
Research

Developing Critical Thinking Skills in High Ability Adolescents: Effects of a Debate and Argument Analysis Curriculum

April Walker¹ & Todd Kettler²

Abstract

This study evaluated the impact of a debate intervention on students' critical thinking. The design-based research project included a quasi-experimental, one group pre-test, post-test design. Results indicated small effect sizes on critical thinking using the Cornell Critical Thinking Test-Level X ($d = 0.40$) and an assessment of argument analysis ($d = 0.41$). High-ability students entered the project with stronger critical thinking than general-education students ($d = .82$; $d = .41$), and high-ability students appeared to benefit more from the intervention as the performance gaps increased in the post-test phase ($d = 1.08$; $d = .80$) suggesting possible aptitude-treatment interactions or the Matthew effect. Qualitative data indicated that students learned to (a) think on the spot, (b) analyze arguments, (c) see other perspectives, and (d) construct counter-arguments. This study corroborates previous research that indicated a relationship between high ability and critical thinking.

Key Words: gifted and talented, high ability, middle school, argumentation, debate, critical thinking, argument analysis, adolescents, design-based research

Introduction

It seems well established in the 21st century that critical thinking, communication, and collaboration are goals for all learners to prepare them to be successful in a rapidly changing world (Partnership for 21st Century Skills, 2004). While there will always be fields that rely more heavily on specific skills, being able to think critically and make well-reasoned decisions is becoming increasingly crucial to success across all domains of work. Students live in an age of global connections, instant access, and vast amounts of information (Collins & Halverson, 2009; Thomas & Brown, 2011). How do they decide which information is valid, reliable, significant, and applicable? Through formal argumentation opportunities such as debate, students can learn to communicate and think critically.

Defining Critical Thinking

While there are many different definitions of critical thinking, most of these definitions include a need for thinking rationally and using logic and reasoning to back up decisions, judgments, or arguments. Critical thinking was deconstructed in the Cambridge Assessment Taxonomy of Critical Thinking Skills and Processes, where critical thinking was defined as “analytical thinking which

¹Corresponding author, PhD Student, Department of Curriculum and Instruction, University of North Texas, USA; aprilwalker@gmail.com, ORCID: 0000-0002-8672-3394

² Assoc. Prof., Department of Educational Psychology, Baylor University, USA, ORCID: 0000-0003-3816-242X

underlies all rational discourse and enquiry...As an academic discipline, it is unique in that it explicitly focuses on the processes involved in being rational" (Black, 2008, p. 7). The elements this taxonomy recognized as components of critical thinking were analyzing arguments, evaluating arguments and claims for aspects such as plausibility and relevance, drawing conclusions about the claims, constructing one's own arguments, and reflecting on one's own reasoning.

Considering these definitions of critical thinking and the role that critical thinking is believed to play in the 21st century, the nature of curriculum design potentially changes to emphasize simulated experiences requiring critical thinking (Cone et al., 2016; Everett, Anderson, Wright, & Fontana, 2018). Where curriculum may once have been a course of study that just presented knowledge and then assessed student's absorption of that knowledge, now curriculum should be a series of experiences where students confront information and make judgments about what matters and use perspective to understand degrees of belief and take informed action (Ennis, 2013, 2018; Kettler, 2016).

The Framing of Learning Opportunities Matters

When designing learning opportunities where improved critical thinking is a wanted learning outcome, instructional methods do make a difference. A comprehensive meta-analysis of critical thinking research found that instructional approaches can be related to students' critical thinking outcomes (Abrami et al., 2008). The studies were categorized using Ennis's (1989) typology of four instructional methods to teaching critical thinking: (a) general, (b) infusion, (c) mixed-method, and (d) immersion. The infusion and mixed methods instructional approaches to teaching critical thinking were found to be the most effective. What this demonstrates is that when trying to improve students' critical thinking, direct instruction related to critical thinking processes is needed, and students also benefit from opportunities to apply critical thinking to the content areas and authentic situations. Direct instruction without application (the general approach) and asking students to use critical thinking skills to advanced and thought-provoking content without direct instruction (the immersion approach) are less effective (Abrami et al., 2008). Students need instruction and scaffolding, as well as opportunities to practice critical thinking to different contexts.

Critical Thinking and Debate

Matters Critical Thinking in Adolescents

When searching for research studies on critical thinking and adolescents, one can find many theoretical articles discussing the benefits of teaching critical thinking or the ways to teach critical thinking in the different disciplines; however, it is harder to locate research studies on how to cultivate the critical thinking skills of adolescents. Studies have explored how critical thinking is taught. When comparing imbedded and direct instruction methods for teaching critical thinking, adolescents have been found to make greater gains with direct instruction (Marin & Halpern, 2011). Activities that scaffold critical thinking and involve dialogue and interaction have been found to improve critical reading in science (Oliveras, Marquez, & Sanmarti, 2013). These studies on critical thinking in adolescents indicate that critical thinking skills are not necessarily intuitive for students and students make greater gains with direct instruction and opportunities to apply these skills in

interactive ways.

Adolescents and Argumentation

Some evidence suggests that adolescents often focus on building their own claim and struggle to evaluate the opponent's claim; thus, establishing a need for instruction in argument evaluation and building counterarguments (Crowell & Kuhn, 2014). Students have also been found to have difficulty distinguishing evidence from explanation when making a claim and forming counterarguments (Hsu, Van Dyke, & Chen, 2015). Students need to not only be taught to argue well, but they also need to be taught the epistemological understanding of what argument is and the power it holds (Kuhn & Crowell, 2011).

Researchers who study argumentation from a psychological perspective often focus on the cognitive development of argumentation. From early childhood, children seem to be able to form arguments and counterarguments and refute others. Being able to consider other perspectives to develop counterarguments and form rebuttals comes later, which is ultimately why argumentation is a skill that needs to be cultivated in education (Rapanta, Garcia-Mila, & Gilabert, 2013). One reason some adolescents may struggle with argumentation skills is that children mature at different rates; some adolescents reach the formal operational stage of cognitive development sooner than others, making them more able to understand abstract concepts and the cognitive demands of argumentation sooner than students who remain at the concrete operational stage of cognitive development (Hsu, Van Dyke, & Chen, 2015). Age and cognitive development does seem to be a factor in students' ability to construct arguments and identify fallacies. Eleventh graders were found to have more epistemological understanding than seventh and ninth graders and were able to identify more fallacies (Weinstock, Neuman, & Glassner, 2006). It was hypothesized that eleventh graders were better able to identify the fallacies due to more educational experiences with argumentation and more time spent using informal reasoning in domain-specific environments.

