

Spring 2014

Improving Argumentation Through Goal Instructions in Asynchronous Online Discussions

Yekaterina Prudchenko
Old Dominion University

Follow this and additional works at: https://digitalcommons.odu.edu/stemps_etds



Part of the [Adult and Continuing Education Commons](#), [Instructional Media Design Commons](#), and the [Online and Distance Education Commons](#)

Recommended Citation

Prudchenko, Yekaterina. "Improving Argumentation Through Goal Instructions in Asynchronous Online Discussions" (2014). Doctor of Philosophy (PhD), Dissertation, STEM Education & Professional Studies, Old Dominion University, DOI: 10.25777/f0dc-s038
https://digitalcommons.odu.edu/stemps_etds/87

This Dissertation is brought to you for free and open access by the STEM Education & Professional Studies at ODU Digital Commons. It has been accepted for inclusion in STEMPS Theses & Dissertations by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.

**IMPROVING ARGUMENTATION THROUGH GOAL INSTRUCTIONS IN
ASYNCHRONOUS ONLINE DISCUSSIONS**

by

Yekaterina Prudchenko
B.A. August 2005, University of Southern California
M.A. December 2012, Western New Mexico University

**A Dissertation Submitted to the Faculty of Old Dominion University in
Partial Fulfillment of the Requirements for the Degree of**

DOCTOR OF PHILOSOPHY

EDUCATION

OLD DOMINION UNIVERSITY
January 2014

Approved by


Amy Adcock (Director)


Tammie M. Milliken (Member)


Ginger S. Watson-Papelis (Member)

ABSTRACT

IMPROVING ARGUMENTATION THROUGH GOAL INSTRUCTIONS IN ASYNCHRONOUS ONLINE DISCUSSIONS

Yekaterina Prudchenko
Old Dominion University, 2014
Director: Dr. Amy Adcock

Argumentation incorporated into class discussions can improve students' problem solving skills and enhance their epistemic and conceptual understanding. Research indicates students sometimes need scaffolding such as goal instructions to improve their argumentation skills. This study examined the effectiveness of different types of goal instructions on participants' argumentation achievement. In particular, the study compared the effects of minimal, moderate, substantial, and no goal instructions in asynchronous online discussions on participants' argumentation achievement, as measured by development, balance, and explanatory discourse scores. The study also tried to understand participants' experiences of the goal instructions by comparing the differences in emergent themes across goal instructions groups.

Ninety-seven undergraduate students participated in three debates and posted responses to an open-ended qualitative question over a three-week period. The study found significant differences in the balance scores between minimal, moderate, and substantial goal instructions and no goal instructions, indicating that goal instructions are effective in facilitating responses that consider both sides of an issue. In particular, findings suggested that goal instructions with any level of specificity are more effective in creating balance in argumentation than no goal instructions and that minimal goal

instructions are more effective than moderate and substantial goal instructions in encouraging participants to present both sides of an issue. While the study did not find significant differences in explanatory discourse scores, the differences were close enough to significance to suggest that goal instructions did have some positive effect on helping participants consider other people's perspectives in a constructive way and build on each other's ideas.

Quantitative analysis of codes across goal instructions groups revealed participants who received limited instructions focused their discussions on the environment itself while participants who received extended instructions focused their discussions on the impact that debates had on them. Therefore, it is likely that more extended instructions made an impact on encouraging participants to think about their views and consider other people's perspectives.

The study did not find significant differences in development scores or differences in participants' perceptions across goal instructions groups. However, there are indicators that suggest that participants might have dismissed many aspects of moderate and substantial goal instructions, and additional research is needed to confirm these conclusions. Additional research on goal instructions using different methods for evaluating quality of argumentation is also needed to confirm the results of this study.

Keywords: argumentation, goal instructions, asynchronous discussion board, argumentation development, balance, explanatory discourse.

Copyright, 2014, by Yekaterina Prudchenko, All Rights Reserved

ACKNOWLEDGMENTS

It would not have been possible to write this doctoral thesis without the help and support of the kind people around me. I would like to express my gratitude to my advisor Dr. Amy Adcock for her advice, continuous support, motivation and patience. I would also like to thank Dr. Tammi Milliken and Dr. Ginger Watson, members of my PhD committee, without whose assistance and dedication this project would not have been possible. Last but not least, I would like to thank my husband, my parents and Charlie for their personal support and patience at all times. This dissertation is dedicated to my grandmother Galya. For any errors or inadequacies that may remain in this work, of course, the responsibility is entirely my own.

