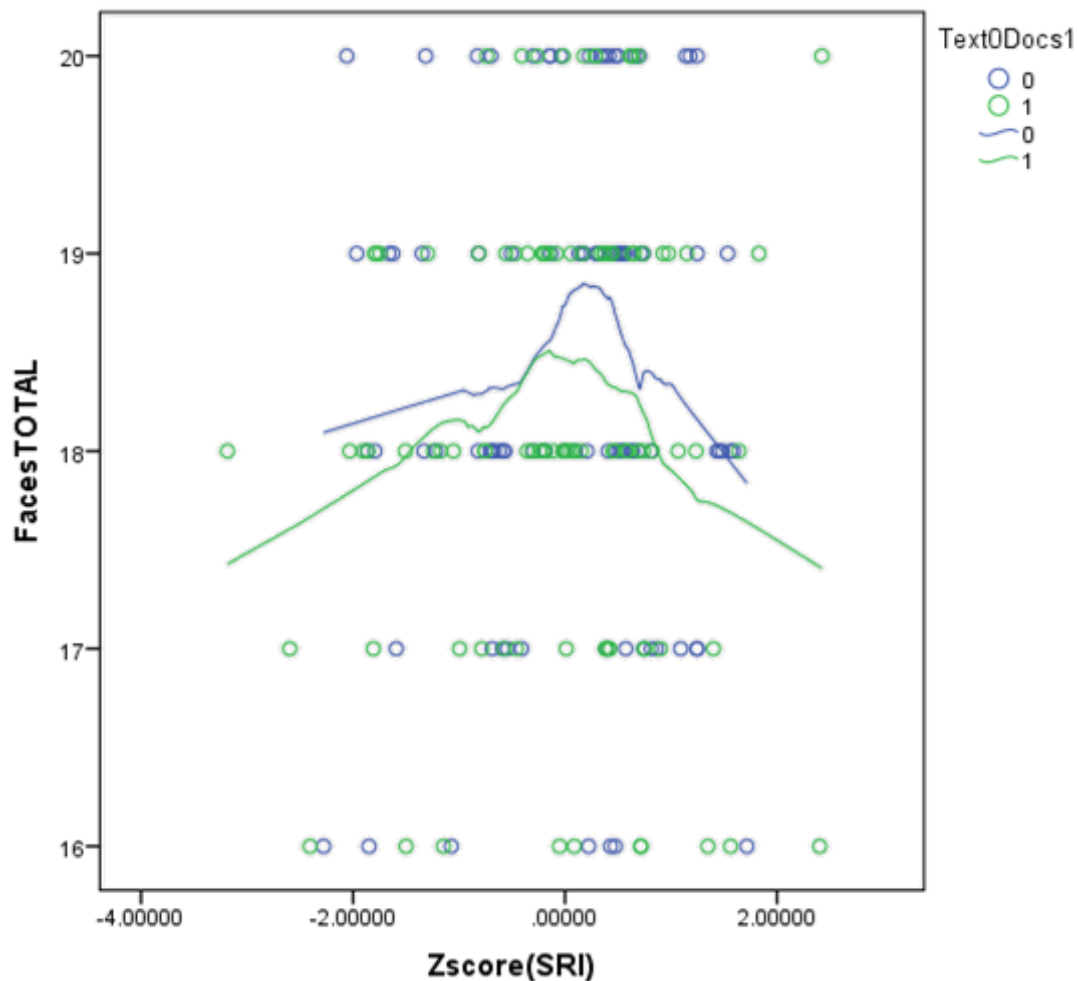


Figure 6. Scatterplot of the Faces Test (y-axis) and SRI scores (x-axis).

SRI scores alone did not fully predict how well students comprehended the text, neither did time reading fully predict how well students comprehended the text. The researcher created a variable of standardized reading times and SRI scores combined in order to gain a more complete perspective of when individuals scored higher on ToM and for what text groups. In a scatterplot on the Faces Test scores and Z score (SRI and time reading), the narrative group scored higher than the document group on the Faces Test at

all times combined with SRI scores except those that are approximately one SD above the mean (see Figure 7).

The regression analysis indicates that the difference in the Faces Test scores between the two text groups was not statistically significant. However, an analysis of three scatterplot graphs above provide more nuance. These three scatterplot graphs (Figure 5, 6, and 7) appear to indicate that for most times, individuals assigned to read the narrative text scored higher on the Faces Test. However, there appears to be a group of individuals assigned to read the historical documents with higher SRI scores and

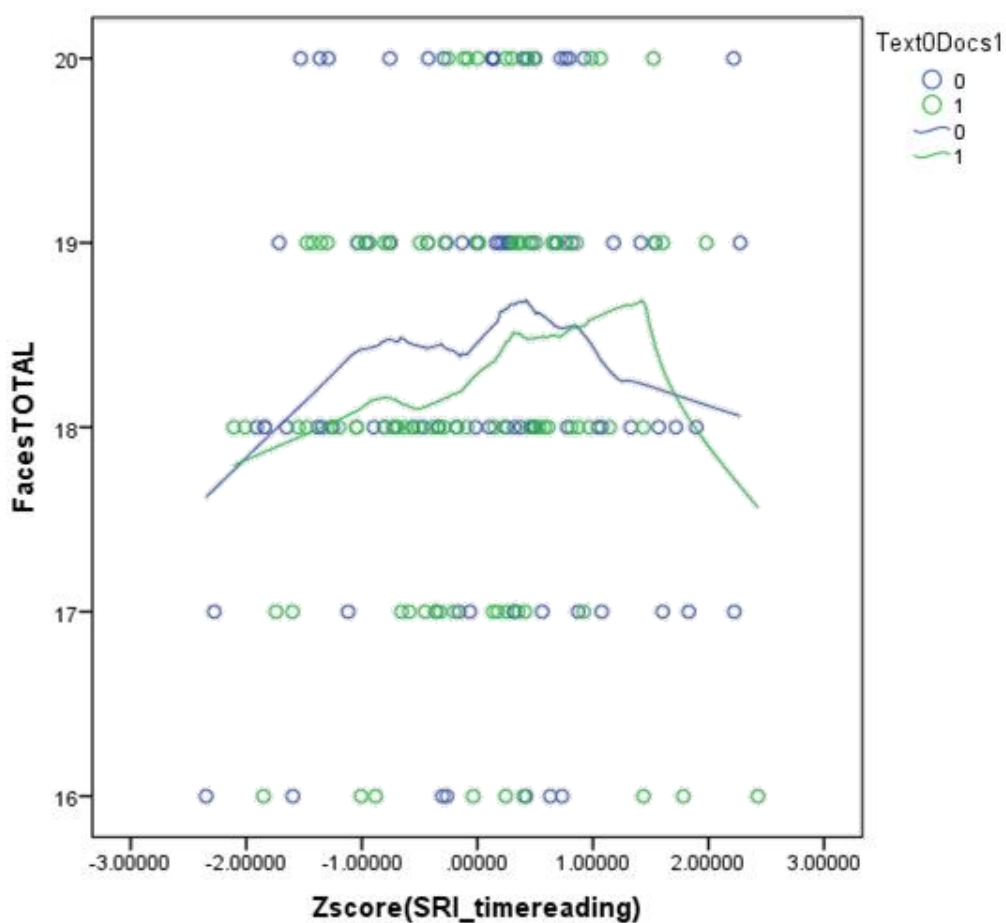


Figure 7. Scatterplot of the Faces Test (y-axis) and SRI and Time Reading Scores (x-axis).

higher reading times who appear to have outperformed comparable individuals in the narrative group on the Faces Test.

Influence of Historical Texts on Empathic Concern

The second research question was: To what extent do different types of historical texts influence adolescents' EC? Readers of the narrative account scored slightly higher than readers of the historical documents on EC as measured by the Empathic Emotions group in the Emotions Survey (narrative, $N = 107$, $M = 14.47$, $SD = 3.84$; historical documents $N = 105$, $M = 14.73$, $SD = 4.07$).

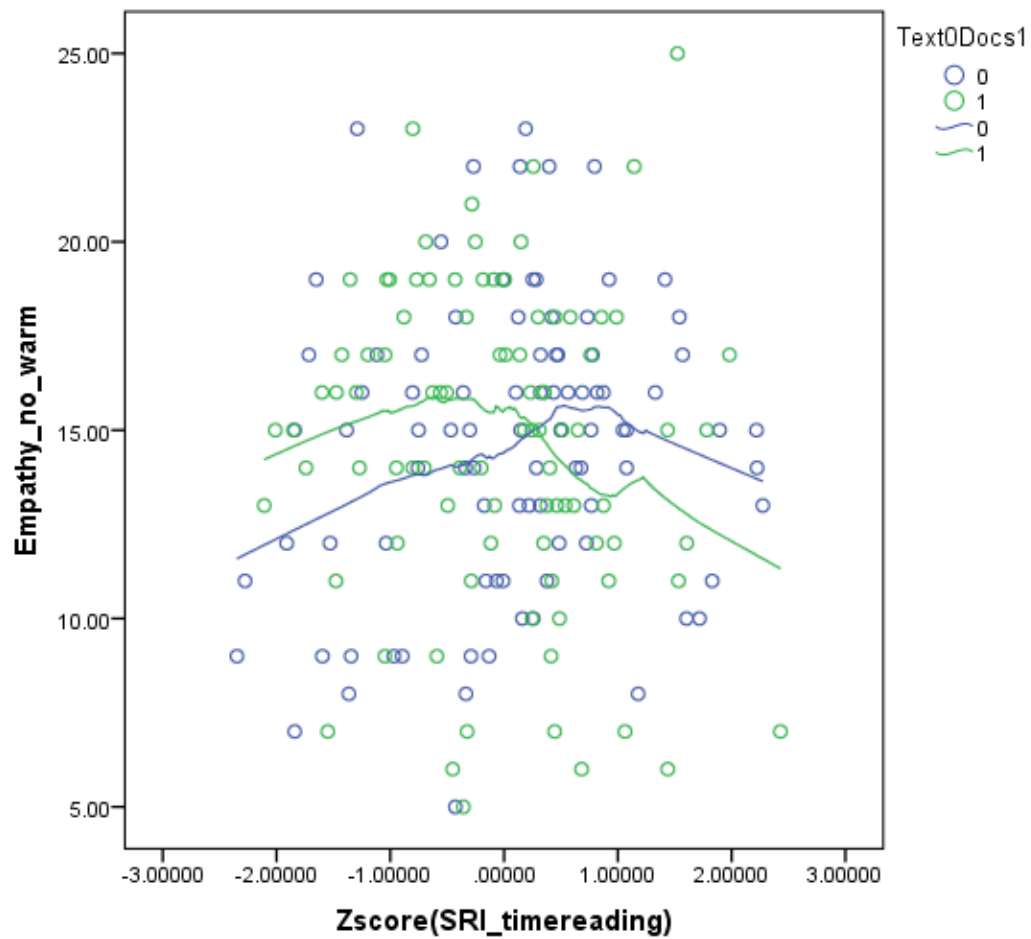
A regression investigated if this difference was statistically significant. The ART and the DANVA were used as control variables in step one of the regression. Step one was statistically significant, $F(2, 209) = 5.951$, $p = .03$. Step two included the experimental variable, the text assignment. The text assignment did not explain a significant amount of additional variance, $F(1, 208) = .990$, $p = .321$, $\Delta r^2 = .004$, suggesting that there was not a statistical difference between the two text groups on Empathic Emotions. Finally, in step three, the teacher variable was added to the regression. The teacher variable did not account for a significant amount of additional variance, $F(1, 207) = 1.143$, $p = .286$, $\Delta r^2 = .005$. This indicates that the potential teacher effect may have had on students' Faces Test scores, was not a statistically significant amount. Table 14 provides each of the coefficients for the regression.

The researcher created a scatterplot with Empathic Emotions as the y-axis and SRI scores combined with reading times as the x-axis (*Figure 8*). The Loess line for each

Table 14

Coefficient Values for Three Steps of a Regression on Empathic Emotions

Variable	β step 1	β step 2	β step 3
DANVA	-.207*	-.220*	-.218*
ART	.133*	.131	.122
Text Type		.068	.070
Teachers			.072
	$R^2 = .054^*$	$R^2 = .058$	$R^2 = .064$
	$\Delta R = .054^*$	$\Delta R = .004$	$\Delta R = .005$

*Figure 8.* Scatterplot of empathic emotions (y-axis) and SRI and Reading Times (x-axis).

group indicates a similar inverse parabolic relationship. For the historical document group, individuals who had low SRI scores and who read for less time reported more empathetic emotions; however, this peaked and decreased much sooner than the historical narrative group.

Influence of Historical Texts on Historical Perspective Taking

The third research question was: To what extent do different types of historical texts influence adolescents' HPT? Readers of the narrative account scored slightly higher than readers of the historical documents on HPT (narrative, $N = 79$, $M = 5.1$, $SD = 1.98$; historical documents $N = 83$, $M = 4.9$, $SD = 2.14$).

The researcher created a regression to investigate if this difference was statistically significant. The SRI, ART, time reading, and gender were used as control variables in step one of the regression. Step one was statistically significant, $F(4, 153) = 14.84$, $p = .000$. Step two included the experimental variable, text assignment. Text assignment did not explain a significant amount of additional variance, $F(1, 152) = .155$, $p = .694$, $\Delta r^2 = .001$, indicating that there was not a statistical difference between the two text groups on HPT. Finally, in step three, the teacher variable was added to the regression. The teacher variable did not account for a significant amount of additional variance, $F(1, 151) = .310$, $p = .578$, $\Delta r^2 = .001$. This indicates that the potential influence of teacher effect on students' HPT scores, was not a statistically significant amount. The regression coefficients for each step are included in Table 15.

In order to investigate each category of the rubric, the researcher ran the same

Table 15

Coefficient Values for Three Steps of a Regression on Historical Perspective-Taking

Variable	β step 1	β step 2	β step 3
SRI	.443**	.447**	.449
ART	.121	.118	.123
Time Reading	.119	.121	.097
Gender	.118	.119	.119
Text Type		.028	.026
Teachers			-.046
	$R^2 = .280^*$	$R^2 = .280$	$R^2 = .282$
	$\Delta R = .280^*$	$\Delta R = .001$	$\Delta R = .001$

regression for each category. Individuals assigned to read the historical documents scored statistically higher on the amount of evidence provided, $F(1, 150) = 5.050$, $p = .026$, $r^2 = .028$. No significant difference was found between group assignment and the other components of the HPT rubric, which are perspective taking, $F(1,150) = .992$, $p = .321$; context, $F(1,150) = .007$, $p = .935$; and inference, $F(1,150) = .531$, $p = .467$. The researcher ran the same regression on HPT without evidence provided, which only included the combined scores for inference, context, and perspective-taking. In this regression, text assignment accounted for much less variance than HPT with evidence included, $F(1, 150) = 510$, $p = .476$. This suggests that most of the difference in scores between the two text groups was caused by scores on the single component of evidence. Including evidence was the least related component of HPT to empathy.

It is interesting to note that of the three emotion groups (Positive Emotions, Negative Emotions, and Empathic Emotions), the only emotion group correlated with

HPT is Empathic Emotions (Empathic Emotions with HPT $r(212) = .166, p < .01$). A Pearson correlation was conducted for each text group and each component of the HPT, the complete HPT without evidence, and word count. When only the narrative group participants were selected, the correlation was much stronger for HPT, $r(79) = .276, p = .014$; and was much weaker when only the historical document group was selected, $r(83) = .071, p = .522$. These results are displayed in

Table 16.