Very little critical-analytical thinking is present in student discussions without the scaffolding of a teacher or an intervention, but discussions have also been found to be important in promoting critical-analytical thinking (Murphy, Rowe, Ramani, & Silverman, 2014). Researchers have also studied the role of teachers in students' argumentative skills. When teachers focus solely on direct instruction, students provide less elaboration and raise fewer questions; at the same time, students provide more reasoning and ask more questions when they are encouraged to think more deeply (Hsu, Van Dyke, & Chen, 2015). When studying the effects of teacher guidance on collaborative argumentation of seventh graders, it was found that the intervention groups who had more teacher guidance than the control groups led to more elaborated reasoning, evidence, and counterarguments in the argumentative process of the students (Hsu, Van Dyke, & Chen, 2015). These different studies indicate that argumentation is a valuable skill that students may struggle to develop completely on their own, but that students can reach greater depths of thinking through scaffolding and interventions designed to cultivate critical thinking and argumentation skills.

Debate as an Instructional Method

Debate, as an instructional method, offers opportunities for students to learn and practice critical

thinking, develop their communication skills, and develop deeper understanding and mastery of content through active engagement. The nature and format of a debate lends itself to opportunities to teach and allow students to practice critical thinking skills. Debate is offered in many formats (especially for competition). Traditionally debate involves an issue with two distinct opposing views. A resolution is given; one team affirms the resolution and one team negates the resolution. Debates can be done in teams or as individuals. Prior to the debate, each team builds a case to prove or disprove the resolution using evidence (Scott, 2008). The debate will usually begin with the affirmative presenting their side of the resolution, and then the negative presenting the counterargument. Usually, each side is then given the opportunity to cross-examine and offer rebuttal to the opposing side. Debate offers students an opportunity to build empathy and possibly (at least temporarily) reduce their bias because they are forced to look at multiple viewpoints of an issue (Kennedy, 2009). In a debate, students may be asked to defend positions they may or may not personally agree with. They will be forced to consider reasons and evidence to support the claim they are being asked to make. Debaters also will have to consider the other viewpoint in order to evaluate the quality, reliability, and rationality of the arguments and evidence given by the opposing side.

Studies have looked at using forms of debate in the classroom and measured student perception of debate on its impact to their communication and critical thinking skills (Oros, 2007). Overall, these studies have found the perceptions to be very positive with students feeling they improved their confidence, gained insight into seeing multiple perspectives, and improved their argumentation skills. Participation in competitive debate shows even greater improvement in critical thinking skills than just instruction in argumentation (as cited in Allen, Berkowitz, Hunt, & Loudon, 1999; Bellon, 2000).

Aptitude-Treatment Interaction Theory

Aptitude-treatment interaction theory predicts that some instructional strategies will be more or less effective for learners based upon the aptitudes, or specific abilities, of the learners (Cronbach & Snow, 1977). One of the general principles derived from the theory suggested that highly structured learning environments tend to be more effective with lower ability students, and less structured learning environments tend to be more effective with higher ability students (Snow, 1989). Whether the variation resides in elements of the learning environment (structured or unstructured) or the complexity of the instruction intervention, aptitude-treatment interaction theoretically proposes that some learning tools may work better with high ability students. VanTassel-Baska (2000) argued that the principles associated with differentiated instruction have their roots in aptitude-treatment interaction theory and the theory of individual differences. Based on the evidence that critical thinking skills are positively associated with cognitive ability (Kettler, 2014), it is reasonable to predict the presence of an aptitude-treatment interaction in which the debate critical thinking intervention yields greater growth in high ability students than in their general education peers.

Design-Based Research

Design-based research (DBR) is a relatively new approach in educational research (Anderson &

Shattuck, 2012) with a goal of bridging some existing gaps between educational theory and educational practice (Bakker & van Eerde, 2015). A distinguishing feature of DBR is the simultaneous work of designing educational tools while also conducting context-based research on the effectiveness of those tools. Educational environments are complex and include variables typically beyond the control of the research team. However, laboratory-based educational research while more controlled may fail to yield evidence-based practices that thrive within authentic and complex learning environments. DBR involves the participants (e.g. teachers) in the design process with the intent of effective integration of learning interventions well-suited for actual learning spaces (Plomp & Nieveen, 2007).

The present study employs the following features of DBR: (a) researcher/teacher design of an authentic assessment protocol for argument analysis and a domain-specific intervention; (b) testing both what works (quasi-experimental design) as well as why it works (qualitative data from both the teacher and students), and (c) simultaneously developing and testing theory of critical thinking in adolescents. The potential benefits of design-based research in gifted education is relatively unexplored (Jen, Moon, & Samarapungavan, 2015), but an ongoing study from the Center for Talent Development at Northwestern University has been successfully employing principles of design-based research (Olszewski-Kubilius, et al., 2018) in the design of curriculum and services to develop academic talent in middle school students.

Research Questions

In design-based research, the design of educational materials is a crucial part of the research. In this study, two educational designs were investigated: (a) the debate curriculum focusing on argument analysis and critical thinking, and (b) the argument analysis scoring protocol. The purpose of the study was to design and empirically evaluate a teaching approach capable of improving students' skills at critical thinking/argument analysis. Data were gathered and analyzed to answer the following research questions:

1. Does the use of formal debate as an instructional strategy improve students' critical thinking skills?
2. What are the perceptions of students participating in debate about if and how debate influenced their critical thinking skills and argument analysis skills?
3. Do high ability students (> 95th percentile achievement) respond similarly or differently than other peers to the debate instructional strategy? In other words, is there any evidence of an aptitude-treatment interaction effect within the debate intervention?

Method

Participants

There were 17 students who participated in the debate intervention study. Eight (47%) were female, and nine (53%) were male. Fourteen (82%) of the students were White, and three (18%) students were Asian. The average age of participants was 13 years and 5 months at the beginning of the intervention. Eight of the students were in grade 7, and nine of the students were in grade 8 at the

time of the study. Seven of the students were classified as high ability based on scores at or above the 95th percentile in either reading or mathematics on the Comprehensive Testing Program (CTP-4, Educational Records Bureau, 2011). The high ability students ($n = 7$) had mean percentile rank scores of 98.6 in mathematics and 90.0 in reading. The other students ($n = 10$) had mean percentile rank scores of 70.8 in mathematics and 66.9 in reading.