TABLE OF CONTENTS

	Page
LIST OF TABLES.....	vii
CHAPTER I: INTRODUCTION AND LITERATURE REVIEW.....	1
INTRODUCTION.....	1
LITERATURE REVIEW.....	3
PURPOSE OF RESEARCH.....	11
CHAPTER II: METHOD AND ANALYSIS.....	13
METHOD.....	13
CODING AND SCORING.....	20
CHAPTER III: ANALYSIS AND RESULTS.....	32
QUANTITATIVE DATA.....	32
QUALITATIVE DATA.....	38
SUMMARY OF FINDINGS.....	43
CHAPTER VI: DISCUSSION AND CONCLUSIONS.....	46
RESEARCH QUESTIONS.....	46
CONCLUSION.....	51
LIMITATIONS.....	51
FUTURE RESEARCH.....	53
REFERENCES.....	55
APPENDICES.....	65
APPENDIX A: INFORMED CONSENT FORM.....	65
APPENDIX B: DEMOGRAPHIC SURVEY.....	70
APPENDIX C: WEEK 1 TREATMENTS.....	71
APPENDIX D: WEEK 2 TREATMENTS.....	72
APPENDIX E: WEEK 3 TREATMENTS.....	73
APPENDIX F: DEVELOPMENT SCALE.....	74
APPENDIX G: BALANCE SCALE.....	75
APPENDIX H: EXPLANATORY DISCOURSE SCALE.....	76
VITA.....	77

LIST OF TABLES

Table	Page
1. Descriptive Statistics of Participants	15
2. Timeline of Study and Debate Prompts	17
3. Research Questions and Analysis Methods	21
4. Quantitative Data Codebook	23
5. Interpretation of Development Scale	26
6. Interpretation of Balance Scale	27
7. Interpretation of Explanatory Discourse Scale.....	29
8. Qualitative Data Codebook	30
9. MANOVA Descriptive Statistics	33
10. MANOVA Results	34
11. Levene’s Test for Equality of Variances	34
12. Unadjusted ANOVA – Development and Balance	35
13. Adjusted Outcome for Homogeneity of Variance	36
14. Pairwise Comparison Tests – Balance	37
15. Explanatory Discourse Descriptives	38
16. ANOVA – Explanatory Discourse	38
17. Codes by Treatment Group (Type of Goal Instructions)	39
18. Codes Organized into Themes	40

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

Introduction

An argument is a tool that teaches individuals how to think because it requires learners to engage in the deeper epistemological levels of learning (Newton, Driver, & Osborne, 1999) in order to find rational resolutions to questions, issues, and problems (Siegel, 1995). Research has shown that collaborative argumentation, the process of constructing and critiquing arguments with other learners, improves learners' conceptual understanding and problem solving skills by allowing groups of learners to reason at a higher-level collaboratively than each individual would otherwise (Anderson et al., 2001; Chinn, 2006, Golanics & Nussbaum, 2008; Nussbaum, 2002; Pilkington & Walker, 2003; Veerman, 2003; Vygotsky, 1981). However, the process of collaborative argumentation is only conducive for learning if learners can construct strong arguments. Arguments are considered strong if they include evidentiary support, alternative theories, counterarguments, and rebuttals (Kuhn, 1991). Unfortunately, research suggests that many students struggle with various aspects of argumentation (Kuhn, 1991; Means & Voss, 1996). For example, learners rarely qualify their claims or make counterarguments (Kuhn, Shaw, & Felton, 1997; Mandl, Gruber, & Renkl, 1996) and few disagree with their peers' positions (Koschmann, 2003).

A scaffold is an instructional support that constrains the learners' responses, allowing them to perform above their abilities or carry out tasks that might otherwise be too difficult (Collins, Brown, & Newman, 1989; Jonassen, 1999; Woods, Bruner, &

Ross, 1976). Goal instructions are scaffolds that tell students how to complete particular tasks. In educational situations where students are asked to debate a particular topic by proposing arguments, goal instructions can act as a guide for the construction of the argument. By constraining the students' abilities to put forth arguments so that they can perform above their capacities (Jonassen), goal instructions encourage learners to generate better arguments (Collins, Brown, & Newman, 1989; Johnson, 1989; Woods, Bruner, & Ross, 1976). A current review of research into goal instructions suggests that they improve the quality of learners' argumentation (Golanics & Nussbaum, 2008; Nussbaum, 2002; Nussbaum, 2005; Nussbaum & Kardash, 2005). Goal instructions help learners generate more counterclaims and rebuttals (Nussbaum & Kardash, 2005) and have a positive effect on argumentation development and exploration of opposing views (Golanics & Nussbaum, 2008).