In order to understand some of the relevant characteristics of participants who scored higher on HPT scores, the researcher performed a series of Pearson correlations and ANOVAs. Individuals with higher HPT scores also scored higher on the SRI (ANOVA: $F(17,157) = 3.462, p < .001$; Pearson Correlation: $r(158) = .478, p < .001$), and the ART (ANOVA $F(17, 161) = 2.124, p = .009$; Pearson Correlation: $r(162) = .307, p < .001$). In addition, individuals who scored higher on HPT, self-reported higher levels of Empathic Emotions (narrative ANOVA: $F(16, 78) = 1.189, p = .302$; narrative Pearson

Table 16

Correlations for the Historical Perspective-Taking Components, Word Count, and Empathic Emotions

Component	Empathic emotions for the narrative group	Empathic emotions for the historical document group
Context	.218	.084
Perspective	.294*	.014
Inference	.348**	.108
HPT (context, perspective & inference combined)	.322*	.079
Word Count	.288*	-.051

Note. Narrative HPT $N = 68$; Historical Document HPT $N = 77$; Narrative word count $N = 95$; Historical Document word count $N = 97$.

* $p < .05$.

** $p < .01$.

correlation ($r(79) = .276, p = .014$; document ANOVA ($F(16, 75) = 1.700, p = .061$; document Pearson correlation ($r(93) = .085, p = .418$). There did not appear to be any relationship with high HPT and high scores on the Faces Test or the DANVA (Faces Test: $F(17, 159) = .623, p = .870$; DANVA: $F(17, 154) = .801, p = .801$). One of the highest correlations with HPT was word count ($r(162) = .597, p < .001$). This indicates that after individuals self-reported higher levels of Empathic Emotions, they went on to write more words in their paragraphs.

Post-Hoc Analyses

The researcher performed three post-hoc analyses in order to better understand some curious trends with the data. These include the amount of reading time for the participants with different SRI scores, a phi analysis of author recognition with the Faces Test and Empathic Emotions, and an analysis of birth sex.

Time Reading as a Variable

There was not a statistical difference for the amount of time that each text group spent reading the texts, $t(210) = .453, p = .651$. However, there is evidence for a difference in how long students with different SRI scores read each of the different texts.

The researcher standardized the Z scores for each of the three eighth-grade history classes and then summed them. A scatterplot was created with SRI scores as the y-axis and time reading as standardized Z score as the x-axis for each of the text groups. These two scatterplots are portrayed in Figure 9 and Figure 10.

These two scatterplots show that students with the lowest SRI scores on average

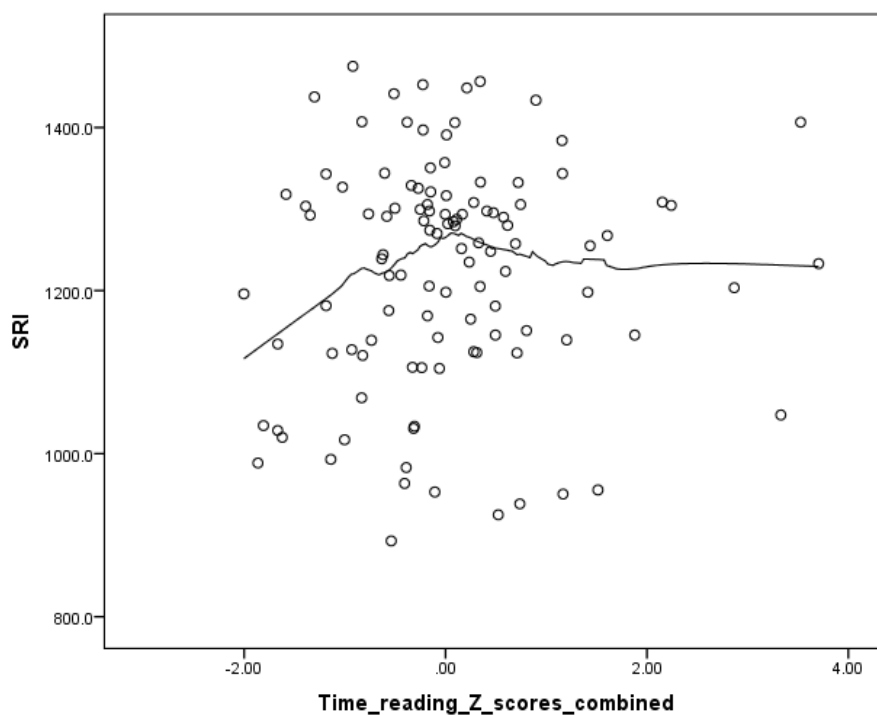


Figure 9. Scatterplot of SRI (y-axis) and Time Reading (x-axis) for the Narrative Group.

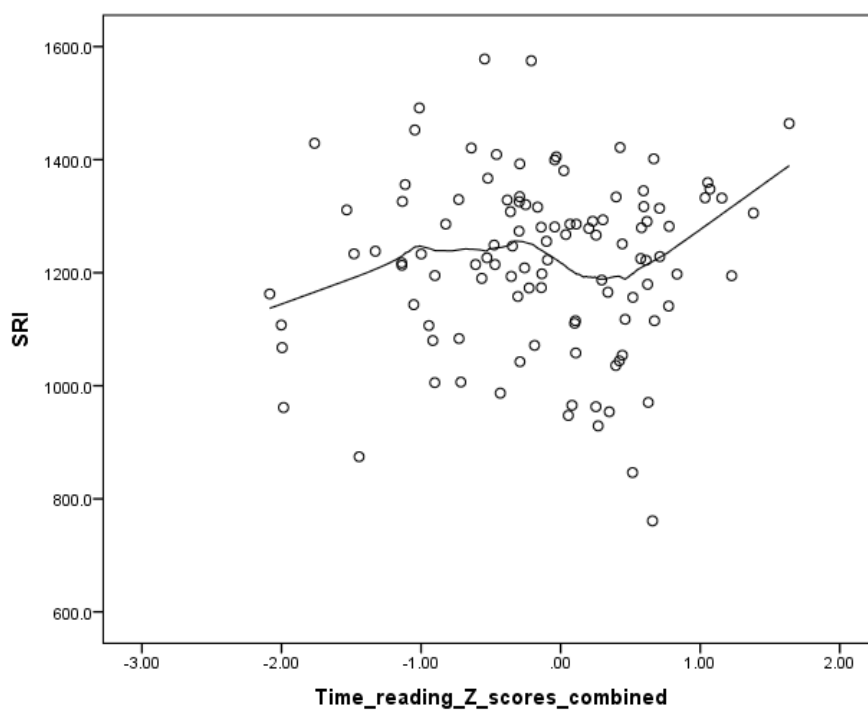


Figure 10. Scatterplot of SRI (y-axis) and Time Reading (x-axis) for the Document Group.

also read for the shortest lengths of time. In the narrative group, individuals with the highest SRI scores appear to have read for an average length of time, after which, there appears to be an association with lower SRI scores and longer reading times.

In contrast, Figure 10 for the historical document group, indicates that individuals with the highest SRI scores appear to have read for some of the greatest amounts of time. To ascertain if the final steep incline in Figure 10 was statistically significant, the researcher selected all participants who read the historical documents and read for a longer time than the mean and less than two *SD* above the mean, and then performed a Pearson correlational analysis. The results showed that SRI scores above the mean and below two *SD* were positively correlated for historical document readers, $r(38) = .31, p = .030$. In contrast, for narrative readers who read for a longer time than average and for less time than two *SD* above the mean, a negative relationship between longer reading and higher SRI scores, $r(44) = -.264, p = .084$, was found. These two Pearson correlations suggest that after the mean, higher reading times are associated with lower SRI scores for the narrative group, while higher reading times are associated with higher SRI scores for the document group.

Based on these findings, the researcher hypothesized that the narrative group would have higher SRI scores in the first *SD* above the mean, and the historical document group would have higher SRI scores in the second *SD* above the mean. In order to test this hypothesis, the researcher performed a *t* test with only participants who read longer than the mean and less than one *SD* above the mean. The findings showed that readers of

the historical narrative text averaged higher SRI scores, $t(76) = 2.682, p = .009$. However, when only those participants who read for at least one *SD* above the mean and less than two *SD* above the mean were included in the analysis, a statistical difference was found between historical document readers and narrative readers in the SRI scores, with historical document readers having higher SRI scores, $t(12) = -1.898, p = .082$. In the second *SD* above the mean, historical document readers averaged higher SRI scores.

Again, it is important to note that there is not a statistically significant difference between average reading times for the narrative or historical documents text groups as a whole, $t(210) = .453, p = .651$. Rather, these analyses (the scatterplot graph, the Pearson regressions, and the t tests), all provide insights for differences in how students of different reading abilities read the two different historical texts. Specifically, students with higher SRI scores read the narrative text for an average length of time, after which, longer amounts of time were associated with students with lower SRI scores. In contrast, students with higher SRI scores read the historical documents for the longest amounts of time on average.

Analysis of Specific Authors

Past studies have found a consistent positive relationship with higher ToM scores and higher author recognition scores for specific authors with adult participants (Mar et al., 2006, Kidd et al., 2016). The researcher sought to ascertain if this was also true for adolescent participants and for what authors. The researcher investigated if recognition of individual authors demonstrated a relationship with the Faces Test and Empathic Emotions. The researcher used a phi test in SPSS, which explains the strength of a

relationship of a dichotomous categorical variable with another variable. In this case the dichotomous categorical variable was whether the participants selected the author or not, and the other variable was their Faces Test score. This analysis revealed six authors with a p value equal to or less than .1 in a phi test with the Faces Test as displayed in Table 17.

One of the authors, Jeff Kinney, has a negative relationship with the Faces Test.

A phi analysis on authors and empathic concern found that eight authors had a p value equal to or less than .1 (see

Table 18). Past research has emphasized the relationship of specific authors with ToM (Mar et al., 2006, Panero et al. 2016). The present analysis indicates that there may also be a relationship for the recognition of specific authors as a proxy for reading exposure and the disposition for feeling empathic concern. In other words, the authors that a person recognizes appears to connected to a person's disposition for experiencing Empathic Emotions.

Table 17

Phi Test with the Faces Test and Select Authors

Author	Phi	<i>P</i>
Brandon Sanderson	.262	.024
Jeff Kinney	.260	.026 (negative relationship)
JK Rowling	.215	.134
R.L. Stine	.237	.064
Scott O'Dell	.233	.075
Scott Westerfeld	.244	.050
S. E. Hinton	.247	.045

Table 18

Phi Test with Empathic Concern and Select Authors

Author	Phi	<i>P</i>
James Dashner	.382	.041
Jane Austin	.401	.018
John Flanagan	.513	.000
L. M. Montgomery	.347	.145
Lois Lowry	.356	.108
Louis Sachar	.414	.010
Margaret Peterson Haddix	.389	.031
Orson Scott Card	.460	.001
Stephanie Meyer	.464	.001
Veronica Roth	.366	.077

Birth Sex of Participants

One of the biggest differences identified in this study was the difference between adolescents who were born male or female. Participants for this study included 134 females and 103 males. Females outperformed males on the ToM assessments (Faces Test, female: $M = 18.27$, $SD = 1.420$; Faces Test, male: $M = 17.74$, $SD = 1.281$; DANVA female $M = 21.55$, $SD = 1.816$; DANVA male $M = 20.52$, $SD = 1.747$; Faces Test: $t(210) = -2.838$, $p = .005$; DANVA: $t(210) = -4.138$, $p < .001$).

Evidence suggests that female participants completed the study with greater fidelity. Females wrote 42 more words on average than males on the writing task (word count female $M = 236$, word count male $M = 194$; $t(210) = -2.838$, $p = .005$), invested more than a minute longer in writing and more than two minutes in reading, and scored higher on the text-specific comprehension questions, $t(164) = -1.934$, $p = .055$.

Furthermore, females completed the entire study more than 234 seconds slower on average (almost 4 minutes $t(210) = -2.107, p = .036$). In addition, males were significantly more likely to self-report experiencing more Positive Emotions after reading the account of the Salem witch crisis, which may also be an indication of a lack of engagement, $t(210) = 2.819, p = .005$. Most importantly, females scored significantly higher than males for the Faces Test (Cronbach's alpha of females on the = .329, Cronbach's alpha of males = -.13). The higher Cronbach's alpha score indicates that females demonstrated much greater reliability on the Faces Test than males.

Finally, males and females did not significantly differ in literacy ability as measured by the ART (ART ($t(210) = -.803, p = .423$) and the SRI ($t(210) = .437, p = .663$). This indicates that the differences in reading and writing were likely differences of fidelity rather than literacy ability. Based on nine measures (word count, time reading, time writing, text-specific comprehension questions, total time, Positive Emotions, Faces Test, the DANVA, and Cronbach's alpha on the Faces) it appears that females completed the study with greater fidelity. The means and *SDs* are shown in Table 19.

Table 19

Means and Standard Deviations for Adolescent Males and Females

Sex	<i>N</i>	Faces			DANVA		Word count		Reading time		Comprehension questions (<i>z</i>)		Positive emotions		ART	
		<i>M</i>	<i>SD</i>		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Males	91	17.74	1.28		20.52	1.75	194	106	510	384	-.07	1.02	10.71	2.77	8.24	4.57
Females	121	18.23	1.42		21.55	1.82	236	112	641	415	.17	.79	9.64	2.74	8.77	4.84

CHAPTER V

DISCUSSION

In this study individuals were randomly assigned to read one of two different texts: A historical narrative, or a collection of historical documents, both pertaining to the Salem Witch Trials. After the participants read their assigned text they were assessed on theory of mind (ToM), empathic concern (EC), and historical perspective-taking (HPT). A statistical difference between the two groups would have provided causal evidence that text type had differential effects on the dependent variables: ToM, EC, or HPT. No statistical difference was found between the two text groups on any of the three dependent variables. A lack of statistical difference between the two text groups can indicate one of three likely possibilities: the instruments lacked sufficient sensitivity, the texts did not impact the dependent variable in question, or the texts impacted the dependent variable equally.

Although the results for the research questions were null, this study contributes to our understanding of the use of historical texts in the history classroom with adolescents. In the next section, the researcher will provide evidence (from the Coh-metrix analysis as well as textual evidence) for why it appears that students of different estimated comprehension abilities appeared to have read the text for different lengths of time. Then the researcher will discuss the relationship of reading certain texts and experiencing EC and HPT. Next, the researcher explains the challenges of reliability in measuring ToM for an adolescent population and then describes the challenge of using authors as a proxy for reading exposure for adolescents. Finally, the researcher concludes with a discussion of

the limitations, future research and a conclusion.

Textual Analysis

According to the study design, one group of participants was randomly assigned to read the historical documents and another group was randomly assigned to read the historical narrative. Consequently, differences between the groups can more likely be attributed to the differences in the texts. Although no statistical differences for the dependent variables between the two text groups were found, there was a difference in reading times for participants with higher SRI scores (see Figures 9 and 10). Subsequent statistical analyses revealed that individuals with higher SRI scores read the narrative text for an average length of time, while individuals with higher SRI scores tended to read the historical documents for longer amounts of time.

Before the study was administered, the two texts received comparable overall Lexile scores and Flesch-Kincaid grade levels. However, these two overall measurements may have been insufficient to compare the two texts. Individuals with higher SRI scores may have read the documents for longer periods of time, because the historical documents may have been more difficult to comprehend. Below the researcher discusses the Coh-metrix evidence, textual evidence, and evidence from historical pedagogy that provide support that the historical documents may have been more difficult to comprehend.