Because of the design-based research approach, the teacher was also considered a participant in the study. The teacher who taught the speech course which included the debate intervention was a 32-year old, white female who had eight years of teaching experience. The teacher had teaching certifications as a Generalist for early childhood through grade 8 and English/Language Arts for grades 8-12, as well as supplemental certifications in English as a Second Language and Gifted and Talented Education. The goal of the study was to design an intervention and assessment protocol that were well-adapted to the complex nature of a middle school classroom and easily implemented in a language arts curriculum. Thus, the teacher participated in the design of the intervention and the design of the argument analysis assessment protocol.

Research Design

Design-based research is interventionist in nature, and in this study, the research team developed and tested a specific teaching and learning protocol (structured debate) representative of critical thinking pedagogy. A quasi-experimental, one-group, pre-test -- post-test design (Shadish, Cook, & Campbell, 2002) was used to test the effects of the debate intervention. The critical thinking dependent variable was measured in two ways: (a) the Cornell Critical Thinking Test-Level X and (b) researcher-designed argument analysis tasks. Kline (2009) demonstrated that one way to improve the one-group, pre-test -- post-test design is to use multiple pre-test or post-tests. Thus, to strengthen the design, we used repeated measures on the argument analysis task in both the pre-test and the post-test phases of the design, as well a standardized test of critical thinking—Cornell Critical Think Test-Level X.

Data Collection

Prior to beginning the debate intervention or the pre-test argument analysis tasks, participants completed a standardized assessment of critical thinking—the Cornell Critical Thinking Test-Level X (CCTTX; Ennis, Millman, & Tomko, 2005). At the end of the project, after all argument analysis tasks had been completed, participants again completed the same form of the CCTTX. In addition to the standardized assessment of critical thinking, the research team developed and used an authentic assessment protocol to measure students' argument analysis skills, a recognized component of critical thinking (Black, 2008). Participants completed six argument analysis tasks prior to receiving the intervention, and then completed six more argument analysis tasks after receiving the intervention. No more than two argument analysis tasks were administered in any single week. Thus, for the pre-test and post-test phases, the argument analysis tasks were completed over several weeks. For the argument analysis tasks, articles were selected from the debate section from 6 issues of *The New York Times Upfront* magazine. The reading selections from NYT Upfront presented two short articles written to present opposing viewpoints on a contemporary issue relevant

to adolescents. Each article was approximately 350-400 words in length. The Lexile level of the articles ranged from 1100L to 1200L. Students completed the Analyzing Author's Claim protocol for each article.

Because design-based research is concerned with both the outcome of the intervention and the context and processes with which it was tested, interviews were conducted with the four students who demonstrated the greatest improvement from pre-test to post-test on the critical thinking argument analysis tasks. This sequential (quantitative then qualitative) mixed-method design theoretically uses the qualitative data to help explain the quantitative data (Creswell, Plano Clark, Gutmann, & Hanson, 2003; Steckler, McLeroy, Goodman, Bird, & McCormick, 1992).

Data Analysis

To score the argument analysis tasks, a rubric was created to measure argument analysis (see Table 1). The rubric generated a total score as well as scores in four sub-categories: recognizing arguments and claims, categorizing the components of the argument by identifying reasons, categorizing the components of the argument by identifying evidence, and recognizing different types of reasoning. Each section is scored on a four-point scale ranging from zero to three. Thus, each argument analysis task generated a total score ranging from 0 to 12. Two research assistants were trained to use the rubric and score the argument analysis tasks. The authors provided the 90-minute training for using the rubric. It included an overview of the project, direct instruction on the four categories of argument analysis on the rubric, and practice scoring and calibration discussions on each task scored. The practice scoring tasks used the same articles but were completed by students in a pilot test of the argument analysis tasks. The research assistants scored practice tasks until they demonstrated thorough understanding of the scoring rubric. During the scoring of the argument analyses of the study, the tasks were randomized so that the research assistants did not know whether a task was completed during pre-test or post-test phase. Also, identification codes were used so that the research assistants had no information about the participant who completed each task.

Table 1. Argument Analysis Rubric

	3	2	1	0
Recognizing Arguments and Claims	The student clearly identified the author's main <i>claim</i> and argument in a clear and concise manner.	The student identified the author's main claim and argument but could have articulated the claim and argument more clearly.	The student attempted to identify the main claim but appeared to have misunderstood the author.	The student did not recognize the author's main claim.
Categorizing the Components of the Argument: Identifying Reasons	The student clearly identified all the <i>reasons</i> offered by the author to support the claim.	The student identified all the reasons offered by the author but could have articulated them more clearly.	The student identified some of the reasons offered by the author to support the claim, but not all of them.	The student did not identify the reasons offered by the author to support his/her claim or listed reasons that do not seem to support the claim.

Table 1 (continued). Argument Analysis Rubric

	3	2	1	0
Categorizing the Components of the Argument: Identifying Evidence	The student described <i>examples of evidence</i> for all the reasons offered by the author and <i>cited specific quotes from the author</i> .	The student offered examples of evidence for all the reasons offered by the author but could have provided quotes or elaborated more clearly.	The student offered examples of evidence for some of the reasons offered by the author, but not all. Student also could have elaborated more clearly.	The student did not identify evidence to support the reasons or offered conflicting evidence (which may indicate lack of student's understanding of the argument).
Recognizing Different Types of Reasoning	The student identified <i>the types of reasoning</i> offered by the author and offered his/her own reflection of the effectiveness of the types used.	The student identified the types of reasoning offered by the author but could have reflected more clearly on the effectiveness of these types of reasoning.	The student attempted to identify the types of reasoning used by the author but did not identify all of them or struggled to identify the types correctly.	The student did not identify the types of reasoning offered by the author.

Instruments

Cornell Critical Thinking Test-Level X. The CCTTX was first published in 1985 and is now in the 5th revised edition (Ennis, et al., 2005). The CCTTX is recommended for students in grade 4 through early high school. The test also includes a Level Z which is recommended for high school and college students. In a recent meta-analysis (Abrami et al., 2008) examining instructional interventions and measurements of critical thinking, the two forms of the Cornell Critical Thinking Test were noted as two of the more common instruments used to measure critical thinking. The CCTTX allows students 50 minutes to complete the assessment yielding a raw score between 0 and 71. The CCTTX measures the following aspects of critical thinking: (a) induction, (b) observation and credibility, (c) deduction, and (d) recognizing assumptions. Though it is a general measure of critical thinking, it does not specifically identify argument analysis as a feature of its measurement design. The CCTTX administration guide lists reliability estimates from previous studies ranking from .67 to .90 with a median estimate of .80, and Kettler (2014) reported a Cronbach $\alpha = .89$ with a population of both general education and students. The observed internal consistency estimate for the pre- and post-tests high ability administered in the present study was Cronbach $\alpha = .83$