In the current movement towards distance education, where much of the coursework involves discussion board interaction, there is a tendency to ignore the design of instruction. This tendency leaves students to learn on their own and vulnerable to not gaining appropriate thinking and reasoning skills (Morrison & Anglin, 2006). Research into goal instructions on asynchronous online discussion boards is still in its infancy, and the relationship between specificity of goal instructions and quality of argumentation is a major unknown. Prior research suggests that the more specific the goal instructions the better the overall quality of argumentation, but this research is limited to only one study with only two levels of specificity (Nussbaum, 2005). Additional research into the specificity with which goal instructions could be designed

was needed because effective designs of these scaffolds may be essential in improving the quality of argumentation. Thus, the study aimed to fill this gap in the research.

The study, conducted using online discussion boards, examined what effect goal instructions with four different degrees of specificity had on participants' argumentation achievement and their perception of the instructions. In particular, it examined the effects of minimal, moderate, substantial, and no goal instructions on participants' argumentation achievement, as measured by development, balance, and explanatory discourse scores. The study also attempted to understand how goal specificity shapes learners' experiences with the activity. Because prior research suggests that the more specific the goal instructions the better the overall quality of argumentation, the researcher anticipated that argumentation quality would increase with goal instruction specificity (Golanics & Nussbaum, 2008; Nussbaum, 2005).

Literature Review

The following review analyses the relevant literature, demonstrates a gap in the research, and provides a rationale for the study. In particular, it provides a brief introduction to argumentation and collaborative argumentation, state the assumptions of the research study, and discuss scaffolding, all relevant studies on goal instructions, and the purpose of the research.

Argumentation

Because argumentation is central to the process of thinking and reasoning, and to the development of conceptual understanding (Chin & Osborne, 2010), it is believed to play an important role in students' learning in both science and humanities.

Argumentation gives individuals tools that teach them how to think and these tools, in

turn, facilitate meaningful, deep learning. Argumentation is the way that individuals resolve questions and disputes. It requires learners to argue the basis on which claims are made and engage in the deeper epistemological levels of learning (Newton, Driver, & Osborne, 1999) in order to find rational resolutions to questions, issues, and problems (Siegel, 1995). Argumentation also modifies learners' underlying beliefs and allows others to identify and refute misconceptions (Baker, 1999).

Students learn through argumentation because it facilitates conceptual change (Asterhan & Schwarz, 2007; Baker, 1999; Nussbaum & Sinatra, 2003; Wiley & Voss, 1999). Conceptual change is the process of altering learners' understanding of a particular topic in order to accommodate new perspectives and reorganize the conceptual framework that encompassed them. For example, a debate club requires students to consider other people's arguments and positions, altering their understanding of a particular topic and reorganizing their conceptual frameworks. As a result, learning environments that incorporate argumentation enhance learners' conceptual and epistemic understanding of a particular topic (Duschl & Osborne, 2002; Wiley & Voss, 1999). The overall quality of argumentation, or argumentation achievement, is evaluated using three scales or outcome measures: how well arguments are developed (development); how well arguments present both sides of an issue (balance); and how well participants consider other people's perspectives (explanatory discourse) (Golanics & Nussbaum, 2008; Nussbaum & Kardash, 2005).

Collaborative Argumentation. Collaborative argumentation is the process of working together to construct and critique arguments (Anderson et al., 2001; Golanics & Nussbaum, 2008; Nussbaum, 2002; Pilkington & Walker, 2003; Veerman, 2003). By

requiring students to collaborate and share knowledge, collaborative argumentation improves their conceptual understanding and problem solving skills (Chinn, 2006) and allows a group of learners to reason at a higher-level together than each individual would otherwise (Vygotsky, 1981). Furthermore, it allows learners to construct and reconstruct their views on a particular topic by engaging in cognitive conflict and applying domain specific knowledge to resolve the conflict (Bell, Grossen, & Perret-Clermont, 1985; Johnson & Johnson, 2004; Koschmann, 2003; Slavin, 1996).

Strong Arguments. The process of argumentation is only effective when learners construct strong arguments. An argument is considered strong if it offers evidence to support theories, generates alternative theories, makes counterarguments, and rebuts alternative theories (Kuhn, 1991). Counter-argumentation is an essential factor in developing a strong argument because it requires learners to look beyond evidence that only supports their positions and re-examine their ways of thinking (Leitau, 2000). Therefore, a strong counterargument is a critical component of a strong argument and the ability to construct arguments and counterarguments is an essential skill for effective collaborative argumentation (Andriessen, Baker, & Suthers, 2003; Erkens, Andriessen, & Peters, 2003; Jarvela & Hakkinen, 2003; Leitao, 2000; Nussbaum & Schraw, 2007).