Coh-Metrix Analysis

Syntactic simplicity, word concreteness, and deep cohesion. The historical narrative

text scored higher than the historical documents on three Coh-metrix categories: syntactic simplicity; word concreteness; and deep cohesion (see

Table 4). A high score for syntactic simplicity indicates that sentences tend to be short and simple (Graesser, McNamara, & Kulikowich, 2011). Short simple sentences are typically easier for readers to construct a mental model of the propositions (Kintsch, 1998). In addition, short simple sentences tend to place less demand on the reader's working memory because there are fewer pieces of information that the reader has to process at a time (Kintsch, 1998).

Second, the Coh-metrix analyses explains that the high score in word concreteness indicates that words in the historical narrative were less abstract and had high imageability. Less abstract references are typically easier to comprehend because the reader can more effortlessly connect concrete references with their background knowledge (Graesser et al., 2011; Kintsch, 1998). For example, one neurological study found that participants had a "more accessible semantic network for concrete words than for abstract words" (Fliessbach, Weis, Klaver, Elger, & Weber, 2006, p. 1413). Third, the high score in deep cohesion indicates more explicit causal relationships, which are also typically easier to comprehend (Graesser et al., 2011).

Narrativity for the historical document. In contrast, the historical documents only scored significantly higher than the historical narrative in one Coh-metrix category: narrativity. The Coh-metrix analysis explains that a score high in narrativity indicates that it is more "story like," which is typically easier for readers to comprehend (Graesser, et al., 2011). This finding is surprising, considering that the documents do not have a single author, nor do they explicitly include a narrative arc of setting, conflict, and resolution.

The documents do not provide a continuous thread of events. In addition, the historical documents do not have as many causal references as the historical narrative, which, according to the deep cohesion Coh-metrix score, is an important aspect of narrativity (Wolf, 2008).

Instead, the historical documents included in this study are a collection of seven documents that are disparate, disjointed, and disconnected. The historical documents would likely be classified as “traces” (Seixas & Peck, 2004) because they lack context and they do not provide a complete narrative. For example, the first document is a short excerpt of a sermon about the reality of witchcraft. The second document is a diary entry about a girl who appears to be bewitched. There is nothing in the text that explicitly connects the two documents into a coherent story. The third document is trial transcript about a girl being interrogated by a judge. Again, there is nothing that explicitly connects the trial transcript with the diary entry or the sermon.

Thompson (1999) defines narrative as a “chain of events, occurring in time and space, and linked by cause and effect” (p. 10). In the case of the documents, it can be argued that this chain of events is broken between each document and readers are left to construct the narrative or the chain of events for themselves.

Narrativity for the historical narrative. On the other hand, the narrative text is written by a single author, intentionally as a story for children (Zumbusch, 2009). In this study, the historical narrative would likely be classified as an ‘account’ because it provides a complete narrative of the story from beginning to end (Seixas & Peck, 2004). The narrative text follows a narrative arc. It explicitly describes the setting (e.g., “The

Salem witch-hunts took place in Eastern Massachusetts. In the 1600s...”). It introduces the conflict (e.g., “the Parrises lived peacefully until...[their daughter and niece] started to have fits”). It explains how the conflict builds (e.g., “Tituba, Good, and Osborne were put in jail, but the matter did not end there”). Then the conflict climaxes (e.g., “In the following months, eighteen people were hanged for witchcraft”), and finally resolves (e.g., “the governor forgave everyone still jailed for witchcraft”).

This plot structure resembles Freytag’s (1900) dramatic arc, which is used in countless nursery stories, plays, movies, and story plots (Cutting, 2016), that participants were likely well-acquainted with. Therefore, the historical narrative could have worked as a textual schema, which Kintsch also defines as ‘rhetorical superstructures’ wherein each of the different elements of the story intuitively fit into the student’s mental structure of a typical narrative (Anderson, 1984). A familiar schema provides an easy way for a text to activate a reader’s background knowledge and for the reader to understand how to structure the information in the text (Anderson, 1984). Kintsch (1998) explains that textual schemas are helpful in the formation of the textual macrostructure and the situation model. He further explains that “narratives in our culture have a basic exposition-complication-resolution structure” that are used by children as young as four years old (Kintsch, 1998, p. 68).

Furthermore, each of the elements of the story in the historical narrative appear to be more connected as a single thread or chain that travelled through the story, which is another essential component of narrative (Thompson, 1999). For example, the text discusses how the conflict culminates into a crisis where 19 people were executed and

five died in prison. The narrative text continues:

In time, people started doubting the accusations. The girls started naming even more powerful people. Increase Mather a powerful Minister questioned if stories about being hurt by specters [or a witch's spirit] should even be used in trials. People were then found innocent. In May 1693, the governor forgave everyone still jailed for witchcraft. (p. 18)

From this passage, a reader may be able to understand that accusations directed towards more powerful people, as well as skepticism of the use of evidence of witnesses testifying that they saw someone's specter attack them, began to be challenged, which led to the dénouement of the crisis. These causal references appear more explicit in the historical narrative than in the historical document. Considering that the narrative text followed a narrative arc, more closely connected the separate elements of the story in a single chain of events, and received a higher score for causation, it is surprising that it scored lower in narrativity.

Coh-metrix narrativity proxy. The Coh-metrix analysis is a powerful tool, but it is not infallible (Graesser et al., 2011). The creators of the Coh-metrix analysis warn about its limitations, explaining that “computers obviously cannot identify and scale texts on all levels of language, discourse, and meaning” (Graesser et al. 2011, p. 223). Furthermore, they report that the Coh-metrix analysis only accounts for 67%, or about two thirds of the variance (Graesser et al. 2011). This means that although the Coh-metrix analysis is an impressive and powerful analysis, it still has plenty of room (about 33%) for error.

The Coh-metrix analysis appears to fail to capture one of the most salient characteristics of the historical documents: their fragmentary disjointed nature. The

documents included in this study jump from a speech to a diary entry, to a courtroom interrogation, etc. This highlights a major weakness of a computer analysis that cannot comprehend the meaning and therefore cannot follow the thread of a story. Instead of following the thread of a story, the creators of the Coh-metrix analysis, measured components of thousands of texts, and found that typical narrative texts include a greater number of verbs, adverbs, pronouns, personal and third-person pronouns, higher word frequency of specific words, and references to intentional actions (Graesser, et al. 2011). Most narrative texts in general likely have these characteristics, because they are usually about people (Zunshine, 2006).

The Coh-metrix description of narrativity is appropriate for the historical documents, which include a diary description of a girl apparently bewitched, two petitions written from first-person perspectives, two court transcripts that are essentially dialogues of speech, and a public confession. The documents are full of verbs, personal and third-person pronouns, references to intentional action, and frequently repeat words like “witch” and “witchcraft.” For example, the researcher counted 94 personal and third-person pronouns in the historical documents and only 29 in the historical narrative.

Therefore, the historical documents may be more narrative in the sense that they contain more human elements such as personal pronouns, verbs, and intentional action. But they are not narrative in the sense that they do not provide a well-connected account that explicitly explains the events of the crisis from beginning to end. In contrast, the historical narrative may be less narrative in the sense that it provides more information and has fewer human references in the forms of pronouns, verbs, and intentional action,

but it is more of a narrative in the sense that it provides an account of the Salem witch crisis from beginning to end, and more closely bridges each of the elements of the story together for the reader. As a result, even though the historical documents score higher in narrativity one can be justifiably skeptical about whether they are actually more ‘story-like’ and easier to read.

Likely Demands on Background Knowledge and Inferencing

The historical narrative was intentionally designed for a specific audience, where assumptions about background knowledge are more likely to be accurate. The narrative explains concepts that would likely have been unfamiliar to the reader such as who the Puritans were and their understanding of witchcraft, before using these concepts later in the text. Then the text moves forward and builds on the concepts in a logical sequence.

In contrast, historical documents by their very nature are almost never originally constructed for adolescent students living in the present age. Much of the information the reader needs in order to comprehend the documents is not explicitly contained in the text. For example, in the first document, a sermon by an “influential” preacher, the preacher declared the reality of witchcraft. The text itself did not explain what a sermon was, nor did it explicitly state what kind of influence such a sermon could have had on Puritan society. The reader would have had to infer from the sermon that most Puritans believed in witchcraft, expected to find witchcraft, and that Puritans used witchcraft as an explanation for tragedies or strange behavior. This is likely a big inferential jump for an adolescent reader.

The next document, a diary entry of a minister visiting the Parris household, describes the strange erratic behavior of Mr. Parris' niece as she flapped her arms, ran into the fire and said she saw people who were not present. The documents included an introduction that says that "girls were powerless in Puritan society." From this description, students may have been able to infer the thrill that this powerless female adolescent must have received by commanding the rapt attention of her father and other powerful leaders of the community, by exploiting their belief in witchcraft and acting out as if she were bewitched. Again, this would have been a challenging inference.

The intertextual and intra-textual (across documents) inferences that were required for participants to comprehend the historical documents would likely have placed greater demand on students' background knowledge because the documents leave so much more of the context out of the story. Without the context provided and explained, the readers must fill-in the gaps, primarily according to their own background knowledge (Ashby & Lee, 1987; VanSledright, 2004). The readers would have to juggle several bits of information in their short-term working memory (Kintsch, 1998) in order to connect information across documents. To make these connections would likely have required close reading and rereading (Wineburg, 2001). However, the similarity in reading times for participants of the two different texts indicate that most students read the historical documents for the same amount of time as the historical narrative. They likely read the documents just as if they were reading a narrative text, straight through one time, from beginning to end.

Reading Like a Historian Requires Scaffolding or Support

Analyzing the documents entails what researchers have termed as reading or thinking like a historian (Nokes, 2013; Wineburg, 2001). Crucially, researchers have emphasized the importance of providing scaffolding or support to help students read like historians. Foster (2001) speaks of the importance of allowing students to work in small groups, in conjunction with whole-class discussions and small-group discussions. He speaks of the importance of the teacher in “selecting appropriate materials, asking probing questions, stimulating thoughtful investigation, leading whole-class discussions, and maintaining the momentum of inquiry” (Foster, 2001, p. 178). Nokes (2014) demonstrated that with sufficient scaffolding, students as young as 10 years-old are capable of reading historical documents. This study did not provide support or scaffolding to the students as recommended by previous research (Foster, 2001; Nokes, 2013; Wineburg, 2001). Therefore, students were left on their own to construct coherent meaning across the separate documents.

It appears that without scaffolding or support, only a select group of participants in the present study invested the extra time that a close reading of the historical documents may have required for comprehension and commensurate activation of ToM. Previous research provides evidence that greater support and scaffolding can lead to greater engagement and higher quality analyses of historical documents (Foster, 2001; Nokes, 2013; Nokes & De La Paz, 2018; Wineburg, 2001).

Textual Analysis Conclusion

Above, the researcher provides evidence that the narrative was likely easier to read. This evidence is from a Coh-metrix analysis that showed that the narrative scored higher than the documents on syntactic simplicity, word concreteness, and deep cohesion. Furthermore, the researcher provides textual evidence that the narrative followed a narrative structure, and that a Coh-metrix analysis that gave the historical documents a higher score in narrativity may be problematic. Finally, the researcher provides evidence from historical pedagogical experts that students typically require scaffolding and support to successfully read historical documents.

Researchers and theorists explain that ToM is activated by literary fiction because comprehension of the text challenges the readers to construct socially complex situation models, make social inferences, and engage in nested minds (e.g., imagining a person thinking about another person; Kidd & Castano, 2013, 2016; Mar & Oatley, 2008; Zunshine, 2006). Based on the Coh-metrix analysis score for narrativity, the historical documents seem to share more of these characteristics than the narrative text in that the documents were higher in personal pronouns, verbs, and intentional action.

However, if the historical documents were more difficult to comprehend, then fewer students would likely have successfully comprehended the text. If fewer students comprehended the text, then fewer students would have constructed social situation models of the text and would likely not have activated their ToM. If the historical documents were indeed more challenging for most participants, this might explain why relatively few readers of the historical documents scored higher in ToM (see Figure 4, 5,

and 7). It is interesting to note that according to these same three graphs, a subgroup with higher SRI scores, who read the historical documents longer, also scored higher on the Faces Test than the narrative participants.

Theory of Mind, Empathic Concern, and Historical Perspective-Taking

Empathy is often promoted as if it were a single construct, but empathy can be defined in several different ways. For example, Batson (2011) found eight major different uses of the term in the research literature. In addition, some researchers have directly linked historical empathy and clinical empathy (Endacott & Brooks, 2013, 2018). This study provides evidence that ToM is separate and distinct from EC, and that ToM and EC are distinct and separate from HPT.

Despite researchers equating historical empathy with clinical empathy, as earlier shown in **Error! Reference source not found.**, the Faces Test only demonstrated a weak, not statistically significant, positive relationship with self-reported empathic concern, $r(212) = .036$, $p = .602$, and a weak, not statistically significant, negative relationship with HPT, $r(161) = -.039$, $p = .621$. This may be due to a lack of sensitivity with the HPT rubric and scorers, lack of sensitivity with the ToM assessment, as well as the fact that HPT was mediated by literacy motivation and abilities because it was assessed through a writing task.

The one correlational relationship found between the three constructs of empathy was for EC and HPT, $r(212) = .166$, $p = .034$. As discussed above, this relationship was almost entirely a result of the narrative group (narrative group for HPT & EC $r(68) =$

.322, $p < .05$) in comparison to the historical document group (historical document group for HPT & EC $r(83) = .071$, $p = .522$).

As highlighted in the literature review, some historians have expressed experiencing EC while researching and writing about people in the past (McCullough, 2003; Stack, 2017). Their empathic experiences appear to be both an impetus for their research and a consequence of their research (McCullough, 2003; Stack, 2017). In others words, it appears to be the love and fascination for people in the past that drives their research, and through their research they appear to achieve cognitive and emotional empathic experiences (McCullough, 2003; Stack, 2017).

The findings of this study provide empirical evidence of a positive relationship of EC and HPT for certain texts. Individuals who self-reported higher levels of EC for the narrative group, went on to write paragraphs that demonstrated higher scores with more context, perspective taking, and above all, engaged in more inferences. This shows that there may be a relationship in experiencing EC and HPT while writing about certain events through certain texts.

In addition, EC was correlated to word count for the narrative text and word count was highly correlated with HPT scores. A larger number of individuals assigned to read the narrative text and who experienced greater empathic concern wrote more words. Those who wrote more usually scored better on the HPT categories than those who wrote fewer words. Writing scores are often highly correlated with word count (Kobrin, Hui Deng, & Shaw, 2007). It is interesting to note then, that word count, for the narrative text is associated with EC.

Assessing Theory of Mind for Adolescents

The ToM instruments appear to be inadequate for adolescents. The DANVA received higher reliability scores than the Faces Test (Faces Test Cronbach's $\alpha = .178$, DANVA Cronbach's $\alpha = .586$). Furthermore, the two instruments were only weakly correlated, $r(212) = .142$, $p < .05$. The lower reliability score for the Faces Test may be a result of a decline in motivation because the Faces Test was administered on the second day. It may also be connected with the issue of having an adult actress or having only one alternative distractor. In contrast, the DANVA was administered on the first day, and it employs child actors, with three alternative distractors. The low reliability scores could be due to the fact that the Faces Test used an adult actress and that it assessed for 10 complex emotions, while the DANVA used child male and female actors and it only assessed for four basic emotions.

Previous studies have found that the use of actors of different ethnicity or cultural background can affect a person's score in discerning emotions (Adams et al., 2010). This study provides evidence it is also more challenging for adolescents to recognize facial emotions of individuals with the opposite gender. Males and females did not demonstrate a significant difference in reliability for the DANVA when both male and female actors were portrayed, but there is a significant difference in reliability for the Faces Test when only a female actress was portrayed. This provides evidence that ToM ability may be influenced by the gender of the actors.