Argument Analysis Rubric. The argument analysis rubric (AAR) was designed for use in this study; thus, one of the goals of this design-based project was to gather and report psychometric data for the AAR. There were 17 participants and a total of 177 argument analysis tasks scored using the AAR. Each AAR task included four sub-scores on a scale of 0 to 3 resulting in a total score ranging from 0 to 12. A two-way, random effects intraclass correlation coefficient (ICC) model based on mean rating ($k = 2$) was used to estimate the consistency of the two raters who were trained to use the AAR (McGraw & Wong, 1996). A high degree of reliability was found between Rater A and Rater B. The average measure ICC was .82 [95% CI: .78, .86] ($F(175, 1575) = 5.64, p < .001$). ICC values between 0.75 and 0.90 are considered to indicate good reliability and an equivalent Cronbach α estimate of .823 (Koo & Li, 2016; Shrout & Fleiss, 1979). Even with the good reliability,

a third trained rater was used for more conservative estimates in cases ($n = 47$) where Rater A and Rater B differed by more than three points on the total score (scale of 0 to 12). In those cases, the final score for the student argument analysis became the mean of all three raters rather than two raters. A second analysis on the 47 argument analyses scored with three raters ($k = 3$) revealed moderate reliability between the three raters on the total scores for those 47 tasks yielding average measure ICC of .55 [95% CI: from .26, .73] ($F(46, 92) = 2.20, p = .001$).

Missing data. Due to participant absence from class some students did not complete all 12 tasks. Six participants completed 10 of 12 tasks, and three students completed eight, seven, and six tasks respectively for a missing data rate of 13.3%. The missing data were estimated to be missing at random (MAR) because the missing values (y) depended on the observed variable ($x =$ absent from class) but not on ($y =$ performance on the argument analysis) (Little & Rubin, 2014). Thus, data imputation techniques were used to preserve all the data in this small sample. Specifically, imputation via linear interpolation was used because of the time-series nature of multiple pre- and post-test observations (Salgado, Azevedo, Proença, & Vieira, 2016). Interpolation is the process of using existing data for each student to estimate missing data values for that same student. Because the study compared pre- and post-assessment values, the research team performed the interpolation for missing data using only data available in either the pre- or post-assessment phase of performance, depending upon where the missing data occurred.

Qualitative Analysis of Interviews

The four students who demonstrated the greatest improvement in their mean scores on the critical thinking argument analysis tasks between pre and post argument analysis tasks were interviewed using semi-structured interviews. The structured interview questions were as follows: (a) What do you feel you have learned in debate class? (b) In what ways did your ability to analyze arguments improve in this debate class? (c) Why do you think your scores improved on the argument analyses tasks? (d) Do you feel your argument analysis skills have improved? Why? How? (e) What were the most important skills you learned during the debate process? and (f) How might you use the argument analyses skills you learned in the debate program in the future? Additional questions asked in the interviews were follow up questions based on student responses to the structured questions.

The interview questions were analyzed using thematic analysis (Braun & Clark, 2006). An inductive method of coding was used which means that codes were generated by reviewing the data without trying to fit the codes into preconceived notions (Braun & Clark, 2006). Themes were determined by looking for patterns of coding across the interviews and by considering the relevance of the codes to the research questions. The thematic analysis was conducted using the six step process suggested by Braun and Clark (2006): (a) the interviews were transcribed and then read and reread to note initial ideas; (b) initial codes were created after reading through the interview transcripts; (c) themes were generated from among the codes; (d) themes were reviewed; (e) themes were defined; and (f) a report of findings was created.

Results

Descriptive data for each variable is presented in Table 2 for all students as well as groups of (a) male and female students and (b) general education students and gifted and talented students. Additionally, Table 3 presents descriptive data for each of the 12 argument analysis tasks (AA). Because of the small sample size, the research team paid close attention to the skewness and kurtosis coefficients indicating distributions to satisfy the assumptions of our analyses. The tasks for argument analysis were randomly ordered for this study. While each task's reading came from the same source with similar Lexile levels, the students' feedback and the data suggested that some readings may have been more challenging to analyze than others—specifically pre-test task 4 and post-test task 1. Student responses on each item of the AA tasks varied from short phrases of only a few words to several sentences with substantial detail. Observation and student feedback suggested that when students had a strong interest in the topic (e.g. school uniforms) they responded more completely than when they had weak interest in the topic (e.g. voter turnout).

Table 2. Descriptive Data for Variables and Groups

		N	Minimum	Maximum	Mean	St. Dev.
All Students	CCTT Pretest	17	36	60	47.35	6.06
	CCTT Posttest	17	37	58	49.71	5.79
	Analysis Task Pre	17	4.25	9.56	7.37	1.34
	Analysis Task Post	17	4.33	10.70	8.01	1.77
Female Students	CCTT Pretest	8	36	60	45.38	6.91
	CCTT Posttest	8	37	58	48.25	7.27
	Analysis Task Pre	8	5.36	9.56	7.71	1.29
	Analysis Task Post	8	5.75	10.33	8.70	1.56
Male Students	CCTT Pretest	9	39	55	49.11	4.94
	CCTT Posttest	9	43	56	51.00	4.09
	Analysis Task Pre	9	4.25	8.61	7.06	1.38
	Analysis Task Post	9	4.33	10.70	7.39	1.79
High Ability Students	CCTT Pretest	7	44	60	50.14	5.46
	CCTT Posttest	7	48	58	53.00	3.00
	Analysis Task Pre	7	5.36	8.61	7.69	1.16
	Analysis Task Post	7	5.75	10.70	8.79	1.73
General Education Students	CCTT Pretest	10	36	55	45.40	5.93
	CCTT Posttest	10	37	56	47.40	6.26
	Analysis Task Pre	10	4.25	9.56	7.14	1.47
	Analysis Task Post	10	4.33	9.42	7.45	1.65

Table 3. Descriptive Data for Pre and Post Argument Analysis Tasks

Analysis Task	Minimum	Maximum	Mean	St. Deviation	Skewness	Kurtosis
Pre 1	4.00	9.50	7.46	1.56	-.83	.03
Pre 2	5.00	10.00	7.25	1.43	.20	-.53
Pre 3	3.50	11.50	7.17	1.84	.27	1.10
Pre 4	1.00	8.33	5.35	2.06	-.56	-.34
Pre 5	3.50	11.00	7.87	1.97	-.61	-.20
Pre 6	5.50	11.00	8.95	1.68	-.63	-.26
Post 1	2.00	9.50	6.87	2.40	-.75	-.63
Post 2	1.50	10.70	7.91	2.53	-1.07	.93
Post 3	6.00	12.00	9.14	1.82	-.23	-.97
Post 4	4.33	11.50	7.84	1.95	.03	-.62
Post 5	5.00	11.50	8.72	1.94	-.35	-1.02
Post 6	1.00	11.00	7.55	2.89	-.96	.21