Unfortunately, research indicates that students struggle with many aspects of argumentation (Kuhn, 1991; Means & Voss, 1996). Research on collaborative argumentation shows that many learners rarely qualify their claims or make counterarguments (Kuhn, Shaw, & Felton, 1997; Mandl, Gruber, & Renkl, 1996). Therefore, students are not giving their peers reasons to consider their positions or

reasons why they should reconsider their own. Studies have also found that learners rarely disagree with their peers and instead merely restate their own positions (Andriessen, 2006; Koschmann, 2003). Furthermore, learners also rarely challenge other people's claims and rarely respond to other people's challenges (Baker, 2003; Cho & Jonassen, 2002; Jeong & Joung, 2007; Oh & Jonassen, 2007). Collaborative argumentation in asynchronous environments allows students the time and space to properly reflect on their peers' positions, internalize their arguments, and possibly reconsider their own positions, but many students do not take advantage of these opportunities.

Cognitive Conflict and Development. One of the assumptions of this research study is that conflict and consideration of alternative perspectives are essential ingredients for facilitating inquiry, reflection, and deeper understanding (Johnson & Johnson, 1987; Wiley & Voss, 1999). Conflict and the subsequent inquiry, the essential elements of argumentation, are the result of the interrelationship between ideas and the responses they generate.

Another assumption of this research is that collaborative argumentation, in which learners examine and evaluate alternative perspectives, facilitates conceptual change (Dole & Sinatra, 1998; Schwarz, Neuman, & Biezuner, 2000; Baker, 2003) or deepens their understanding of a subject matter (Alexopolou & Driver, 1996; Bell & Linn, 2000). Cognitive disequilibrium is the result when learners encounter information that generates a contradiction within their existing cognitive structures (Piaget, 1977). To resolve this conflict, they work to accommodate and assimilate new information. These processes require individuals to construct new schema, which facilitates the individuals'

cognitive development. As a result, collaborative argumentation helps learners make connections between ideas, reflect on meaning, and consider and possibly adopt alternative viewpoints.

The final assumption is that collaborative argumentation, the process of considering challenges, counterchallenges and others' positions, is more productive than adversarial argumentation, the process of persuading others of one particular and predetermined point of view (Mercer, 1996; Golanics & Nussbaum, 2008). One way that designers can influence the process of argumentation and help facilitate collaborative rather than adversarial argumentation in online discussions is by incorporating scaffolds into the discussion prompts. These scaffolds take the form goal instructions.

Scaffolding

Learners can be encouraged to make better arguments using scaffolding, a type of guidance that helps learners carry out tasks that might otherwise be too difficult (Collins, Brown, & Newman, 1989; Woods, Bruner, & Ross, 1976). Scaffolds are instructional supports that constrain learners' arguments so that they can perform above their capacities (Jonassen, 1999). As a result, learners are able to achieve higher levels of understanding than they could otherwise (Jonassen & Kim, 2010). A type of scaffolding, which is particularly appropriate for asynchronous learning environments, is goal instructions.

Goal Instructions

Goal instructions, placed at the end of a discussion prompt, are short statements that tell students how to complete a particular task. Goal instructions may be effective scaffolds because they activate collaborative argumentation schema that help learners

identify patterns of locating and correcting problems within the argument (Nussbaum, 2002).

Goal Instructions and Persuasive Essays. The effects of goal instructions on writing were first studied in a traditional classroom environment (Ferretti, MacArthur, & Dowdy, 2000). The study examined the effects of elaborated goal conditions on the quality of persuasive essays about controversial topics. Participants, fourth and sixth grade students, were asked to write a letter that persuaded others to agree with their position. The study found that sixth-grade students who were provided with the same general goal and explicit sub-goals based on the elements of argumentative discourse (elaborated goal condition) produced more persuasive essays and included a greater number of argumentative elements in their essays than both sixth-grade students in the general goal condition and fourth-grade students in both goal conditions.