There are other possible explanations for the low reliability scores. The Faces Test includes only one distractor, so all participants only selected between two options.

This increased the random chance of guessing correctly. The low reliability on the Faces Test could have been related to a ceiling effect. Most items were answered correctly by 95% percent of the participants. Four items were answered correctly by all respondents. Multiple items were only missed once or twice. Very little variance on several items increases the impact of the small variance that exists. It is likely that some of this small amount of variance was a single person who accidentally clicked incorrectly. If there had been much more variance, then random mistakes likely would not be as impactful.

The ceiling effect for the Faces Test and the DANVA result in instruments with more variance with lower scores, while individuals with higher abilities usually scored about the same. In other words, individuals with higher ToM abilities usually scored perfect or almost perfect while lower scores were spread out over a range of scores. In other words, the Faces Test and the DANVA were not as sensitive for students with higher ToM abilities. These issues discussed above including the low Alpha Cronbach scores, the ceiling effects, the issue of using an adult female actress, and only a single distractor all contribute to these ToM instruments being less effective for adolescents.

The Author Recognition Test

The ART has consistently correlated with ToM in prior studies (Mar et al., 2006; Panero et al. 2016). It is interesting that the relationship between the ART and ToM appears to be weaker in this study. Again, the ART was not significantly correlated with the Faces Test, $r(212) = .086, p = .215$; or the, DANVA, $r(212) = .129, p = .061$. This may be a result of the specific juvenile authors selected in this version of the ART.

Different types of literature have different degrees of correlation with the Faces Test (Kidd & Castano, 2016; Mar et al, 2006). For example, Mar et al., found that nonfiction authors are not positively correlated with ToM, while fictional authors are. In addition, juveniles may not pay as close attention to authors when reading a novel so that this instrument may be a less effective proxy for this age. The lack of a statistical correlation is also likely connected to the low reliability of the ToM instruments. Low reliability is indicative of greater statistical noise, which decreases the strength of a correlational relationship.

Research has emphasized the correlation of the ART with ToM (Mar et al., 2006; Panero et al. 2016), but perhaps just as powerful is the ART's relationship with EC. A phi test above indicates that the recognition of only three authors demonstrated a statistically significant relationship with scores on the Faces Test. In contrast, for Empathic Emotions, there were seven authors whose recognition demonstrated a p value of .05 or lower on a phi test. Furthermore, three of these authors demonstrate p values of .001 or lower (see Table 17 and

Table 18). This suggests a possible relationship between the recognition of certain authors and the disposition for experiencing empathic concern while reading. This is an important finding because EC may be more important than ToM for moral behavior among eighth-grade students. Prior research has more closely connected EC to moral behavior (Batson, 2011), while ToM has demonstrated mixed results in regard to moral behavior (Bloom, 2016). For example, or ToM, can be a tool employed by con artists, liars and bullies (Bloom, 2016), while EC appears to impel individuals to help those in need (Batson, 2011).

Computer Reading

Having students log into an online platform through laptops, in order to administer a study with more than 200 participants with three different teachers has its advantages. It is much easier and it is much more uniform than administering the study orally or through a paper survey. However, it does have potential drawbacks, such as the lack of face-to-face discussions among and between teacher and students. This study did not employ any social scaffolding. Students did not read or discuss in groups. Reading texts is a type of social activity (Allen, 2011; Fuller, 2008) and discussing texts in real time can be a powerful engagement and scaffolding tool (Fuller 2008).

Furthermore, reading on a computer screen may be less effective than reading from hard copy texts. Mangen, Walgermo, and Brønnick (2013) found that students scored lower on comprehending texts when they read on a computer screen in comparison to when they read on paper. Carr (2011) argues that readers on screens are more likely to scan and skip through the text quicker, without engaging in deep slow

thinking.

Of course, in a face-to-face classroom, it is not likely a teacher would provide historical documents without providing additional instructional support. Rather, a teacher would likely provide definitions and questions, have students read and discuss together in small groups, check for student understanding and help explain the meaning of the texts when students appear to be struggling. As a result, administering the study purely through a computer may have limited generalizability to typical classrooms.

Limitations

There are several limitations with this study in regard to a lack of a control group, the use of only two specific texts, the use of a specific adolescent population, a short-term study design, and the ceiling effect of the ToM instruments.

The researcher expected that the ToM pretest, the DANVA, would serve as a robust control for the Faces Test, because both the DANVA and the Faces Test ostensibly entail the same ability: discerning emotions from facial expressions. However, in order to be a robust control variable, the DANVA would need to be highly correlated to the Faces Test. This would have provided greater statistical power in the regression. With an accurate baseline ToM ability, one would be able to be more confident that differences between an individual's pretest and posttest scores are a result of the intervention.

If the DANVA and the Faces Test had been highly correlated as anticipated, a control group would not have been as necessary. As a result, the decision was made to not include a control group. But the DANVA and the Faces Test were only weakly correlated. Therefore, not having a control group was major limitation to the study design. It would have been helpful to have a baseline score on the Faces Test from a control group that had not read either text. It is possible that a control group would have demonstrated lower scores than both text groups, which would provide evidence for an effect on ToM for both texts. Or, possibly a control group would have demonstrated the same score as both experimental groups, indicating no effect of either text type.

This study was limited by the amount of time available in a single class period (61-62 minutes). For this reason, the texts were limited to about a 1,000 words, and most

students only had about 30 minutes remaining to write a paragraph. As a result, this study only sought to discern an immediate short-term effect of reading historical texts on empathy. However, adolescents may require longer texts in order to sufficiently activate ToM. Furthermore, the design for this study does not directly relate to potential long-term effects. That would require a much greater dosage of reading certain texts over longer periods of time.

This study was limited by the lack of an appropriate adolescent instrument. The DANVA appeared to be designed for a younger population (Nowicki & Duke, 1994) while the Faces Test appeared to be designed primarily to identify adults with Autism (Baron-Cohen et al., 1997). This demonstrates a need for an adolescent ToM instrument, with male and female adolescent actors and actresses that would be more challenging than the DANVA and Faces Test but easier than the adult version of the Reading the Mind in the Eye Test (Baron-Cohen et al., 2001). Such an instrument may experience less of a ceiling effect and higher reliability.

This study used two different types of texts, classified broadly as a narrative text, and a collection of historical documents. However, it is very difficult to generalize to entire categories of texts. Other narrative texts and historical documents or combinations thereof may be more or less effective in activating ToM. Understanding the effect of different types of texts on ToM will require many more studies employing a much greater range of texts.

These findings are specific to this particular predominantly White, upper-class, adolescent population. It will take many other studies to discern how these findings can

generalize to other segments of the population. Based on the school's state standardized tests, the students at this school tend to have much higher comprehension scores than other typical students their age. As a result, other students drawn from other populations would likely have a more difficult challenge to comprehend these texts.

Future Research

This study should be replicated with different texts, especially longer texts. It is possible that a greater treatment dosage, in this case of the texts, could increase the effect. Furthermore, a study that employs greater number of participants would also increase the power of the study.

As discussed above, individuals with higher comprehension ability, who read the historical documents for longer times also demonstrated higher ToM scores. With the design that this study employed it is impossible to prove causation, but it is possible that a deeper analysis of the historical documents produced a greater activation of ToM. If so, then perhaps it is also possible to broaden the potential benefits to individuals with lower SRI scores by providing greater support in comprehending the documents. Perhaps assignments that are tailored to the historical documents, that help students make inferences from each document and then help students to connect the documents into a cohesive narrative, could produce greater activation of ToM with the historical documents for a greater number of participants. One could assess students who are not provided the extra support and who are provided the extra support. Past studies demonstrate that students need scaffolding or some type of extra support to successfully

analyze historical documents (Foster, 2001).

In this study, the texts were used as the intervention and a writing task was used as an assessment of HPT. However, it would be interesting to measure the effect of historical writing on ToM and emotions. One could employ a study design with a writing task as part of the intervention prior to a ToM and EC assessment. Perhaps, writing activates ToM and EC more effectively than just reading the text.

For this study, the researcher used a collection of modified excerpts of historical documents that may have inhibited the construction of a coherent situation model. It would also be valuable to use a single first-person primary source, such as an extensive diary entry. Batson (2011) often used first-person accounts to activate EC participants. A longer single entry would perhaps provide more opportunity for the reader to understand the author's perspective, and become emotionally invested.

This study was purely a quantitative study, yet it did collect a large amount of qualitative data in the form of student essays. The researcher should analyze these essays looking for qualitative differences between the two groups. Furthermore, the researcher should find clues regarding how students practice historical thinking and agency in trying to understand historical events and the individuals within these events.

This study was designed to measure a short-term effect. Five other studies found short-term immediate improvements in ToM through the engagement of certain activities (Black & Barnes, 2015a, 2015b; Bormann & Greitemeyer, 2015; Kidd & Castano, 2013; Kidd et al., 2016). Another study provides evidence of an enduring effect that lasted at least week (Pino & Mazza, 2016). It would be valuable to provide more longitudinal

evidence of the long-term changes that reading can have on students. What kinds of effects will the reading of a single book or multiple books have on ToM, EC, and HPT over the span of months or years?

It is important to find ToM instruments with higher reliability. The most commonly applied ToM instrument is the Reading the Mind in the Eye's Test (RMET, Baron-Cohen et al., 2001). In a pilot study, this researcher used the RMET but calculated low reliability scores (Cronbach's $\alpha = .56$). As a result, the researcher applied the DANVA and the Faces Test in this study, but these were comparable or worse. In addition, the RMET did not demonstrate a ceiling effect. It appears that the RMET is more sensitive to a range of abilities. There are several other ways that one can assess for ToM, from the Director's Task (Dumontheil et al., 2010), to Morphed Faces (Schweinberger, Burton, & Kelly, 1999) or to false-belief scenarios (Saxe & Kanwisher, 2003). Above the researcher discussed the need to develop a ToM instrument that is better designed for an adolescent population. It would be interesting to engage some of these other instruments with adolescents in the reading of history. The application of various validated psychology instruments in the investigation of historical empathy could help strengthen the case that empathy really is connected to some of the practices in history.

Finally, this study employed a fairly homogenous upper middle class participant population. Future studies should engage participant populations with different ethnicities and levels of SES.

Final Conclusion

This study is based on the theoretical premise that transitioning adults and adolescents today have significant deficits in their empathetic dispositions and abilities (Konrath, 2011; Smith et al., 2011; Turkle, 2015; Twenge & Campbell, 2009).

Adolescents should be targeted for intervention because adolescence is a period of dramatic neurological development pertaining to empathy (Blakemore, 2008).

Furthermore, this study assumes that social and emotional learning can be accomplished in conjunction with academic learning (CASEL, 2015). Certain types of literature have demonstrated a causal effect in eliciting empathy as defined as ToM (Kidd & Castano, 2013). This study was performed in middle school classrooms to test the hypothesis that the reading of historical primary documents could more effectively activate empathy whether measured as theory of mind (ToM), empathic concern (EC), or historical perspective-taking (HPT) in comparison to a historical narrative text.

Middle school student participants were randomly assigned to read either a historical narrative text or a collection of historical primary documents. Students were then assessed on the three dependent variables: theory of mind (ToM), empathic concern (EC), or historical perspective-taking (HPT). There was not a statistical difference found for any of the dependent variables. This could mean that the instruments were not sufficiently sensitive, or that the texts activated the dependent variables equally or not at all. This study demonstrates the challenges involved in using three different teachers with 15 different classes, and using psychological instruments to determine the potential effect of texts on students' cognition and emotions.

Although this study was not successful in providing clear answers for the research questions, researchers need to gain greater understanding of the relationships between literacy and empathy for adolescents. Empathy is crucial for morality. Society is raising younger generations in a way that undermines their empathetic development. As a result, there is a crisis of empathy among adolescents. It is therefore incumbent for researchers to discover, and for educators to apply, ways of engendering greater levels of empathy for the rising generation.

REFERENCES

- Adams, R. B., Jr., Rule, N. O., Franklin, R. G., Jr., Wang, E., Stevenson, M. T., Yoshikawa, S., ... Ambady, N. (2010). Cross-cultural reading the mind in the eyes: An fMRI investigation. *Journal of Cognitive Neuroscience*, 22(1), 97-108.
- Allen, G. (2011). *Intertextuality* (2nd ed.) New York, NY: Routledge.
- American Psychological Association (2002). *Developing adolescents: A reference for professionals*. Washington, DC: Author.
- Anderson, R. C. (1984). Role of the reader's schema in comprehension, learning, and memory. In R. C. Anderson, J. Osborn, & R. J. Tierney (Eds.), *Learning to read in American schools: Basal readers and content texts* (pp. 243-257). Hillsdale, NJ: Erlbaum.
- Anderson, C. A., Bushman, B. J., Donnerstein, E., Hummer, T. A., & Warburton, W. (2015). SPSSI research summary on media violence. *Analyses of Social Issues and Public Policy (ASAP)*, 15(1), 4-19. doi:10.1111/asap.12093
- Armstrong, K. (2006). *The great transformation: The beginning of our religious traditions*. Toronto, Canada: Vintage Canada.
- Ashby, R., & Lee, P. (1987). Children's concepts of empathy and understanding in history. In C. Portal (Ed.) *The history curriculum for teachers* (pp. 62-88). Lewes, UK: The Falmer Press.
- Asher, S. R., & Taylor, A. R. (1981). Social outcomes of mainstreaming: Sociometric assessment and beyond. *Exceptional Education Quarterly*, 1(4), 13-30.
- Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, 3(3), 193-209.
- Bandura, A., Underwood, B., & Fromson, M. E. (1975). Disinhibition of aggression through diffusion of responsibility and dehumanization of victims. *Journal of Research in Personality*, 9(4), 253-269.
- Bargh, J. A., & Chartrand, T. L. (2000). The mind in the middle: A practical guide to priming and automaticity research. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 253-285). New York, NY: Cambridge University Press.

- Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The “Reading the Mind in the Eyes” Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 42(2), 241-251.
- Baron-Cohen, S., Wheelwright, S., & Jolliffe, A. T. (1997). Is there a “language of the eyes”? Evidence from normal adults, and adults with autism or Asperger syndrome. *Visual Cognition*, 4(3), 311-331.
- Barton, K. C. (2005). Primary sources in history: Breaking through the myths. *Phi Delta Kappan*, 86(10), 745-753.
- Barton, K. C. (2008). Narrative simplifications in elementary students’ historical thinking. L. S. Levstik, & K. C. Barton (Eds.) *Researching history education: Theory, method, and context* (pp. 181-208). New York: NY: Routledge.
- Barton, K. C., & Levstik, L. S. (2004). *Teaching history for the common good*. Mahwah, NJ: Erlbaum.
- Batson, C. D. (2011). *Altruism in humans*. New York, NY: Oxford University Press.
- Batson, C. D., Chang, J., & Orr, R. (2002). Empathy, attitudes, and action: Can feeling for a member of a stigmatized group motivate one to help the group? *Personality & social psychology bulletin*, 28(12), 1656-1666. <https://doi-org.dist.lib.usu.edu/10.1177/014616702237647>
- Batson, C. D., Eklund, J. H., Chermok, V. L., Hoyt, J. L., & Ortiz, B. G. (2007). An additional antecedent of empathic concern: Valuing the welfare of the person in need. *Journal of Personality & Social Psychology*, 93(1), 65-74.
- Bear, G. G., & Watkins, J. M. (2006). Developing self-discipline. In G. G. Bear & K. M. Minke (Eds.), *Children’s social and emotional learning needs III: development, prevention, and intervention* (pp. 29-44). Bethesda, MD: National Association of School Psychologists.
- Belfield, C., Bowden, A. B., Klapp, A., Levin, H., Shand, R., & Zander, S. (2015). The economic value of social and emotional learning. *Journal of Benefit-Cost Analysis*, 6(3), 508-544.
- Benson, P. L. (2006). *All kids are our kids: What communities must do to raise caring and responsible children and adolescents* (2nd ed.). San Francisco, CA: Jossey-Bass.
- Benson, P. L., Scales, P. C., Leffert, N., & Roehlkepartain, E. C. (1999). *A fragile foundation: The state of developmental assets among American youth*. Minneapolis, MN: Search Institute.