Improving Critical Thinking Skills

For the first research question, the hypothesis that the 12-week debate curriculum would improve students' critical thinking skills was tested? Two separate measures were used to test the hypothesis that the debate curriculum would improve critical thinking skills. For both measures, normality of the data was verified by analyzing skewness and kurtosis as well as visual inspection of distributions (histograms). For argument analyses, pretest skewness was -0.61 and kurtosis was 1.33. Posttest skewness was -0.44 and kurtosis was -0.46. For the CCTTX pretest, skewness was -1.08 and kurtosis was -0.14. Posttest skewness was 0.27, and kurtosis was 0.75. Paired sample *t*-tests were used to compare the pre-test and post-test scores. Using the CCTTX there was a mean difference between the pre-test and the post-test of 2.36 [95% CI: -1.62, 6.34], $t(31.5) = 1.16$, $p = .13$. The estimated effect size of the intervention with this sample was $d = 0.40$ [95% CI: -0.28, 1.07], a small effect by standard interpretation of Cohen's *d*. Second, using the multiple-measure, argument analysis task, there was a mean difference between the set of pre-tests and the set of post-test observations of 0.64 [95% CI: -0.42, 1.70], $t(31.5) = 1.19$, $p = .12$. The estimated effect size of the intervention with this sample was $d = 0.41$ [95% CI: -0.27, 1.09], again, a small effect by standard interpretation of Cohen's *d*.

Both the CCTTX and the argument analyses measures indicated a small effect. In other words, students demonstrated slightly better critical thinking skills in the post-test phase than they did in the pre-test phase. Practice effects may occur when participants are exposed to an assessment task multiple times even without feedback (Wesnes & Pincock, 2002). In this study, students did not receive any feedback on their argument analyses task performance. The data suggests a slight positive slope in average performance at both the pre-test and post-test phases; thus, it should be assumed that the group of participants improved their argument analysis performance slightly during both the pre- and post-test phases of the study. Being conservative in the analysis, the research team acknowledges this to be potential evidence of a practice effect. Inferences of causation require three conditions: (a) relationship condition between the variables (constructing and critiquing debates is related to critical thinking), (b) temporal antecedent condition (The first six argument analyses preceded the intervention and the second six argument analyses were conducted after the 12-week debate intervention), and (c) lack of an alternative explanation condition (Shadish, et al., 2002). The first two conditions for causal inference are clearly met in this study; however, the practice effect may offer a potential alternative explanation for the estimated small effect-size improvements between the pre-test and post-test phases.

Understanding the Process of Improving Critical Thinking Skill

Qualitative data generated from interviews were analyzed using thematic analysis techniques to better understand the process of how students improved their critical thinking skills. Several themes were found in the interview data around the student perceptions of the debate curriculum and its impact to their argument analysis skills. The categories and themes (see Table 4) were generated around students' perceptions of how debate: (a) improved their critical thinking skills, (b) benefits and skills the students believed they gained from debate, (c) perceptions of how and why

their argument analysis skills improved, and (d) the relevance of the skills learned in debate to their future.

Table 4. How the Debate Intervention Improved Critical Thinking Skills

Themes	Evidentiary Examples
How to think on the spot	Students mentioned that debate taught them to think on the spot. One participant specifically mentioned, "I think I learned more about how debate works and especially with rebuttals and stuff I have learned to think on the spot and use my points against somebody else's" (Participant 090325). A second participant noted, "Probably because in a debate you learn how to think on the spot and what to say in response and take their argument and really break it down and think about what you can say in response and that really helped" (Participant 131708).
Being able to break down arguments	Students discussed that debate helped them to break down arguments into different parts and evaluate each argument as a good or bad argument or an argument with faulty reasoning. The two students identified as gifted and talented focused more on how they learned how to break down arguments. One participant explained, "I'm better like I said at finding when an argument is bad, doesn't hold up, or when it has faulty reasoning. In the same vain, I'm better at realizing when I have a bad argument...like when I realize 'Oh, my stance here isn't really supported.' And you know I have to change outlooks based on that (Participant 040721).
Seeing multiple points of view	All of the students mentioned that participating in debate helped them to see multiple sides of an issue. This helped them to be prepared for whichever side they had to argue and to help them refute the other side during a debate and make their own points stronger. One participant stated, "Well I guess I'm better at identifying when an argument is bad and when an argument just doesn't hold up or fallacies that they use. I'm better at forming an argument that actually has meaning behind it and looking at it from all sides and making sure it is actually a good argument" (Participant 040721).
Knowing how to use an opponents' points against them	Almost all of the students mentioned that participating in debate and practicing helped them learn how to break down arguments made by an opponent and use things the opponent said against them in the refutation process. One participant stated, "I'm able to understand what they are actually saying. Basically, I know the opposite side to that so I know what I need to say so I know what's going to make my case sound better" (Participant 010521).
Note-taking skills	Students indicated that with practice they began to understand debate more and in the process they got better at note-taking in the rounds. Their notes helped them to plan how to refute their opponent. One participant noted, "I think so because last year at the beginning of the year when you gave me that I wasn't sure what to do. But afterwards I began to understand it more, so I can write stuff down more and actually know what to say" (Participant 090325). She later explained more, "I'm not really sure but maybe taking notes because now they are more organized, I write everything down now, so it helps when I am going to say something and I'm next" (Participant 090325).
Listening skills	The students mentioned that a big part of the process in debate is becoming a better listener to be able to take good notes and plan out responses to opponents. One participant said, "Yeah even in like simple arguments with other people, you can see what they are saying and take other points and evidence and make your side sound better. And also show that you are still listening to them" (Participant 010521).
Confidence	Almost all of the students mentioned that they have gained more confidence in themselves as speakers and debaters through the process. "I feel like I have learned to be more confident in myself whenever I'm speaking. Like a few years ago whenever I had a speaking assignment I was absolutely awful at it because I was so nervous. I was so self-conscious, but I feel like debating and public speaking really helped gain confidence through that. And it also helped me articulate myself better" (Participant 040721).