Goal Instructions and Argumentative Essays. Another study investigated approaches that encourage undergraduates to consider more counterarguments when writing argumentative essays on TV violence (Nussbaum & Kardash, 2005). Researchers conducted two experiments and provided students with directions for different kinds of essays. The first experiment found that participants who received specific goal instructions generated more counterarguments and rebuttals than the control group. The second experiment by Nussbaum & Kardash (2005) focused on the purpose of constructing arguments and researchers examined the concept of persuasion. Findings in the second experiment showed that instructions asking students to persuade had a negative effect on both the quality of essays as a whole and on the number of reasons that students provided to support their counter arguments. In other words,

students actually believed that identifying counter arguments made their own arguments less persuasive. The study also found that the text outline, though only effective for students with less extreme prior attitudes about the topic, counteracted the negative effects of persuasion instructions and increased the overall quality of argumentation. Furthermore, the study also found that students who were instructed to produce counterclaims (reasons why others may disagree) and rebuttals (reasons why those reasons are wrong) generated more counterclaims, rebuttals, and reasons that supported their rebuttals (Nussbaum & Kardash, 2005). This finding indicated that specific goal instructions generated more counterarguments and rebuttals than general goal instructions and, as a result, the authors concluded that specific goal instructions are more effective in facilitating better argumentation than general goal instructions. These findings are consistent with other research that found that setting specific short-term goals facilitated better writing (Page-Voth & Graham 1999). Since participants rarely provide counterclaims, counterclaims were not evaluated for quality. Instead, the mere existence of a counterclaim or an alternative argument was indicative of better argumentation.

Goal Instructions of Two Levels of Specificity. One of the first studies to examine the effectiveness of goal instructions on students' reasoning and argumentation in asynchronous online discussions also asked students to argue about TV violence (Nussbaum, 2005). Undergraduate students were separated into three conditions that varied according to the kind of goal instructions students were presented with in the discussion prompt. One group was given general goal instructions (to persuade one another or to explore an issue) and another group was given specific goal instructions (to

generate as many reasons as possible or to generate counterarguments and rebuttals). Students in the third group did not receive any goal instructions. The study also controlled for the need for cognition, a measure of the students' disposition to think, and found that it predicted total argument claims and depth.

The study found that the general goal instruction 'to persuade' produced elaborated but more adversarial and somewhat better supported arguments and that the specific goal instruction 'to generate as many reasons as possible' which produced the most deep and contingent arguments (Nussbaum, 2005). The study also found that the goal 'to generate as many reasons as possible' resulted in balanced discourse in which both sides of an issue were evaluated and explored almost evenly. This result was a significant finding because it suggested a method for fostering collaborative argumentation without teaching students the rules of good argumentation.

The study also found that the general goal 'to explore' and the specific goal 'to generate counterarguments and rebuttals' were ineffective (Nussbaum, 2005). While the previous study found that specific goal instructions that asked for counter arguments were an effective tool in argumentative essay writing (Nussbaum & Kardash 2005), this study suggested that the finding was as a result of awkward goal instructions. In particular, the goal instructions directed students to generate their counter arguments right after proposing their arguments and the author suggested that this direction may have been difficult to implement in conversation.

Goal Instructions and Elaborated Questions. Since a previous study found that the goal instruction 'generate as many reasons as possible' resulted in more balanced argumentation in an online environment (Nussbaum, 2005), another study was

conducted to try to replicate this finding (Golanics & Nussbaum, 2008). In addition, this study also examined the effects of elaborating on possible lines of reasoning within the question prompt, and the role of prior attitudes, knowledge and interest. The study asked learners to construct an argument about wearing uniforms in public schools and crossed the goal instruction to generate as many reasons as possible (goal/no goal) with an elaborated question (elaborated/unelaborated question) in a 2 x 2 randomized design. Elaborated questions are prompts that briefly mention arguments on both sides of the issue, helping learners generate connections among ideas and between ideas and the learners' prior knowledge (Wittrock, 1992). The study also randomly assigned half of the students to complete a preliminary attitude survey.

The study found that the goal instruction positively affected argument development and exploration of opposing views for high-issue knowledge students, when prior knowledge was controlled (Golanics & Nussbaum, 2008). The study also found that high issue knowledge students benefited from the elaborated question which asked students to generate as many reasons as possible. In particular, high issue knowledge students produced more explanatory discourse when they were presented with the goal instruction condition and more balanced and better developed individual arguments when they were presented with both the goal instruction condition and the elaborating question condition.

Purpose of Research

Gaps exist in our understanding of how goal instructions should be designed to facilitate learning. The purpose of this mixed-methods research study on online discussion boards was to examine what effect different types of goal instructions had on

participants' argumentation achievement and to understand the impact of participants' experiences of the goal instructions. The aim of this research was to contribute to better designs of online discussion prompts, ones which are more effective at enhancing participants' conceptual and epistemic understanding of a given topic.