- Berrueta-Clement, J. R. (1984). *Changed lives: The effects of the Perry Preschool Program on youths through age 19: Monographs of the high/scope Educational Research Foundation, number eight*. Ypsilanti, MI: High/Scope Foundation.
- Black, J. E., & Barnes, J. L. (2015a). The effects of reading material on social and non-social cognition. *Poetics*, 52, 32-43. doi:10.1016/j.poetic.2015.07.001
- Black, J., & Barnes, J. L. (2015b). Fiction and social cognition: The effect of viewing award-winning television dramas on theory of mind. *Psychology of Aesthetics, Creativity, And The Arts*, 9(4), 423-429. doi:10.1037/aca0000031
- Blakemore, S. (2008). The social brain in adolescence. *Nature Reviews Neuroscience*, 9(4), 267-277. doi:10.1038/nrn2353
- Blakemore, S., & Choudhury, S. (2006). Development of the adolescent brain: Implications for executive function and social cognition. *Journal of Child Psychology and Psychiatry*, 47(3-4), 296-312.
- Bloom, P. (2016). *Against empathy: The case for rational compassion*. New York, NY: Harper Collinss.
- Boddington, T. (1980). Empathy and the teaching of history. *British Journal of Educational Studies*, 28(1), 13-19.
- Bormann, D., & Greitemeyer, T. (2015). Immersed in virtual worlds and minds: Effects of in-game storytelling on immersion, need satisfaction, and affective theory of mind. *Social Psychological and Personality Science*, 6(6), 646-652. doi:10.1177/1948550615578177
- Bramlett M.D., & Mosher W.D. (2002). Cohabitation, Marriage, Divorce, and Remarriage in the United States. National Center for Health Statistics. *Vital Health Stat* 23(22).
- Brooks, S. (2009). Historical empathy in the social studies classroom: A review of the literature. *Journal of Social Studies Research*, 33(2), 213-234.
- Broughton, S. H., Sinatra, G. M., & Nussbaum, E. M. (2013). "Pluto has been a planet my whole life!" Emotions, attitudes, and conceptual change in elementary students learning about Pluto's reclassification. *Research in Science Education*, 43(2), 529-550.
- Bruner, J. (1991). The narrative construction of reality. *Critical Inquiry*, 18(1), 1-21.
- Buchanan, C. M., Eccles, J. S., & Becker, J. B. (1992). Are adolescents the victims of raging hormones? Evidence for activational effects of hormones on moods and behavior at adolescence. *Psychological Bulletin*, 111(1), 62-107.

- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6(1), 3-5.
- Carey, S., Diamond, R., & Woods, B. (1980). Development of face recognition: A maturational component? *Developmental Psychology*, 16(4), 257 – 269.
- Carlyle, T. (1841/1897). *On heroes and hero-worship and the heroic in history*. London, UK: Chapman and Hall.
- Carr, N. (2011). *The shallows: What the Internet is doing to our brains*. New York, NY: Norton.
- Collaborative for Academic, Social and Emotional Learning (CASEL). (2015). *CASEL guide: Effective social and emotional learning programs*. Chicago, IL: Author.
- Catalano, R. F., Berglund, M. L., Ryan, J. A. M., Lonczak, H. S., & Hawkins, J. D. (2002). Positive youth development in the United States: Research findings on evaluations of positive youth development programs. *The Annals of the American Academy of Political and Social Science*, 591(1), 98-124.
- Christensen, L. B., Johnson, B., & Turner, L. A. (2014). *Research methods, design, and analysis* (12th ed.). Upper Saddle River, NJ: Pearson.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Earlbaum.
- Collette, J. (2012). *A primary document analysis-centered curriculum in a residential treatment center*. Retrieved from <https://digitalcommons.usu.edu/etd/>
- Common Sense Media. (2016). *Technology addiction: Concerns, controversy, and finding balance*.
- Common Core State Standards Initiative. (2010). *Common Core State Standards for English Language Arts & Literacy in History/Social Studies, Science, and Technical Subjects*. Retrieved from www.corestandards.org
- Cook, C. R., Gresham, F. M., Kern, L., Barreras, R. B., Thornton, S., & Crews, S. D. (2008). Social skills training for secondary students with emotional and/or behavioral disorders: A review and analysis of the meta-analytic literature. *Journal of Emotional and Behavioral Disorders*, 16(3), 131-144.
- Cordova, J. R., Sinatra, G. M., Jones, S. H., Taasobshirazi, G., & Lombardi, D. (2014). Confidence in prior knowledge, self-efficacy, interest and prior knowledge: Influences on conceptual change. *Contemporary Educational Psychology*, 39(2), 164-174.

- Curtin, S. C., Ventura, S. J., Martinez, G. M., (2014). Recent declines in nonmarital childbearing in the United States. *NCHS data brief, no 162*. Hyattsville, MD: National Center for Health Statistics.
- Cutting, J. (2016). Narrative theory and the dynamics of popular movies. *Psychonomic Bulletin & Review*, 23(6), 1713-1743. <https://doi-org.dist.lib.usu.edu/10.3758/s13423-016-1051-4>
- Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44(1), 113-126. doi:10.1037/0022-3514.44.1.113
- Davis, O. L. (2001). In pursuit of historical empathy. In O. L. Davis, E. A. Yeager, & S. J. Foster (Eds.), *Historical empathy and perspective taking in the social studies* (pp. 1-12). New York, NY: Rowman & Littlefield.
- Davis, O. L., Yeager, E. A., & Foster, S. J. (Eds.). (2001). *Historical empathy and perspective taking in the social studies*. New York, NY: Rowman & Littlefield.
- Day, R. R., & Park, J. S. (2005). Developing Reading Comprehension Questions. *Reading in a foreign language*, 17(1), 60-73.
- Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews*, 3(2), 71-100.
- Decety, J., & Jackson, P. L. (2006). A social-neuroscience perspective on empathy. *Current Directions in Psychological Science*, 15(2), 54-58.
- Djikic, M., Oatley, K., & Moldoveanu, M. C. (2013). Reading other minds: Effects of literature on empathy. *Scientific Study of Literature*, 3(1), 28-47.
- Donahue, M. L., & Press, G. (2014). Perspective-taking and reading comprehension of narratives: Lessons learned from "The Bean." In C.A. Stone, E. R., Silliman, B. J. Ehren, & G. P. Wallach (Eds.), *Handbook of language and literacy: Development and disorders* (pp. 323-338). New York, NY: Guilford.
- Doppen, F. H. (2000). Teaching and learning multiple perspectives: The atomic bomb. *The Social Studies*, 91(4), 159-169.
- Douglas, F. (1845). *The narrative of the life of Frederick Douglas an American Slave*. Boston, MA: Anti-Slavery Office.
- Dumontheil, I., Apperly, I. A., & Blakemore, S. J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science*, 13(2), 331-338.

- Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development, 82*(1), 405-432.
- Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., & Mac Iver, D. (1993). Development during adolescence: The impact of stage-environment fit on young adolescents' experiences in schools and in families. *American Psychologist, 48*(2), 90-101.
- Eisenberg, N. (2000). Emotion, regulation, and moral development. *Annual Reviews in Psychology, 51*, 665-697.
- Eisenberg, N. (2003). Prosocial behavior, empathy, and sympathy. In M. H. Bornstein, L. Davidson, C. L. M. Keyes, & K. A. Moore (Eds.), *Well-being: Positive development across the life course* (pp. 253-265). Mahwah, NJ: Erlbaum.
- Eisenberg, N., Carlo, G., Murphy, B., & Van Court, P. (1995). Prosocial development in late adolescence: A longitudinal study. *Child Development, 66*(4), 1179-1197.
- Eisenberg, N., & Valiente, C. (2002). Parenting and children's prosocial and moral development. In M. H. Bornstein (Ed.), *Handbook of parenting* (pp. 111-142). Mahwah, NJ: Erlbaum.
- Endacott, J. L. (2010). Reconsidering affective engagement in historical empathy. *Theory & Research in Social Education, 38*(1), 6-47.
- Endacott, J. L. (2014). Negotiating the process of historical empathy. *Theory & Research in Social Education, 42*(1), 4-34.
- Endacott, J., & Brooks, S. (2013). An updated theoretical and practical model for promoting historical empathy. *Social Studies Research and Practice, 8*(1), 41-58.
- Endacott, J., & Brooks, S. (2018). Historical empathy. In S. Metzger & L. Harris (Eds.) *The Wiley international handbook of history teaching and learning* (pp. 203-226). Hoboken, NJ: Wiley. doi:10.1002/9781119100812
- Endacott, J. L., & Pelekanos, C. (2015). Slaves, women, and war! Engaging middle school students in historical empathy for enduring understanding. *The Social Studies, 106*(1), 1-7.
- Erickson, F. (2004). *Talk and social theory: Ecologies of speaking and listening in everyday Life*. Malden, MA: Polity Press.
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*(2), 211-245.

- Feldman, S. S., & Elliott, G. R. (Eds.). (1990). *At the threshold: The developing adolescent*. Cambridge, MA: Harvard University Press.
- Flavell, J. H. (1977). *Cognitive development*. Englewood Cliffs, NJ: Prentice Hall.
- Fliessbach, K., Weis, S., Klaver, P., Elger, C. E., & Weber, B. (2006). The effect of word concreteness on recognition memory. *Neuroimage*, 32(3), 1413-1421.
- Fong, K., Mullin, J. B., & Mar, R. A. (2013). What you read matters: The role of fiction genre in predicting interpersonal sensitivity. *Psychology of Aesthetics, Creativity, and the Arts*, 7(4), 370-376.
- Foster, S. J. (2001). Historical empathy in theory and practice: Some final thoughts. In O. L. Davis, E. A. Yeager, & S. J. Foster (Eds.), *Historical empathy and perspective taking in the social studies* (pp. 167-182). New York, NY: Rowman & Littlefield.
- Freytag, G. (1900). *Freytag's technique of the drama, an exposition of dramatic composition and art* (3rd ed.). Chicago, IL: Scott, Foresman and Company.
- Frith, C., & Frith, U. (2005). Theory of mind. *Current Biology*, 15(17), R644-R645.
- Fuller, D. (2008). Reading as social practice: The Beyond the Book research project. *Popular Narrative Media*, 1(2), 211-217.
- Gehlbach, H. (2004). *A new perspective on perspective taking: A study of social perspective taking and its association with conflict resolution, historical empathy, and social studies achievement*. Dissertation Abstracts International Section A: Humanities and Social Sciences. ProQuest Information & Learning. Retrieved from: <http://dist.lib.usu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2004-99019-023&site=ehost-live>
- Giddens, A. (1984). *The constitution of society*. Berkeley, CA: University of California Press.
- Gildemeister, C., (2016). *Protecting children or protecting Hollywood? A twenty-year examination of the effectiveness of the TV content ratings system*. Los Angeles, CA: Parent Television Council.
- Gioia, D. (2008). *To read or not to read: A question of national consequence*. Washington, DC: National Endowment for the Arts.
- Goleman, D. (2005). *Emotional intelligence*. New York, NY: Bantam.
- Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods*, 36(2), 193-202.

- Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, 40(5), 223-234.
- Greene, J. P., Kisida, B., & Bowen, D. H. (2014). The Educational value of field trips. *Education Next*, 14(1), 78-86. Retrieved from <http://dist.lib.usu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1027941&site=ehost-live>
- Greenberg, M. T., Domitrovich, C., & Bumbarger, B. (2001). The prevention of mental disorders in school- aged children: Current state of the field. *Prevention & Treatment*, 4, 1-62.
- Hartup, W. W. (1983). The peer system. In P. H. Mussen (Ed.). *Handbook of child psychology* (Vol. 4, pp. 103-196). Hoboken, NJ: Wiley.
- Hatzenbuehler, M. L., Corbin, W. R., & Fromme, K. (2008). Trajectories and determinants of alcohol use among LGB young adults and their heterosexual peers: results from a prospective study. *Developmental Psychology*, 44(1), 81.
- Hoffman, M. L. (1985). Interaction of affect and cognition in empathy. In C. E. Izard, J. Kagan, & R. B. Zajonc (Eds.), *Emotions, cognition, and behavior* (pp. 103-131). New York, NY: Cambridge University Press.
- Hoffman, M. L. (2000). *Empathy and moral development: Implications for caring and justice*. New York, NY: Cambridge University Press.
- Hoffman, M. L. (2008). Empathy and prosocial behavior. *Handbook of Emotions*, 3, 440-455.
- Horowitz, J. L., & Garber, J. (2007). The prevention of depressive symptoms in children and adolescents: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, 74, 401-415.
- Huijgen, T., van de Grift, W., Van Boxtel, C., & Holthuis, P. (2017). Teaching historical contextualization: the construction of a reliable observation instrument. *European Journal of Psychology of Education*, 32(2), 159-181.
- Jenkins, K. (1991). *Re-thinking history*. New York, NY: Routledge.
- Keating, D. P. (1990). Adolescent thinking. In S. S. Feldman, G. R. Elliott, S. S. Feldman, & G. R. Elliott (Eds.), *At the threshold: The developing adolescent* (pp. 54-89). Cambridge, MA: Harvard University Press.

- Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, 11(1), 32-38.
- Kidd, D. C., & Castano, E. (2013). Reading literary fiction improves theory of mind. *Science*, 342, 377-380. doi:10.1126/science.1239918
- Kidd, D. & Castano, E. (2017). Different stories: how levels of familiarity with literary and genre fiction relate to mentalizing. *Psychology of Aesthetics, Creativity & the Arts*, 11(4), 474-486.
- Kidd, D. C., & Castano, E. (2017). Panero et al. (2016): Failure to replicate methods caused the failure to replicate results. *Journal of Personality and Social Psychology*, 112(3), e1-e4. doi:10.1037/pspa0000072
- Kidd, D., Ongis, M., & Castano, E. (2016). On literary fiction and its effects on theory of mind. *Scientific Study of Literature*, 6(1), 42-58.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York, NY: Cambridge University Press.
- Kobrin, J. L., Hui Deng, & Shaw, E. J. (2007). Does quantity equal quality? The relationship between length of response and scores on the SAT essay. *Journal of Applied Testing Technology*, 8(1), 1-15.
- Kohlberg, L. (1971). Stages of moral development. *Moral Education*, 1, 23-92.
- Kohlmeier, J. (2006). "Couldn't she just leave?": The relationship between consistently using class discussions and the development of historical empathy in a 9th grade world history course. *Theory & Research in Social Education*, 34(1), 34-57.
- Konrath, S. H., O'Brien, E. H., & Hsing, C. (2011). Changes in dispositional empathy in American college students over time: A meta-analysis. *Personality and Social Psychology Review*, 15(2), 180-198.
- Lamm, C., Batson, C. D., & Decety, J. (2007). The neural substrate of human empathy: Effects of perspective-taking and cognitive appraisal. *Journal of Cognitive Neuroscience*, 19(1), 42-58.
- Lee, H. (1960). *To kill a mockingbird*. New York, NY: Warner Books.
- Lee, P., & Ashby, R. (2001). Empathy, perspective taking, and rational understanding. In O. L. Davis, E. A. Yeager, & S. J. Foster (Eds.) *Historical empathy and perspective taking in the social studies* (pp. 21-50). Lanham, MD: Rowan and Littlefield.