Table 4 (continued). How the Debate Intervention Improved Critical Thinking Skills

Themes	Evidentiary Examples
Self-awareness	The students mentioned that practice and taking good notes helps them be aware of where they need to improve for the next round. One participant said, "Yeah because sometimes towards the end I write it down and sometimes I'll look down and realize at the end that I said something I shouldn't have said. I'll look over it so it helps to get better next time" (Participant 090325).
Knowing how to write a debate case so others can understand it	Students mentioned that one of the most important skills they gained in debate was how to write and structure a debate case. When asked about one of the most important skills they gained, one participant responded, "probably how to write a debate case and how to structure it so others can understand it" (Participant 010521).
Practice	Students noted that practice was how they improved at debate. They felt practicing gave them more opportunity to build self-awareness and improve their critical thinking skills. One participant mentioned, "I think that just debating in general you can keep working and know what you need to work on and practice at" (Participant 131708).
Presentations and Public Speaking	Students mentioned that the primary way they felt debate would help them in the future would be in presentations and public speaking. One participant stated, "I feel like it is skills you can use in other classes too like presentations and talking in front of other people" (Participant 131708).

Several themes were found in the data demonstrating how the students believe the debate curriculum improved their critical thinking skills. Students discussed that debate taught them how to think on the spot, how to break down arguments, how to see multiple points of view, and how to use an opponent's points against them. From the debate intervention, students believed they improved in note-taking skills, listening skills, confidence, and self-awareness. They also felt that they learned how to write a debate case so others can understand it. Students perceived that their argument analysis skills improved due to preparing for and engaging in debate, and they believed that practicing debate rounds was much more influential than just observing other students debating. They reflected after each round and made decisions about how to improve for the next round. Students believed debate would help them in the future when they engaged in presentations and public speaking opportunities.

High Ability Student Effects

For the third research question, the hypothesis was tested that students with high cognitive ability would demonstrate greater growth during the intervention than students with average-range cognitive ability. As noted above, research question one indicated small effects for the debate intervention—on average, critical thinking performance improved for all students in the study. Prior to the intervention, there were slight differences in critical thinking ability between the high ability students and the general education students in the study as indicated on both the CCTTX and the argument analyses (AA) (see Table 2). On post-test analyses with both measures, the mean differences between the high ability students and the general education students increased. Specifically, on the CCTTX, the pre-test mean difference was 4.74 [95% CI: -1.29, 10.78], $t(15) = 1.68$, $p = .12$, $d = 0.82$ [95% CI: -0.18, 1.83], and on the AA the pre-test mean difference was 0.56 [95% CI: -0.81, 1.92], $t(15) = 0.84$, $p = .42$, $d = 0.41$ [95% CI: -0.57, 1.38]. However, during post-testing after the debate intervention, the mean difference on the CCTTX increased to 5.60 [95% CI: 0.68, 10.5], $t(13.7) = 2.46$, $p = .03$, $d = 1.08$ [95% CI: 0.04, 2.11]. Similarly, the post-test mean difference more than doubled on the AA to 1.34 [95% CI: -0.43, 3.11], $t(15) = 1.61$, $p = .13$, $d = 0.80$ [95% CI: -0.20, 1.80].

To look deeper into these differences, the research team examined the correlation matrix between the post-test dependent variables and three demographic variables representing cognitive ability (see Table 5). Participants in this sample were classified as high ability based on exceptional metrics of ability in either mathematics or reading achievement (> 95th percentile). Exceptional mathematics ability was most strongly related to critical thinking performance on the CCTTX accounting for 37% of the variance. However, there was no significant relationship between mathematics ability and the argument analysis tasks where mathematics ability only accounted for an estimated 4% of the variance. Notably, on both measures of critical thinking, mathematics ability accounted for more variance in critical thinking than reading ability as measured by either reading achievement or Lexile scores. Students' Lexile scores accounted for at best 16% of the variance on the CCTTX, but neither reading ability metric accounted for more than 3% of the variance on the AA tasks.

Table 5. Correlation Matrix for Demographic and Post-Test Variables (n =17)

	CCCT Post	AA Post	Lexile	Reading	Math
CCTT Post	1				
AA Post	.166	1			
Lexile	.395	.186	1		
Reading	.301	.187	.959**	1	
Math	.609**	.201	.517*	.475	1

CCTT Post = Cornell Critical Thinking Test Level X post-test; AA Post = argument analysis task post-test; Lexile = numeric representation of an individual's reading ability; Reading = national percentile rank on a reading achievement test; Math = national percentile rank on a mathematics achievement test.

* $p < .05$; ** $p < .01$

Discussion

Developing critical thinking skills among all students has become increasingly popular as an educational goal in the context of ubiquitous information technologies. Despite the interest, there has been sparse empirical evidence documenting promising practices for developing critical thinking skills or measuring critical thinking skill growth. This small-scale, design-based study was an initial step in that direction. The data indicated small effect sizes on two separate measures of critical thinking—one of which was a standardized assessment that has been in use for more than 30 years (CCTTX), and another that was a researcher/practitioner-designed authentic assessment tool. With a small sample and broad confidence intervals, those effect sizes should be interpreted cautiously. On the positive side, both authentic measures and standardized measures yielded almost identical pre-post effect sizes lending some confidence to estimated effects of the intervention.

Additionally, causal inference must be carefully considered in this study. While the study was designed to ideally make causal inference about the capacity of the debate intervention to yield improved critical thinking skills, the data suggested a potential practice effect associated with the use of multiple measures in both the pre-test and post-test phases. The most prudent discussion of these results must acknowledge that the small effects for improved critical thinking could be attributed to the debate intervention, but they could also reasonably be attributed to completing multiple (6 to 12) argument analyses tasks. The students were not provided grades or feedback from any of the argument analyses tasks, but the very act of repeating the same task multiple times may have yielded improvement. Fortunately, the research design included two measures of the dependent variable, and the post-test results on the CCTTX also indicated an identical small effect

as did the argument analyses tasks. This dual measurement of the dependent variable lends support to the causal inference of the debate intervention. Regardless, it is worth noting that practicing argument analyses consistently over time may yield increases in general critical thinking capacity.

The qualitative analysis of student interviews provided some contextual understanding of how the debate intervention was related to improvements in critical thinking skills. Each of the themes identified demonstrated how the debate process was connected to the taxonomy of critical skills articulated by Black (2008). Students mentioned that they learned (a) to see issues from multiple points of view, (b) to break down arguments, (c) to construct arguments/debates, and (d) to recognize fallacies or faulty reasoning. Interestingly, students mentioned that they improved their listening skills and note-taking skills as well. Though not specifically recognized in the critical thinking skills taxonomy, skills at listening and noting may be considered pre-requisite or concurrent skills that support the development of critical thinking. The qualitatively derived theme of practice furthers the discussion on the causal connections between doing multiple argument analyses and participating in multiple debates. One might contend that developing critical thinking skills in this project may be equally associated with the debate curriculum and the opportunities to intentionally practice analyzing and constructing arguments through the argument analysis task. In other words, there remains a blurred distinction of whether the argument analysis tasks were simply a measurement tool or an unintentional component of the intervention. On a positive note, the students in the study demonstrated improved critical thinking skills despite the blurred distinction, and subsequent studies might be able to further sort out the causal contributions of the debate and the argument analyses tasks.