Research Questions

One central research question guided this study: What is the effect of goal instructions with different degrees of specificity on learners' argumentation achievement in online settings? More specifically, the study considered the following research questions:

(a) What are the effects of minimal, moderate, substantial and no goal instructions on participants' development scores?

(b) What are the effects of minimal, moderate, substantial and no goal instructions on participants' balance scores?

(c) What are the effects of minimal, moderate, substantial and no goal instructions on participants' explanatory discourse scores?

(d) What are the participants' experiences of minimal, moderate, substantial, and no goal instructions?

CHAPTER II

METHOD AND ANALYSIS

Method

Participants

Participants were recruited from six undergraduate human services courses in offered by a mid-sized mid-Atlantic university during the summer 2013 semester. These participants were chosen purposefully because they could best and most broadly inform the questions studied (Creswell, 2009). All courses were at the upper-division level and offered in an online format, consisting of 2 sections of HMSV368: *Field Observation Human Services*, HMSV447: *Addictions: Theory and Intervention*, HMSV491: *Family Guidance*, HMSV441: *Nonprofit Fund Raising in Human Services*, and HMSV341: *Introduction to Human Services*. The Human Services and Counseling program prepares students to do a wide variety of community services such as helping others to cope with social, personal, and environmental pressures. Participants from these courses were recruited because courses in human services often address controversial topics, similar to those brought up in the debate topics, and are taught online, an educational format which the study focuses on. Finally, these participants were also specifically recruited because, as undergraduates enrolled in upper-division courses, they have significant experience in terms of an educational background, but unlike graduate students, are not yet narrowly focused on one particular subject or area of study (T. Milliken, personal communication, February 26, 2013).

An a priori power analysis was performed using the statistical program A-priori Sample Size Calculator for Multiple Regression (Soper, 2013). It indicated that based on an alpha value of 0.05, a medium effect size of 0.15 yielded a recommended sample size of 84 participants (Cohen, 1988). A total of 97 students ($N = 97$) participated in the study in exchange for extra credit. Their participation was voluntary and instructors provided alternative assignment options to students who wanted to receive extra credit without participating in the study.

Demographic information about all participants was collected using a survey and provided the researcher with information regarding the participants' gender, academic standing, age range, and major (see Table 1). The study participants were 86.50% female and 13.50% male and 65.26% and 33.68% of them were college seniors and juniors, respectively. A total of 47.42% of participants were between the ages 20 – 24, 17.53% between the ages 25 – 29, 13.40% between the ages 30 – 39, 14.43% between the ages 40 – 49, and 7.22% between the ages 50 – 59. Finally, the majority of the study's participants, 80.41%, were human services majors.

Research Design

Participants in each course belonged to the same treatment group, i.e. the groups were intact, and each course was assigned to one of four treatment groups: no goal instructions (control), minimal, moderate, and substantial goal instructions. Minimal and no goal instructions groups each had one course of participants while moderate and substantial goal instructions groups each had two courses of participants because they had fewer total enrolled students. Students participated in three debates on their course's discussion board, each one lasting a week. Participants were instructed to post one

original response and at least two replies to their peers' posts. At the conclusion of the study, participants posted replies to the open-ended qualitative question about the discussion activity in a separate discussion board thread.

Table 1

Descriptive Statistics of Participants

Characteristic	Number	Percent (%)
Total Number of Participants	97	100
Gender		
Female	83	86.50
Male	13	13.50
Skipped	1	1.03
Academic Year		
Freshman	0	0.00
Sophomore	1	1.05
Junior	32	33.68
Senior	62	65.26
Skipped	2	2.06
Age Range		
20-24	46	47.42
25-29	17	17.53
30-39	13	13.40
40-49	14	14.43
50-59	7	7.22
Major		
Human Services	78	80.41
Psychology	6	6.19
Communications	3	3.09
Psychology/ Human Services (minor)	2	2.06
Other majors ^a	7	7.21

Note. ^a health services, therapeutic recreation and human services, communications and human services, high school counseling, occupational and technical studies, criminal justice, and sociology.

Two raters coded and scored each group's responses, blind to the condition, according to scales developed in an earlier study, measuring three different aspects of argumentation quality: development, balance, and explanatory discourse (Golanics & Nussbaum, 2008). The study's dependent variables consisted of the development, balance, and explanatory discourse scores as well as the emergent themes, which came from the qualitative analysis of open-ended question's responses.

Debate Prompts

The four intact groups received the same three prompts over a three-week period (see Table 2): *Should hospitals be mandated to provide birth control? Should doctor-assisted suicides be legal or illegal for terminally-ill patients? Should recreational use of marijuana be legalized?* These questions were generated by talking to one of the researcher's committee members, also a professor in human services, who suggested that the chosen debate questions should address controversial topics in the field, generate debate and arguments, but not be content specific (T. Milliken, personal communication, February 13, 2013).