- Levi, P. (1986). *The drowned and the saved*. New York, NY: Simon & Schuster.
- Levine, P., & Kawashima-Ginsberg, K. (2017). *The republic is (still) at risk-and civics is part of the solution*. Medford, NY: Tufts University.
- Lowenthal, D. (1985) *The past is a foreign country*. Cambridge, MA: Cambridge University Press
- Mangen, A., Walgermo, B. R., & Brønnick, K. (2013). Reading linear texts on paper versus computer screen: Effects on reading comprehension. *International Journal of Educational Research*, 58, 61-68.
- Mar, R., & Oatley, K. (2008). The function of fiction is the abstraction and simulation of social experience. *Perspectives on Psychological Science*, 3, 173-192.
- Mar, R., Oatley, K., Hirsh, J., dela Paz, J., & Peterson, J. (2006). Bookworms versus nerds: Exposure to fiction versus non-fiction, divergent associations with social ability, and the simulation of fictional social worlds. *Journal of Research in Personality*, 40, 694-712.
- Mar, R. A., Oatley, K., & Peterson, J. B. (2009). Exploring the link between reading fiction and empathy: Ruling out individual differences and examining outcomes. *Communications*, 34(4), 407-428.
- Marx, K., Engels, F. (1867/2007). *Capital: A critique of political economy*. New York, NY: Cosimo.
- McCullough, D. (2001). *John Adams*. New York, NY: Simon and Schuster.
- McCullough, D. (2003). David McCullough lecture. In *Humanities: The Magazine of the National Endowment for the Arts*. Retrieved from <https://www.neh.gov/about/awards/jefferson-lecture/david-mccullough-lecture>
- McLanahan, S., & Percheski, C. (2008). Family structure and the reproduction of inequalities. *Annual Review of Sociology*, 34, 257–276.
- Merrell, K. W., & Gimpel, G. (2014). *Social skills of children and adolescents: Conceptualization, assessment, treatment*. New York, NY: Psychology Press.
- Meier, D. (2010). *The use of primary source historical documents, historical reasoning heuristics, and the subsequent development of historical empathy* [e-book]. Ipswich, MA: ProQuest Information & Learning.
- Müller, C. M., & Gmünder, L. (2014). An evaluation of the “Reading the Mind in the Eyes-Test” with seventh to ninth graders. *Journal of Mental Health Research in Intellectual Disabilities*, 7(1), 34-44.

- Neumann, D. J. (2010). "What is the text doing?" Preparing pre-service teachers to teach primary Sources effectively. *The History Teacher*, 43(4), 489-511.
- Nokes, J. D. (2013). *Building students' historical literacies: Learning to read and reason with historical texts and evidence*. New York, NY: Routledge.
- Nokes, J. D. (2014). Elementary students' roles and epistemic stances during document-based history lessons. *Theory and Research in Social Education*, 42(3), 375-413.
- Nokes, J. & De La Paz, S. (2018). Writing and argumentation in history education. In S. Metzger & L. Harris (Eds.) *The Wiley international handbook of history teaching and learning* (pp. 551-578). Hoboken, NJ: Wiley. doi:10.1002/9781119100812
- Nowicki, S., & Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: The diagnostic analysis of nonverbal accuracy scale. *Journal of Nonverbal Behavior*, 18(1), 9-35.
- Nussbaum, M. (1995). *Poetic justice: The literary imagination and public life*. Boston, MA: Beacon.
- Nussbaum, M. (1997). *Cultivating humanity: A classic defense of reform in the liberal tradition*. Cambridge, MA: Harvard University Press.
- Nussbaum, M. (2004). Liberal education and global community. *Liberal Education*, 90(1), 42-47.
- O'Brien, E., Konrath, S. H., Grühn, D., & Hagen, A. L. (2013). Empathic concern and perspective taking: Linear and quadratic effects of age across the adult life span. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 68(2), 168-175.
- Olderbak, S., Wilhelm, O., Olaru, G., Geiger, M., Brenneman, M. W., & Roberts, R. D. (2015). A psychometric analysis of the reading the mind in the eyes test: Toward a brief form for research and applied settings. *Frontiers in Psychology*, 6. Retrieved from <http://dist.lib.usu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=psych&AN=2016-22617-001&site=ehost-live>
- Panero, M. E., Weisberg, D. S., Black, J., Goldstein, T. R., Barnes, J. L., Brownell, H., & Winner, E. (2016). Does reading a single passage of literary fiction really improve theory of mind? An attempt at replication. *Journal of personality and Social Psychology*, 111(5), e46-e54.

- Panero, M. E., Weisberg, D. S., Black, J., Goldstein, T. R., Barnes, J. L., Brownell, H., & Winner, E. (2017). No support for the claim that literary fiction uniquely and immediately improves theory of mind: A reply to Kidd and Castano's commentary on Panero et al. (2016). *Journal of Personality and Social Psychology*, 112(3), e5-e8.
- Parent Television Council. (2013). *Media violence: An examination of violence, graphic violence, and gun violence in the media*. Los Angeles, CA: Author.
- Parker, W. C. (2011). Constructing public schooling today: Derision, multiculturalism, nationalism. *Educational Theory*, 61(4), 413-432.
- Paxton, R. J. (1999). A Deafening Silence: History Textbooks and the Students Who Read Them. *Review of Educational Research*, 69(3), 315-39.
- Payton, J., Weissberg, R. P., Durlak, J. A., Dymnicki, A. B., Taylor, R. D., Schellinger, K. B., & Pachan, M. (2008). *The positive impact of social and emotional learning for kindergarten to eighth-grade students: Findings from three scientific reviews*. Chicago, IL: Collaborative for Academic, Social, and Emotional Learning.
- Pearson, P. D. (2014, October 15). *Dr. P. David Pearson: Implementing the CCSS: Comprehension, Close Reading, and the Common Core*. Retrieved from: <https://www.youtube.com/watch?v=sQjH2Sh0N08&t=512s>
- Pekrun, R., Goetz, T., & Perry, R. P. (2005). *Achievement Emotions Questionnaire (AEQ). User's Manual*. (Available from R. Pekrun, Department of Psychology, University of Munich, Leopoldstrasse 13, D-80802 Muenchen).
- Pew Research Center. (2015). *Parenting in America: Outlook, worries, aspirations are strongly linked to financial situation*. Washington, DC.: The Pew Charitable Trusts.
- Pino, M. C., & Mazza, M. (2016). The use of "literary fiction" to promote mentalizing ability. *Plos One*, 11(8), 1-14. doi:10.1371/journal.pone.0160254
- Pinker, S. (2011). *The better angels of our nature: Why violence has declined*. New York, NY: Penguin.
- Popper, K. (1959). *The logic of scientific discovery*. New York, NY: Routledge.
- Preston, S. D., & de Waal, F. B. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences*, 25(1), 1-20.
- Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. New York, NY: Simon & Schuster. doi:10.1145/358916.361990

- Putnam, R. D. (2015). *Our kids: The American dream in crisis*. New York, NY: Simon & Schuster.
- Putnam, R. D., Campbell, D. E., & Garrett, S. R. (2010). *American grace: How religion divides and unites us*. New York, NY: Simon & Schuster.
- Rand, A. (1964). *The virtue of selfishness*. New York, NY: Signet.
- Ravitch, D. (2003). *The language police: How pressure groups restrict what students learn*. New York, NY: Vintage Books.
- Rideout, V. (2015). *The common sense census: Media use by tweens and teens*. San Francisco, CA: Common Sense Media.
- Rideout, V. J., Foehr, U. G., & Roberts, D. F. (2010). *Generation M2: Media in the lives of 8- to 18-year-olds*. Washington, D.C.: Henry J. Kaiser Family Foundation.
- Roberts, R. (2014). *How Adam Smith can change your life: An unexpected guide to human nature and happiness*. New York, NY: Portfolio/Penguin.
- Rockmore, E. B. (2015, October 21). How Texas teaches history. *New York Times*.
- Rosenberg, E. L. (1998). Levels of analysis and the organization of affect. *Review of General Psychology*, 2, 247-270.
- Rowe, A. D., Bullock, P. R., Polkey, C. E., & Morris, R. G. (2001). Theory of mind impairments and their relationship to executive functioning following frontal lobe excisions. *Brain*, 124(3), 600-616.
- Russell, B., & Prucell, J. (2009). *Online research essentials: Implementing and designing research studies*. San Francisco, CA: Jossey-Bass.
- Sacks, J. (2015). *Not in God's name: Confronting religious violence*. New York, NY: Schocken.
- Samur, D., Tops, M., & Koole, S. L. (2018). Does a single session of reading literary fiction prime enhanced mentalising performance? Four replication experiments of Kidd and Castano (2013). *Cognition and Emotion*, 32(1), 130-144. doi:10.1080/02699931.2017.1279591
- Sargent, J., Dopkins, S., Philbeck, J., & Chichka, D. (2010). Chunking in spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(3), 576-589. doi:10.1037/a0017528

- Savitz-Romer, M., & Bouffard, S. M. (2012). *Ready, willing, and able: A developmental approach to college access and success*. Cambridge, MA: Harvard Education Press.
- Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in “theory of mind.” *Neuroimage*, 19(4), 1835-1842.
- Scholastic. (2014). *SRI college & career technical guide*. New York, NY: Author.
- Schweinberger, S. R., Burton, A. M., & Kelly, S. W. (1999). Asymmetric dependencies in perceiving identity and emotion: Experiments with morphed faces. *Perception & Psychophysics*, 61(6), 1102-1115.
- Seixas, P., & Peck, C. (2004). Teaching historical thinking. In A. Sears & I. Wright (Eds.), *Challenges and prospects for Canadian social studies* (pp. 109-117). Vancouver, British Columbia, Canada: Pacific Educational Press.
- Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: rethinking content-area literacy. *Harvard Educational Review*, 78(1), 40-59.
- Skinner, B. F. (1938) *The behavior of organisms*. New York, NY: D. Appleton-Century Company.
- Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly*, 24(4), 402-433.
- Slovic, P., Västfjäll, D., Erlandsson, A., & Gregory, R. (2017). Iconic photographs and the ebb and flow of empathic response to humanitarian disasters. *Proceedings of the National Academy of Sciences*, 114(4), 640-644.
- Smith, A. (1759/1817). *A theory of moral sentiments*. Boston, MA: Wells and Lilly.
- Smith, C., & Snell, P. (2009). *Souls in transition: The religious and spiritual lives of emerging adults*. New York, NY: Oxford University Press.
- Smith, C., Christoffersen, K. M., Davidson, H., & Herzog, P. S. (2011). *Lost in transition: The dark side of emerging adulthood*. New York, NY: Oxford Press.
- Snow, J. (2011). *The complete research suite: A step-by-step guide to using Qualtrics*. Retrieved from: <https://www.ndsu.edu/gdc/wp-content/pdf/qualtrics-step-by-step-manual.pdf>
- Spencer, H. (1897). *The principles of ethics in two volumes*. New York, NY: Appleton.
- Stack, P. F. (2017, January 9). New book explores an early Mormon dichotomy: Women defending polygamy while pushing a feminist agenda. *The Salt Lake Tribune*.

- Stahl, S. A., & Shanahan, C. H. (2004). Learning to think like a historian: Disciplinary knowledge through critical analysis of multiple documents. In T. L. Jetton & J.A. Dole (Eds.), *Adolescent literacy research and practice* (pp. 94-115). New York, NY: Guilford.
- Stanovich, K. (2000). *Progress in reading: Scientific foundations and new frontiers*. New York, NY: Guilford Press.
- Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading research quarterly*, 24, 402-433. <https://doi-org.dist.lib.usu.edu/10.2307/747605>
- Stenner, A. J., Burdick, H., Sanford, E. E., & Burdick, D. S. (2006). How accurate are lexile text measures? *Journal of Applied Measurement*, 7(3), 307-322.
- Thompson K. (1999). *Storytelling in the new Hollywood*. Cambridge, MA: Harvard Press.
- Tobler, N. S., Roona, M. R., Ochshorn, P., Marshall, D. G., Streke, A. V., & Stackpole, K. M. (2000). School-based adolescent drug prevention programs: 1998 meta-analysis. *Journal of Primary Prevention*, 30, 275-336.
- Turkle, S. (2011). *Alone together: Why we expect more from technology and less from each other*. New York, NY: Basic.
- Turkle, S. (2015). *Reclaiming conversations: the power of talk in a digital age*. New York, NY: Penguin.
- Twenge, J. M. (2006). *Generation Me: Why today's young Americans are more confident, assertive, entitled-and more miserable than ever before*. New York, NY: Free Press.
- Twenge, J. M., & Campbell, W. K. (2009). *The narcissism epidemic: Living in the age of entitlement*. New York, NY: Simon and Schuster.
- Twenge, J. M., & Kasser, T. (2013). Generational changes in materialism and work centrality, 1976-2007: Associations with temporal changes in societal insecurity and materialistic role modeling. *Personality and Social Psychology Bulletin*, 39(7), 883-897.
- Twenge, J. M., Konrath, S., Foster, J. D., Campbell, W. K., & Bushman, B. J. (2008). Egos inflating over time: A cross-temporal meta-analysis of the Narcissistic Personality Inventory. *Journal of personality*, 76(4), 875-902.
- Utah State Board of Education. (2016). *Utah Core Standards: 7-12 Social Studies*. Salt Lake City, UT: Author. Retrieved from www.schools.utah.gov

- Van Boven, L., & Loewenstein, G. (2003). Social projection of transient drive states. *Personality and Social Psychology Bulletin*, 29(9), 1159-1168.
- VanSledright, B. A. (2001). From empathic regard to self-understanding: Im/positionality, empathy, and historical contextualization. In O. L. Davis, E. A. Yeager, & S. J. Foster, (Eds.), *Historical empathy and perspective taking in the social studies* (pp. 51-68). New York, NY: Rowman & Littlefield.
- VanSledright, B. A. (2004). What does it mean to think historically...and how do you teach it? *Social Education*, 68(3), 230-234.
- Walker, R. W., Turley, R. E., & Leonard, G. M. (2008). *Massacre at mountain meadows*. New York, NY: Oxford University Press.
- Watson, P. (2005). *Ideas: A history of thought and invention, from fire to Freud*. New York, NY: Harper Perennial.
- Watson, P. J., Biderman, M. D., & Sawrie, S. M. (1994). Empathy, sex role orientation, and narcissism. *Sex Roles*, 30(9-10), 701-723.
- Watson, P. J., & Morris, R. J. (1991). Narcissism, empathy and social desirability. *Personality and Individual Differences*, 12(6), 575-579.
- West, S. G., Finch, J. F., Curran, P., J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications* (pp. 56-75). Newbury Park, CA: Sage.
- Wilcox, B. W., Cherlin, A. J., Uecker, J. E., & Messel, M. (2012). No money, no honey, no church: The deinstitutionalization of religious life among the white working class. In L. Keister, J. McCarthy, & R. Finke (Eds.), *Religion, work and inequality* (pp. 227-250). Bingley, West Yorkshire, England: Emerald Group.
- Wilson, S. J., & Lipsey, M. W. (2007). School-based interventions for aggressive and disruptive behavior: Update of a meta-analysis. *American Journal of Preventive Medicine*, 33(Suppl. 2S), 130-143.
- Wineburg, S. S. (1991). On the reading of historical texts: Notes on the breach between school and academy. *American Educational Research Journal*, 28(3), 495-519.
- Wineburg, S. (1998). Reading Abraham Lincoln: An expert/expert study in the interpretation of historical texts. *Cognitive Science*, 22(3), 319-346. [https://doi-org.dist.lib.usu.edu/10.1016/S0364-0213\(99\)80043-3](https://doi-org.dist.lib.usu.edu/10.1016/S0364-0213(99)80043-3)
- Wineburg, S. (2001). *Historical thinking and other unnatural acts: Charting the future of teaching the past*. Philadelphia, PA: Temple University Press.