Previous research has supported the benefits of structured debate and argumentation as an instructional strategy (Bellon, 2000; Crowell & Kuhn, 2014; Oros, 2007), and the results of this study corroborate those previous findings. Debate serves as a potentially useful instructional method to build students ability in several aspects of critical thinking—notably in this case, argument analyses. In this study, the debate instructional strategy was used in a speech class and the topics of debate were current events. Further validation of debate as an instructional tool to develop critical thinking needs to integrate the debate strategy with discipline specific topics and courses in fields such as science, social studies, economics, or literature. While structured debate may improve general critical thinking skills, it would be worth investigating whether debate activities improve conceptual understanding of domain-specific content.

The findings of this study indicated a relationship between high ability and critical thinking skills assessed using the CCTTX. These results corroborated two previous studies where critical thinking skills and cognitive ability were associated. Van Tassel-Baska, Bracken, Feng, & Brown (2009) sorted participants into groups based on cognitive ability and found a positive relationship between cognitive ability and critical thinking scores. Similarly, Kettler (2014) demonstrated relationships between critical thinking and high ability specifically using the CCTTX. The relationship between the CCTTX and mathematics achievement in this study with middle school students was $r = .61$ and Kettler found a relationship of $r = .49$ with grade 4 students. The relationship between the CCTTX and reading achievement in this study was $r = .30$ and Kettler found a relationship of $r = .48$. Interestingly the relationships between reading ($r = .19$) and mathematics ($r = .20$) achievement

and the authentic tasks of argument analyses were weaker than the associations between reading and mathematics and the CCTTX. Further investigations of these relationships are warranted to further the discussion of how best to measure growth in students' critical thinking skills over time. Regardless, higher levels of demonstrated ability or achievement continue to be associated with higher levels of critical thinking.

A question of interest in this study focused on theoretical presumption that students with high ability would respond differently to the intervention. As noted in the results, the high ability students scored higher than the general education students in the pre-test phase of the study on both the standardized measure and the authentic measure. Moreover, the observed differences in critical thinking performance between the high ability students and the general education students increased on both measures at the post-test phase. In other words, the performance differences based on cognitive individual differences increased. This differential improvement can be interpreted as an aptitude-treatment interaction or a closely related phenomenon, the Matthew effect. While both groups of students (general education and high ability) improved their critical thinking skills from the pre-test to the post-test phase of the study, the high ability students showed more improvement as a group. It is also worth noting that the general education group in this study were above-average in their own reading and mathematics achievement as a group performing at approximately the 70th percentile in each domain. Thus, it is possible that the differential effects of the study might have even been greater had the general education group been closer to the general education norm of roughly 50th percentile achievement in reading and mathematics.

Limitations

There is no doubt this study is limited by sample size. Larger samples will yield more reliable estimates of effects and narrower confidence intervals for those effects. Larger samples will also allow for deeper componential analyses of critical thinking and argument analyses. One aspect of this design-based study was to design and implement an argument analysis scoring tool. That tool had four-components. Similarly, the CCTTX purports to measure four aspects of critical thinking. With the small sample in the present study, multivariate analyses of the sub-components of those measures were out of reach. However, while analyses of those sub-components will be of interest in subsequent research, the absence of those analyses should not diminish the finding of small effects for this intervention and promising reliability of the argument analyses rubric.

Conclusion

Developing critical thinking skills appears to be an acknowledged educational goal, and research efforts to document instructional tactics to accomplish that goal are needed. Moreover, developing critical thinking skills among gifted and talented students has been a stated educational goal for almost six decades, yet documented evidence of how to accomplish that goal remains sparse. The study indicates that debate as an instructional strategy and the practice of argument analyses can lead to improved critical thinking skills in all students, but especially in those who begin with high cognitive ability.

References

- Abrami, P. C., Bernard, R. M., Borokhovski, E., Wade, A., Surkes, M. A., Tamim, R., & Zhang, D. (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis. *Review of Educational Research, 78*, 1102-1134. doi:10.3102/003465431455106.
- Allen, M., & Berkowitz, S., Hunt, S., & Louden, A. (1999). A meta-analysis of the impact of forensics and communication education on critical thinking. *Communication Education, 48*(1), 18-30. doi:10.1080/03634529909379149
- Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research? *Educational Researcher, 41*, 16-25.
- Bakker, A., & van Eerde, H. A. A. (2015). An introduction to design-based research with an example from statistics education. In A. Kikner-Ahsbals, C. Knipping, & N. Presmeg (Eds.), *Approaches to qualitative research in mathematics education: Examples of methodology and methods* (pp. 429-466). New York, NY: Springer.
- Bellon, J. (2000). A research-based justification for debate across the curriculum. *Argumentation and Advocacy, 36*(3), 161-175. doi: 10.1080/00028533.2000.11951646
- Black, B. (2008). *Critical thinking – a definition and taxonomy for Cambridge Assessment: Supporting validity arguments about critical thinking assessments administered by Cambridge Assessment*. Retrieved from <http://www.cambridgeassessment.org.uk/images/126340-critical-thinking-a-definition-and-taxonomy.pdf>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. doi:10.1191/1478088706qp063oa
- Collins, A., & Halverson, R. (2009). *Rethinking education in the age of technology: The digital revolution and schooling in America*. New, York, NY: Teachers College Press.
- Cone, C., Godwin, D., Salazar, K., Bond, R., Thompson, M., & Myers, O. (2016). Incorporation of an explicit critical-thinking curriculum to improve pharmacy students' critical thinking skills. *American Journal of Pharmaceutical Education, 80*(3), Article 41.
- Creswell, J. W., Plano Clark, V. L., Gutmann, M. L., & Hanson, W. E. (2003). Advanced mixed methods research designs. In A. Tashakkori, & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 209–240). Thousand Oaks, CA: Sage
- Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. Oxford, England: Irvington.
- Crowell, A., & Kuhn, D. (2014). Developing dialogic argumentation skills: A 3-year intervention study. *Journal of Cognition & Development, 15*(2), 363-381. doi:10.1080/15248372.2012.725187
- Educational Records Bureau. (2011). *CTP 4 content standards manual*. Retrieved from https://www.lelycee.org/uploaded/Academics/ERB/2011_CTP_Content_Standards.pdf
- Ennis, R. H. (2013). Critical thinking across the disciplines: The wisdom of CTAC program. *Inquiry, 28*(2), 25-45. doi:10.5840/inquiryct20132828
- Ennis, R. H. (2018). Critical thinking across the curriculum: A vision. *Topai, 37*, 165-184. doi:10.1007/s11245-016-9401-4
- Ennis, R. H., Millman, J., & Tomko, T. N. (2005). *Cornell critical thinking tests* (5th ed., Revised). Seaside, CA: Critical Thinking Company.