The day after the deadline for submitting responses to the third debate, all participants received a prompt, asking them to reflect on the discussion activity (see Table 2): *In no more than a paragraph, please describe your experience of the discussion activity.* This question was generated after a discussion with the researcher's committee member who stated that responses from the qualitative question will triangulate the quantitative data results, providing additional information regarding themes and patterns emerging in each group, and across groups (T. Milliken, personal communication, June 24, 2013). Answers to this question also served as a member

check, giving the researcher the opportunity to get an understanding of the participants' experience of the activity (Creswell, 2009; Merriam, 1998).

Table 2

Timeline of Study and Debate Prompts

Time	Prompts
Week 1: Mon – Fri	Should hospitals be mandated to provide birth control?
Week 2: Mon – Fri	Should doctor assisted suicides be legal or illegal for terminally-ill patients?
Week 3: Mon – Fri	Should recreational use of marijuana be legalized?
Week 3: Fri – Sun	In no more than a paragraph, please describe your experience of the discussion activity.

Treatment Groups

The study had one independent variable (goal instructions) with four levels (none, minimal, moderate and substantial). The exact instructions that participants received for each week of the study are found in Appendix C, D, and E.

No Goal Instructions. This group received only the debate prompt.

Minimal Goal Instructions. This group received the debate prompt and instructions asking them to provide reasons for their positions.

Moderate Goal Instructions. This group received the debate prompt and instructions asking them to provide reasons and evidence for their positions.

Substantial Goal Instructions. This group received the debate prompt and instructions asking them to provide reasons, evidence and assumptions for their positions.

Instruments

Each group's replies were scored using three scales: development, balance, and explanatory discourse (Golanics & Nussbaum, 2008; Nussbaum & Kardash, 2005). Replies from week one were used to conduct rater training and to estimate inter-rater reliability ($\alpha = 0.90$). Two researchers coded discussion postings from weeks two and three, blind to the condition, and assigned scores according to the following scales.

Development. Development, scored on a six-point scale, assessed how well arguments were developed and gave higher scores for originality and evidentiary support (Golanics & Nussbaum, 2008). The development scale is found in Appendix F.

Balance. Balance, scored on a five-point scale, assessed how well arguments presented both sides of an issue (Golanics & Nussbaum, 2008). The balance scale is found in Appendix G.

Explanatory Discourse. Explanatory discourse, scored on a five-point scale, assessed how well participants interacted with one another and considered each member's perspectives in a constructive way (Golanics & Nussbaum, 2008). Because explanatory discourse examines the interaction within the group, one overall score was given to each group. The explanatory discourse scale is found in Appendix H.

These three scales, development, balance and explanatory discourse, were developed and used in two previous studies (Golanics & Nussbaum, 2008; Nussbaum & Kardash, 2005). They are based on the standard model for analyzing arguments, which

examines how claims are used to construct arguments, counter-arguments and rebuttals (Beardsely, 1950; Golanics & Nussbaum, 2008; Inch & Warnick, 2002).

Procedure

The researcher conducted all interactions between her and the participants via email. All students enrolled in the six human services courses in summer 2013 were contacted by email, through Blackboard Learn™, a learning management system, with information about the study and an invitation to participate. Prior to the beginning of the study and during the first week, all participants completed an informed consent IRB release form and demographics survey (see Appendix A and B).

Each class of participants was assigned to one of four treatment groups: no goal instructions (control), minimal, moderate, and substantial goal instructions. Minimal and no goal instructions groups each had one course of participants while moderate and substantial goal instructions groups each had two courses of participants.

The three-week study was conducted entirely using each course's asynchronous threaded discussion board on Blackboard Learn™. Each week's debate opened, at 12:00 a.m. Monday, and the researcher notified all students in each course of its opening using Blackboard's course emailing system. Participants were instructed to post one original response and at least two replies to their peers' posts each week. All participants were given tentative guidelines to post their original replies by Tuesday at 5:00 p.m. and post at least two replies to other people's posts by Friday at 5:00 p.m. These suggested deadlines were used mainly to encourage participants to post their original replies earlier in the week.

Participants participated in a total of three debates on their course's discussion board, each one lasting Monday through Friday. During the study, the researcher also sent out emails to all students on Wednesdays, to remind them to post their replies. On the third Wednesday, the reminder email also contained information about the upcoming post-study question. On Friday, after the debate for week three ended, participants were asked to post their replies to the open-ended qualitative question. The post-study question's discussion thread opened on Friday at 5:00 p.m. and participants were asked to post their replies by Sunday 11:59 p.m. Discussion boards remained opened for all three weeks.