- Wineburg, S. (2010). Historical thinking and other unnatural acts. *Phi Delta Kappan*, 92(4), 81-94.
- Wineburg, S. (2013). Undue certainty: where Howard Zinn's a people's history falls short. *American Educator*, 36(4), 27-34.
- Wineburg, S., Martin, D., & Monte-Sano, C. (2013). *Reading like a historian: Teaching literacy in middle and high school history classrooms*. New York, NY: Teachers College Press.
- Wolters, C. A. (2004). Advancing achievement goal theory: Using goal structures and goal orientations to predict students' motivation, cognition, and achievement. *Journal of Educational Psychology*, 96(2), 236-250.
- Wolf, W. (2008). Chance in fiction as a privileged index of worldviews: A contribution of the world modeling function of narrative fiction. In J. Pier & J. A. G. Landa (Eds.), *Theorizing narrativity* (pp. 165-210). New York, NY: Walter de Greytur.
- Yeager, E. A., & Foster, S. J. (2001). The role of empathy in the development of historical understanding. In O. L. Davis, E. A. Yeager, & S. J. Foster (Eds.), *Historical empathy and perspective taking in the social studies* (pp. 13-20). New York, NY: Rowman & Littlefield.
- Young, A. (2012). Empathy, evolution, and human nature. In J. Decety (Ed.), *Empathy: From bench to bedside* (pp. 21-38). Cambridge, MA: MIT Press.
- Zins, J. E., Weissberg, R. P., Wang, M. C., & Walberg, H. J. (Eds.). (2004). *Building academic success on social and emotional learning: What does the research say?* New York, NY: Teachers College Press.
- Zahavi, D., & Overgaard, S. (2012). Empathy without isomorphism: A phenomenological account. In J. Decety (Ed.), *Empathy: From bench to bedside* (pp. 3-20). Cambridge, MA: MIT Press.
- Zumbusch, A. (2009). *The true story of the Salem witch hunts*. New York, NY: PowerKids Press.
- Zunshine, L. (2006). *Why we read fiction: Theory of mind and the novel*. Columbus, OH: Ohio State University Press.
- Zunshine, L. (2012). Sociocognitive complexity. *Novel: A Forum on Fiction*, 45(1), 13-18.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123(2), 162-85.

APPENDICES

Appendix A

IRB Approval

From: noreply@usu.edu
 Subject: Approval letter from USU IRB
 Date: May 15, 2018 at 10:42 AM
 To: suzanne.jones@usu.edu, jaredpcollette@gmail.com



Institutional Review Board

USU Assurance: FWA#00003308

Exemption #2

Certificate of Exemption



FROM:

Melanie Domenech Rodriguez, IRB Chair

Nicole Vouvalis, IRB Administrator

To: Suzanne Jones, Jared Collette
 Date: May 15, 2018
 Protocol #: 9336
 Title: Hope And Teacher Retention Study 2 Interviews

The Institutional Review Board has determined that the above-referenced study is exempt from review under federal guidelines 45 CFR Part 46.101(b) category #2:

Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior, unless: (a) information obtained is recorded in such a manner that human subjects can be identified, directly or through the identifiers linked to the subjects: and (b) any disclosure of human subjects' responses outside the research could reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, or reputation.

This exemption is valid for three years from the date of this correspondence, after which the study will be closed. If the research will extend beyond three years, it is your responsibility as the Principal Investigator to notify the IRB before the study's expiration date and submit a new application to continue the research. Research activities that continue beyond the expiration date without new certification of exempt status will be in violation of those federal guidelines which permit the exempt status.

As part of the IRB's quality assurance procedures, this research may be randomly selected for continuing review during the three year period of exemption. If so, you will receive a request for completion of a Protocol Status Report during the month of the anniversary date of this certification.

In all cases, it is your responsibility to notify the IRB prior to making any changes to the study by submitting an Amendment/Modification request. This will document whether or not the study still meets the requirements for exempt status under federal regulations.

Upon receipt of this memo, you may begin your research. If you have questions, please call the IRB office at (435) 797-1821 or email to irb@usu.edu.

The IRB wishes you success with your research.

4460 Old Main Hill

Logan, UT 84322-4460

PH: (435) 797-1821

Fax: (435) 797-3769

WEB: irb.usu.edu

EMAIL: irb@usu.edu

Appendix B
School District Approval



Research & Assessment

Hal L. Sanderson, Ph.D., Director

9150 S. 500 West Sandy, UT 84070

T: 801-826-5090 F: 801-826-5065 www.canyonsdistrict.org

January 5, 2018

Jared Collette
Draper Park Middle School
Canyons School District

Dear Jared,

The Canyons School District Research Review Committee has reviewed your request to conduct research titled, "Empathy in the Middle School History Classroom: The Effect of Reading Different Historical Texts on Theory of Mind, Empathic Concern, and Historical Perspective Taking". The committee **has approved** your research for the next 18 months' contingent upon receiving the following documents:

1.

1. USU IRB Approval Form

2. Informed Consent Forms

2.

These documents need to be sent to me before you begin the research.

2.

Let me know if you have any follow-up questions.

Sincerely,

Hal L. Sanderson, Ph. D.
Director of Research & Assessment
Canyons School District

Appendix C

The Introductory Text (Modified) From Reading Like a Historian

Salem Introduction

The Salem witchcraft crisis began during the winter of 1691-1692, in Salem Village, Massachusetts, when a nine and a twelve-year-old girl fell strangely ill. The girls complained of pinching, prickling sensations, knifelike pains, and the feeling of being choked. In the weeks that followed, three more girls showed similar symptoms. Reverend Parris and several doctors began to suspect that witchcraft was responsible for the girls' behavior. They pressed the girls to name the witches who were tormenting them. The girls named three women, who were then arrested. The third accused was Parris's Indian slave, Tituba. Under examination, Tituba confessed to being a witch, and testified that four women and a man were causing the girls' illness. The girls continued to accuse people of witchcraft, including some respectable church members. The new accused witches joined Tituba and the other two women in jail. The accused faced a difficult situation. If they confessed to witchcraft, they could escape death but would have to provide details of their crimes and the names of other participants. On the other hand, it was very difficult to prove one's innocence.

Witchcraft was a cultural tradition that Puritans inherited. In the centuries before this, tens of thousands of women had been tortured and executed as witches in Europe.

The Puritans believed that witches were people that made a covenant or pact with the Devil where he gave them special magical powers in exchange for their soul. They made this covenant by signing their name in the Devil's book. The Puritans believed that when a witch was causing harm to a victim, only the victims were able to see the spirit of the witch that was harming them.

The afflicted girls and women were often kept in the courtroom as evidence while the accused were examined. If they screamed and claimed that the accused witch was torturing them. Judges believed the victims even though the accused witch was not doing anything. Between June and October, twenty people were convicted of witchcraft and killed and more than a hundred suspected witches remained in jail.

Questions:

4. When did the witchcraft crisis begin?
5. Where did the witchcraft crisis begin?
6. What event made it start?
7. According to the Puritans, what was a witch?
8. Why was it really difficult to prove one's innocence?

Appendix D
Historical Documents

Salem Documents

“Discourse on Witchcraft” (Modified)

Cotton Mather, an influential preacher, gave this sermon in Boston 3 years before the Salem Witch Crisis

I will prove that Witchcraft exists. Those who deny it exists argue that they never saw any witches, therefore, there are none. That would be as if you or I said: We never met any robbers, therefore there are none. The scripture mentions witchcraft. Secondly, many people have experienced the horrors of witchcraft.

Reverend Deodat Lawson (modified)

Description: This is a journal account of what a visiting minister witnessed when he visited Reverend Parris on March 19, 1692. This is the very beginning of the witchcraft crisis. Women, especially girls were typically fairly powerless in Puritan society.

“In the beginning of the evening I went to give Mr. Parris a visit. When I was there, his niece Abigail Williams, (about 12 years of age), had a grievous fit; she was at first hurried with violence to and fro in the room sometimes making as if she would fly, stretching up her arms as high as she could. Later she said she saw Goodwin Nurse and said Do you not see her? Why there she stands! And said, Nurse offered her the book, but she was resolved she would not take it, saying often, “I won’t, I won’t, I won’t take it, I do not know what book it is; I am sure it is none of God’s book; it is the Devil’s book for ought I know.” After that, she ran to the fire, and began to throw burning wood about the house, and tried to run against the back of the chimney. They told me she attempted to go into the fire in other fits.

Court Testimony of Abigail Hobbs (Modified)

Below is the testimony of a teenager accused of witchcraft, Abigail Hobbs, on April 19, 1692.

Judge: Abigail Hobbs, you are brought before authority to answer to various acts of witchcraft. What say you? Are you guilty, or not? Speak the truth.

Abigail Hobbs: I will speak the truth. I have seen sights and been scared. I have been very wicked. I hope I shall be better, if God will help me.

Judge: What sights did you see?

Abigail Hobbs: I have seen the Devil.

Judge: How often, many times?

Abigail Hobbs: But once.

Judge: What would he have you do?

Abigail Hobbs: Why, he would have me be a witch.

Judge: Would he have you make a covenant with him?

Abigail Hobbs: Yes.

The Trial of Rebecca Nurse MARCH 24, 1692 (Modified)

Rebecca Nurse is on trial. Just like most Puritans, Nurse also believes in witchcraft, but denies that she is a witch.

Judge: (Questioning one of the accusers) have you seen this Woman hurt you?

Accuser: Yes, her spirit beat me this morning

Ann Putman in a grievous fit cried out that Rebecca Nurse hurt her.

(Others accuse Nurse of hurting her)

A man testified that when Nurse came into the house he was seized twice with an amazed condition.

Mrs. Ann Putnam: Did you not bring the Devil with you? Did you not tell me to tempt God and die? How oft have you eat and drunk your own damnation? Spread out her hands!

(They spread out her hands and the afflicted cry out in pain)

Mrs. Ann Putnam: Do you not see what an awful situation these girls are in? When your hands are loose they are afflicted. Here are these two grown persons now accuse you, what say you? Do not you see these afflicted persons, and hear them accuse you?

Nurse: The Lord knows I have not hurt them: I am an innocent person

Judge: They accuse you of hurting them, and if you think it is not unwillingly but by design, you must look upon them as murderers.

Nurse: I cannot tell what to think of it.

Petition of Mary Easty to Governor and Judge William Phipps (Modified)

Mary Easty was convicted of witchcraft and was waiting for her execution when she wrote this letter to the judge. A petition is a special request.

I petition your honor not for my own life, for I know I must die, but if it be possible that no more innocent blood may be shed. In the way and course you go more innocent bloodshed undoubtedly cannot be avoided. But I plead that you will carefully examine the afflicted persons and keep them apart for a while. I also plead with you to further investigate those who have confessed to being witches. Several have falsely confessed. I know and the Lord knows as will shortly be revealed (at the judgment day) that I have been falsely accused. I am certain that others are also innocent. I beg your honor not to deny this, my humble petition, from a poor dying innocent person. The Lord will bless you for your effort.

Petition of John Proctor

SALEM-PRISON, July 23, 1692. Proctor wrote this letter in prison begging for help from ministers of other towns

Reverend Gentlemen.

Even though we are innocent, our accusers, judges, jury, the church leaders, and all the people in general, are so full of hatred nothing will satisfy them but our innocent blood. Recently five Persons have confessed themselves to be Witches, and do accuse some of

us, of being witches along with them. Two of them would not confess anything till the authorities tied their neck to their heels till the blood was ready to come out of their noses, and this is the only reason for their confession. When they questioned my son William Proctor, he would not confess that he was guilty because he was innocent. So they tied his neck to his heels till the blood gushed out of his nose. They have already taken away our property, and that will not be enough for them without our innocent blood.

John Proctor

Ann Putnam's Confession (1706)

During the Salem Witch Crisis Ann Putnam (the daughter of Mrs. Ann Putnam above) was about 12 years old. She is often considered to be the ringleader of the accusers. About 23 years later Ann Putnam, now a woman in her thirties, stood up in church and read this confession:

In my childhood I was an instrument for the accusing of several persons of a grievous crime, whereby their lives were taken away from them. I now believe they were innocent persons. What was said or done by me I did it not out of any anger or ill-will but out of ignorance. I was deluded [tricked] by Satan. I desire to lay in the dust, for causing, with others, Goodwife Nurse and her two sisters, so sad a calamity to them and their families. For this reason, I beg forgiveness of God, and from all those who I have caused sorrow and offence whose family members were taken away or accused.

Appendix E
Historical Narrative

What Really Happened?

The True Story of the Salem Witch Hunts

By Amelie von Zumbusch

Text excluded because of copyright restrictions

Appendix F

Participant Instructions

Instructions

Thank you for participating in this study! This study is part of a project for Mr. Collette's doctoral dissertation. Please do your best. Parts of the study dealing with history will be graded. The parts of the study that are not directly dealing with history are not graded and are extremely short. But still, please do your best on the entire study.

In this first part, you will be given a historical text. Please read it carefully and try your best to understand what is going on. Once you have finished reading, move onto the next page. You will not be able to go back and reread previous pages. Later, you will be given questions to see how well you understand what you have read.

After the reading and questions, you will be expected to write a paragraph about why people behaved the way they did during the Salem witch crisis. For the paragraph, you will be given the citations but not the text, so read carefully and do not rush through the text. You do not have to memorize the text, but do your best to remember what may be important details. Focus on answering the following question:

Why did people behave the way that they did during the Salem witch crisis?

Appendix G

Author Recognition Test

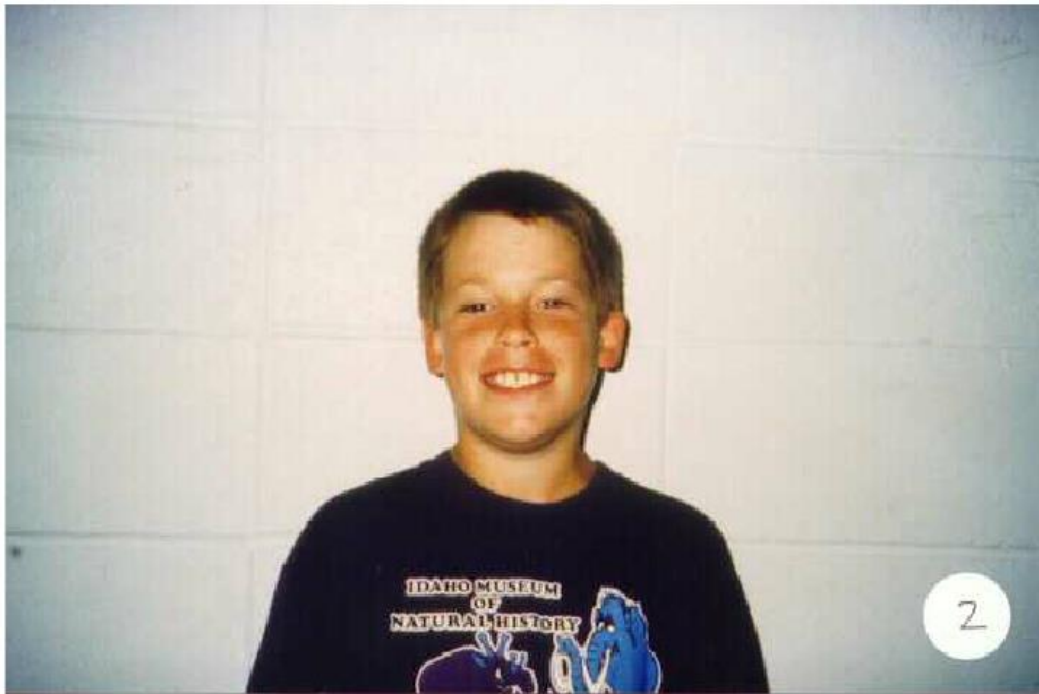
Author Recognition Test (ART)

Directions: Below you will see a list of 63 names. Some of the people in the list are popular juvenile writers and some are not. You are to read the names and put a check mark next to the names of individuals whom you know to be writers. Do not guess. Only check those whom you know to be writers because about half the names are not real authors.