- Everett, A. C., Anderson, O. S., Wright, M. C., & Fontana, M. (2018). Longitudinal assessment of critical thinking skills across a dental curriculum. *Journal of Dental Education, 82*, 921-928. doi:10.21815/JDE.018.088
- Hsu, P., Van Dyke, M., & Chen, Y. (2015). Examining the effect of teacher guidance on collaborative argumentation in middle level classrooms. *RMLE Online, 38*(9), 1-11. doi: 10.1080/19404476.2015.11641185
- Jen, E., Moon, S., & Samarapungavan, A. (2015). Using design-based research in gifted education. *Gifted Child Quarterly, 59*, 190-200. doi:10.1177/0016986215583871
- Kennedy, R. (2009). The power of in-class debates. *Active Learning in Higher Education, 10*, 225-236, doi:10.1177/1469787409343186
- Kettler, T. (2014). Critical thinking skills among elementary school students: Comparing identified gifted and general education student performance. *Gifted Child Quarterly, 58*, 127-136. doi:10.1177/0016986214522508
- Kettler, T. (2016). A differentiated approach to critical thinking in curriculum design. In T. Kettler (Ed.), *Modern curriculum for gifted and advanced academic students* (pp. 91-110). Waco, TX: Prufrock Press.
- Kline, R. B. (2009). *Becoming a behavioral science researcher: A guide to producing research that matters*. New York, NY: Guilford Press.
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155-163. doi:10.1016/j.jcm.2016.02.012
- Kuhn, D., & Crowell, A. (2011). Dialogic argumentation as a vehicle for developing young adolescents' thinking. *Psychological Science, 22*(4), 545-552. doi: 10.1177/0956797611402512
- Little, R. J. A., & Rubin, D. B. (2014). *Statistical analysis with missing data* (2nd ed.). Hoboken, NJ: John Wiley & Sons.
- Marin, L.M. & Halpern, D.F. (2011). Pedagogy for developing critical thinking in adolescents: Explicit instruction produces greatest gains. *Thinking Skills and Creativity, 6*(1), 1-13. doi:10.1016/j.tsc.2010.08.002
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*, 30-46.
- Murphy, P.K., Rowe, M.L., Ramani, G., & Silverman, R. (2014). Promoting critical-analytic thinking in children and adolescents at home and in school. *Educational Psychology Review, 26*(4), 561-578. doi:10.1007/s10648-014-9281-3
- Oliveras, B., Márquez, C. & Sanmartí, N. (2013). The use of newspaper articles as a tool to develop critical thinking in science classes. *International Journal of Science Education, 35*(6), 885-905. doi:10.1080/09500693.2011.586736
- Olszewski-Kubilius, P., Calvert, E., Steenbergen-Hu, S., Chandler, K., Robins, J., Sanguras, L. Y., Cross, T. L., & Cross, J. (2018, April). *A design-based study on accelerated online ELA curriculum for disadvantaged high potential 7th-8th graders*. Paper presented at the annual meeting of the American Educational Research Association annual, New York, NY.
- Oros, A. L. (2007). Let's debate: Active learning encourages student participation and critical thinking. *Journal of Political Science Education, 3*(3), 293-311. doi:10.1080/15512160701558273

- Partnership for 21st Century Skills. (2004). *Moving education forward*. Retrieved from <http://www.21stCenturySkills.org>
- Plomp, T., & Nieveen, N. (Eds.) (2007). *An introduction to educational design research*. Enschede, the Netherlands: SLO.
- Rapanta, C., Garcia-Mila, M., & Gilabert, S. (2013). What is meant by argumentative competence? An integrative review of methods of analysis and assessment in education. *Review of Educational Research*, 83(4), 483-520. doi:10.3102/0034654313487606
- Salgado, C. M., Azevedo, C., Proença, H., & Vieira, S. M. (2016). Missing data. In MIT Critical Data (Eds.), *Secondary analysis of electronic health records* (pp. 143-162). New York, NY: Springer.
- Scott, S. (2008). Perceptions of students' learning critical thinking through debate in a technology classroom: A case study. *Journal of Technology Studies*, 34(1), 39-44.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002) *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420-428.
- Snow, R. (1989). Aptitude-treatment interaction as a framework for research on individual differences in learning. In P. Ackerman, R. J. Sternberg, & R. Glaser (Eds.), *Learning and individual differences* (pp. 13-59). New York, NY: W. H. Freeman.
- Steckler, A., McLeroy, K. R., Goodman, R. M., Bird, S. T., & McCormick, L. (1992). Toward integrating qualitative and quantitative methods: An introduction. *Health Education Quarterly*, 19(1), 1-8.
- Thomas, D., & Brown, J. S. (2011). *A new culture of learning: Cultivating the imagination for a world of constant change*. Charleston, SC: Authors.
- VanTassel-Baska, J. (2000). Theory and research on curriculum development for the gifted. In K. A. Heller, F. J. Mönks, R. J. Sternberg, & R. F. Subotnik, (Eds.), *The international handbook of giftedness and talent* (2nd ed., pp. 345-365). Kidlington, Oxford, UK: Elsevier Science.
- VanTassel-Baska, J., Bracken, B., Feng, A., & Brown, E. (2009). A longitudinal study of enhancing critical thinking and reading comprehension in title I classrooms. *Journal for the Education of the Gifted*, 33(1), 7-37. doi: 10.1177/016235320903300102
- Weinstock, M., Neuman, Y., & Glassner, A. (2006). Identification of informal reasoning fallacies as a function of epistemological level, grade level, and cognitive ability. *Journal of Educational Psychology*, 98(2), 327-341. doi: 10.1037/0022-0663.89.2.327
- Wesnes, K., & Pincock, C. (2002). Practice effects on cognitive tasks: a major problem? *The Lancet Neurology*, 1(8), 473. doi:10.1016/s1474-4422(02)00236-3