Adriana Bus	Jerry L. Johns	Philip Gough
Ally Condie	JK Rowling	Ray Bradbury
Barbara M. Taylor	John Elkins	Richard E. Hodges
Beverly Cleary	John Flannegan	Richard Paul Evans
Brandon Sanderson	John Green	Rick Riordan
Carl Braun	JRR Tolkien	RL Stine
Charles Dickens	Judith Green	Robert Dykstra
Colin Harrison	Judy Blume	Rose-Marie Weber
CS Lewis	Keith Stanovich	Scott O'Dell
David Bloome	L.M. Montgomery	Scott Westerfeld
Diane Barone	Lois Lowry	SE Hinton
Donald J. Leu, Jr.	Louis Sachar	Sheila Valencia
Dorothy Watson	Ludo ver Hoeven	Stephanie Meyer
Eunice N. Askov	Margaret Meek Spencer	Susan B. Neuman
Gary Paulson	Margaret Peterson	Suzanne Collins
George McConkie	Haddix	Thomas Nicholson
Harper Lee	Marion E.D. Jenkins	Veronica Roth
Henrietta Dombey	Michael F. Graves	Victoria Purcell-Gates
James Dashner	Octavio Henao Alvarez	Warwick Elley
James F. Baumann	Orson Scott Card	Peter Bryant
Jane A. Hansen	Jane Austen	Jeff Kinney

Appendix H

DANVA 2 Child Faces



☐ Happy

☐ Sad

☐ Angry

☐ Fearful

next

Appendix I

The Faces Test Examples



SURPRISE

HAPPY



ANGRY

AFRAID

Appendix J
Emotions Survey

Emotions Survey

When people read about the Salem Witch Crisis they may have had a lot of different feelings about it. We're interested in how you felt when you were reading the texts about the Salem Witch Crisis. You, too, may have felt more than one way about it, so please think carefully about each question listed below.

The items below list several emotions that you may have felt when you were reading the texts on the Salem Witch trials. Please read the sentence. Then, for each emotion circle the number that best describes how you felt.

Sentence: *When I read about the Salem Witch Crisis , I felt:*

- | | | | | | |
|----|-------------------|----------|--------------|----------------|----------------|
| 1. | Glad | | | | |
| | strongly disagree | disagree | unsure | agree | strongly agree |
| | 1 | 2 | 3 | 4 | 5 |
| 2. | Softhearted | | | | |
| | strongly disagree | disagree | unsure agree | strongly agree | |
| | 1 | 2 | 3 | 4 | 5 |
| 3. | Mad | | | | |
| | strongly disagree | disagree | unsure agree | strongly agree | |
| | 1 | 2 | 3 | 4 | 5 |
| 4. | Warm | | | | |
| | strongly disagree | disagree | unsure agree | strongly agree | |
| | 1 | 2 | 3 | 4 | 5 |
| 5. | Moved | | | | |
| | strongly disagree | disagree | unsure agree | strongly agree | |
| | 1 | 2 | 3 | 4 | 5 |
| 6. | Sad | | | | |
| | strongly disagree | disagree | unsure agree | strongly agree | |
| | 1 | 2 | 3 | 4 | 5 |
| 7. | Disappointed | | | | |
| | strongly disagree | disagree | unsure agree | strongly agree | |
| | 1 | 2 | 3 | 4 | 5 |
| 8. | Tender | | | | |
| | strongly disagree | disagree | unsure agree | strongly agree | |
| | 1 | 2 | 3 | 4 | 5 |

- | | | | | | |
|-----|-------------------|----------|--------------|----------------|---|
| 9. | Excited | | | | |
| | strongly disagree | disagree | unsure agree | strongly agree | |
| | 1 | 2 | 3 | 4 | 5 |
| 10. | Happy | | | | |
| | strongly disagree | disagree | unsure agree | strongly agree | |
| | 1 | 2 | 3 | 4 | 5 |
| 11. | Bored | | | | |
| | strongly disagree | disagree | unsure agree | strongly agree | |
| | 1 | 2 | 3 | 4 | 5 |
| 12. | Scared | | | | |
| | strongly disagree | disagree | unsure agree | strongly agree | |
| | 1 | 2 | 3 | 4 | 5 |
| 13. | Nervous | | | | |
| | strongly disagree | disagree | unsure agree | strongly agree | |
| | 1 | 2 | 3 | 4 | 5 |
| 14. | Annoyed | | | | |
| | strongly disagree | disagree | unsure agree | strongly agree | |
| | 1 | 2 | 3 | 4 | 5 |
| 15. | Frustrated | | | | |
| | strongly disagree | disagree | unsure agree | strongly agree | |
| | 1 | 2 | 3 | 4 | 5 |
| 16. | Joyful | | | | |
| | strongly disagree | disagree | unsure agree | strongly agree | |
| | 1 | 2 | 3 | 4 | 5 |
| 17. | Angry | | | | |
| | strongly disagree | disagree | unsure agree | strongly agree | |
| | 1 | 2 | 3 | 4 | 5 |
| 18. | Compassionate | | | | |
| | strongly disagree | disagree | unsure agree | strongly agree | |
| | 1 | 2 | 3 | 4 | 5 |
| 19. | Uneasy | | | | |
| | strongly disagree | disagree | unsure agree | strongly agree | |
| | 1 | 2 | 3 | 4 | 5 |

- | | | | | | |
|-----|-------------------|----------|--------------|----------------|---|
| 20. | Irritated | | | | |
| | strongly disagree | disagree | unsure agree | strongly agree | |
| | 1 | 2 | 3 | 4 | 5 |
| 21. | Sympathetic | | | | |
| | strongly disagree | disagree | unsure agree | strongly agree | |
| | 1 | 2 | 3 | 4 | 5 |
| 22. | Upset | | | | |
| | strongly disagree | disagree | unsure agree | strongly agree | |
| | 1 | 2 | 3 | 4 | 5 |
| 23. | Worried | | | | |
| | strongly disagree | disagree | unsure agree | strongly agree | |
| | 1 | 2 | 3 | 4 | 5 |
| 24. | Surprised | | | | |
| | strongly disagree | disagree | unsure agree | strongly agree | |
| | 1 | 2 | 3 | 4 | 5 |

Appendix K

Historical Perspective Taking Rubric

Historical Perspective Taking Rubric

	0	1	2	3
Evidence	No sources	At least one claim backed by a source Evidence not used to persuade	Some or most claims backed by source(s) May not be strong evidence, not clearly used to persuade	Each claim backed by a specific source.
Contextualization	No understanding of context, use of derogatory words to describe people	Weak / implied understanding of historical context	Minimal articulation of relevant historical context	Explicit / strong articulation of relevant historical context
Perspective Taking	No references to internal of thoughts, feelings,	Weak references of internal thoughts & feelings May include egoistic drift	Minimal references of internal thoughts & feelings No egoistic drift	Strong articulation of relevant thoughts, feelings, etc. of people involved. Recognition in how they felt in contrast to how we would feel
Inferences & Imagination	No elaborations / inferences	Little elaboration, missed basic elaboration	Basic / minimal elaboration. Just perfunctory explains the surface meaning	Insightful elaboration, provide deeper inferences and explanations of the evidence

Appendix L

Text-Specific Comprehension Questions

Text-Specific Comprehension Questions

[For the following question, correct option is always 'a' , but in the version given to the students the correct answer and distractors are randomized]

Historical Document Comprehension Questions

Literal Comprehension Questions

1. According to Cotton Mather's sermon in Boston, what did he believe about witches?
 - a. They are real
 - b. They are not real
 - c. He did not know either way
 - d. He did not care either way
2. While being questioned by a judge, what did Abigail Hobbs admit to seeing?
 - a. The Devil
 - b. Witches
 - c. Nothing out of the ordinary
 - d. Angels
3. According to the petition by John Proctor, why are the witness' testimonies against the accused not fair?
 - a. The judge and jury hate the accused
 - b. Some of the witnesses are close friends to the victims
 - c. Some of the witnesses are close friends with the accused
 - d. The witnesses are all being treated with the utmost protection
4. What does Mary Easty claim in a petition about many of the people who have confessed to being witches?
 - a. They are not witches
 - b. They are murderers
 - c. They are good people
 - d. They are not Christian

Inferential Questions:

5. According to the diary of Reverend Deodat Lawson how may Abigail William's behavior be characterized?
 - a. Strange
 - b. Polite
 - c. Friendly
 - d. Shy
6. About 23 years later, how did Ann Putnam feel about her role in the Salem Witch Crisis?
 - a. She regretted it
 - b. She was proud of it
 - c. She did not feel like she did enough to find all the witches
 - d. She believed that although it was sad, it was the right thing

7. In the historical documents that you read, why didn't any of the accused argue that witches do not exist?
 - a. Because even the accused believed in witches
 - b. Because no one had ever considered witches not being real a possibility
 - c. Because they thought it was obvious that witches were not real
 - d. Because they thought it was obvious that the accusers were faking it
8. Although Ann Putnam was only about twelve years old during the Salem witchcraft crisis, it appears that she was one of the main accusers. Based on the trial of Rebecca Nurse, who mostly likely had influenced her to make these accusations?
 - a. Her Mother
 - b. Her father
 - c. Her friends
 - d. Her sister

Historical Narrative Comprehension Questions

Literal Comprehension Questions

1. According to the book, what were Puritans like?
 - a. Strict and hardworking
 - b. Relaxed and lazy
 - c. Hateful and mean
 - d. Kind and patient
2. What did they attempt to do to break the spell against Abigail Williams and Elizabeth Parris?
 - a. Make a cake out of urine and have a dog eat it
 - b. Cast the Devil out by the Holy Spirit
 - c. Burn incense and say prayers
 - d. Beat the Devil out of the two girls
3. What did the accusers/victims claim to see?
 - a. Specters or the witches' spirits
 - b. The witches dancing around a fire
 - c. The witches flying on a broom stick
 - d. The accused engaging in physical violence
4. According to the author, why were some executed for being witches?
 - a. Because they would not confess in being witches
 - b. Because they confessed in being witches
 - c. Because they falsely accused others of being witches
 - d. Because there was proof that they were witches

Inferential Questions

5. According to the author, why did the accused often confess to being witches?
 - a. To protect themselves
 - b. Because they were witches
 - c. Just as joke
 - d. Because they believed they were witches

6. What word best represents the author's view of the Salem Witch Hunts?
 - a. Tragic
 - b. Silly
 - c. Boring
 - d. Exciting
7. The author described the trial of Bridget Bishop as being the first trial. Why was Bishop the first person tried for witchcraft even though she was not the first accused?
 - a. She was the easiest person to convict (to find as guilty)
 - b. She was the oldest
 - c. She was the youngest
 - d. She was the hardest person to convict (to find as guilty)
8. The author seems to believe that witchcraft is
 - a. Not real
 - b. Real
 - c. Harmless
 - d. Fun

CURRICULUM VITAE

JARED COLLETTE

UNIVERSITY ADDRESS: School of Teacher Education and Leadership
 Utah State University
 2805 Old Main Hill
 Logan, UT 84322-2805
 E-mail: jaredpcollette@gmail.com

I. EDUCATIONAL HISTORY

- 2018 *Ph.D. Education, Utah State University. Dissertation*
 Title: Empathy in the Middle School History Classroom: The Effects of Reading Different Historical Texts on Theory of Mind, Empathic Concern, and Historical Perspective-Taking
 Chair: Dr. Suzanne Jones
- 2013 *M.A. Education, Utah State University. Thesis*
 Title: A Primary Document Analysis-Centered Curriculum in a Residential Treatment Center
 Chair: Dr. J. Spencer Clark
- 2007 *B.A. History-Teaching, minor: Music. Brigham Young University, Provo.*

II. PROFESSIONAL HISTORY

- History & Debate Teacher, Draper Park Middle, Draper, UT** 2014-Present
- Teach 8th Grade US History & debate
 - Rated a “highly effective teacher” through a rigorous teacher evaluation
 - Highly involved in extra-curricular activities as a debate coach, chess coach, and in student leadership council
- Research Assistant, School of ITLS, USU** 2015
- Gained extensive experience coding qualitative data
- Social Studies Teacher, Copper Hills Youth Center, West Jordan, UT** 2008 - 2014
- Taught social studies at a lockdown facility to youth in treatment for sexual misconduct, chemical dependency, Autism Spectrum, and criminal behavior
 - Gained experience teaching students with various learning abilities from gifted to intellectually disabled
- Online Curriculum Writer, The American Academy, Salt Lake City, UT** 2013

- Developed original social studies curriculum for a nation-wide online high school

History Teacher, *Valley Mental Health*, Salt Lake City, UT Summers 2007/2008

- Taught US and World History to adolescents at an outpatient treatment center

Full-Time Substitute, *Jordan School District*, West Jordan, UT 2007 - 2008

- Worked in two full-time positions for months at a time teaching 7th grade Utah Studies, 9th grade Geography, 7th grade art, and 8th and 9th grade ceramics

Substitute, *Jordan School District*, Salt Lake County, UT 2007 – 2008

- Discovered great teaching methods by exploring a myriad of classrooms of differing discipline methods, organizational skills, subjects, grades, and pedagogical approaches
- Developed greater classroom management acumen through solving difficult discipline situations

Piano Teacher, *Self-Employed*, Draper & Sandy, UT 2003 – 2014

- Taught students as young as 3 and as old as 50, beginning and advanced piano
- Self-managed a clientele of up to 31 students

Research Assistant, *Joseph Smith Papers Project*, BYU 2006 – 2007

- Collaborated with professional historians in editing, dating, reviewing, and publishing 19th century primary documents
- Gained a deeper appreciation and understanding for 19th century American history

III. SCHOLARY PUBLICATIONS AND ACTIVITIES

Publications – in preparation

Collette, J. P. & Jones, S. H. (in preparation). Can reading history engender empathy? Experiencing empathic concern and historical perspective taking.

Jones, S. H., Campbell, B. D., Collette, J. P. (in preparation). Emotions and teacher efficacy.

Conference Presentations

Collette, J. P., Jones, S. H., Campbell, B. D. (October, 2018). *Empathy in the middle school history classroom*. Paper presented at the annual meeting of the Northern Rocky Mountain Research Association (NRMERA), Salt Lake City, UT.

Collette, J. P. (October, 2017). In the shoes of another: Teaching historical empathy. Workshop session at Utah Council for the Social Studies (UCIRA) conference.

Collette, J. P. (October, 2016). In the shoes of another: Reading and empathy. Workshop presentation at the Utah Council for the International Reading Association.

Collette, J. P. (October, 2015). Bottlenecks of comprehension for fluent decoders. Workshop presentation at the Utah Council for the International Reading Association.

Service

Student Council DPMS 2014-Present

- Organized the program for a Veterans Day Breakfast at DPMS, and participated in numerous fund raisers and activities

Debate Club Coach 2015-Present

- As head coach I have led middle school debate team in multiple tournaments

Chess Club Coach 2015-Present

- Coach middle school students in chess club