Coding and Scoring

All responses were coded and scored by two raters. There were four participants who participated in more than one class' debates. Their responses were eliminated prior to coding and scoring. The quantitative aspect of the study had one independent variable (goal instructions) with four levels (no, minimal, moderate and substantial) and three dependent variables (development, balance, explanatory discourse). Qualitative data consisted of identifying emerging themes from the qualitative question responses and determining whether there were any differences in themes across the different goal instruction groups. Table 3 presents an overview of how the data was coded and scored.

Coding. Dedoose, a web-based quantitative and qualitative data analysis software, was used to develop the codebook, conduct rater training, find the inter-rater reliability score, and code the responses. To achieve the appropriate level of reliability, the researcher enlisted the help of a volunteer rater. Together, the team of two raters

segmented the text, created a codebook, coded the text, assessed the inter-rater reliability, modified the codebook, and coded the responses.

Table 3

Research Questions and Analysis Methods

Research Question	Coding/Scoring
What are the effects of minimal, moderate, substantial and no <i>goal instructions</i> on participants' <i>development</i> scores?	Posts (week 2, 3) coded ^a and scored using 6-point development scale ^b . Data analyzed using MANOVA, one-way ANOVA.
What are the effects of minimal, moderate, substantial and no <i>goal instructions</i> on participants' <i>balance</i> scores?	Posts (week 2, 3) coded ^a and scored using 5-point balance scale ^b . Data analyzed using MANOVA, one-way ANOVA.
What are the effects of minimal, moderate, substantial and no <i>goal instructions</i> on participants' <i>explanatory discourse</i> scores?	Posts (week 2, 3) coded ^a and scored using 5-point explanatory discourse scale ^b . Data analyzed using MANOVA, one-way ANOVA.
What are the emergent themes of minimal, moderate, substantial and no <i>goal instructions</i> ?	Posts (qualitative question) coded according themes. Data triangulated with quantitative data results.

Note. ^a by two raters, blind to the condition, whose inter-rater reliability was kappa = 0.94

^b developed by Golanics and Nussbaum (2008)

Text consisted of all replies posted on each class' discussion board, each week of the study. Text was segmented into individual posts made in reply to the debate questions. To generate the first draft of the codebook, team members focused on

responses from the first week of the study, a portion of the data, which was representative of the whole.

In particular, team members worked independently to examine responses to week one debate question and put forth a set of themes such as “argument,” “reasons,” “counterargument,” and “conclusion,” which focused on specific aspects of a quality argument. Using these themes, the raters came up with a list of eleven codes, taking into account the relevance of the codes to the goals of the study (Hruschka et al., 2004). Codes were separated into three main categories: argument, counterargument, and conclusion. Argument and counterargument categories each had five codes while the conclusion category had only one code (see Table 4). All codes emerged from qualities of good argumentation and were similar to the goal instructions in the fourth treatment group.

After developing the list of codes, the raters then came up with a list of rules, which they used to decide whether a particular discussion board post had or did not have instances of a specific code (Hruschka et al., 2004). In general, the team agreed that the discussion board replies were going to be evaluated as a whole and individual codes were going to be assigned to individual sentences or parts of sentences within the reply. A particular post could have a number of different reasons within one post, and if posts had more than one reason then it was up to the rater to count the number of reasons and include that in the coding. The team also agreed that posts which did not contain reasons or evidence for why others might disagree would not receive any codes under the “counterargument” category and posts which did not have some sort of conclusion at the end of the post or at the very least a restatement of the participant’s statement of opinion

at the end of the post would not receive a “conclusion” code. Furthermore, to get the “statement of opinion” code, the text would have to explicitly identify the participants’ opinion, i.e. “assisted suicide should not be allowed...”

Table 4

Quantitative Data Codebook

Category	Codes
Argument	Statement of opinion Reasons to justify your position Evidence that supports your reasons Assumptions you are making Implications of these assumptions
Counterargument	Reasons why others might disagree with you Evidence that supports these reasons Assumptions that these people are making Implications of their assumptions Why their reasons are wrong
Conclusion	Conclusion

After developing the initial draft of the codebook, the team began the process of coding, reliability assessment, codebook modification, and recoding (Hruschka et al., 2004). Raters took three random postings from each treatment group and applied the codebook collaboratively. After the team reached an understanding as to how to apply the codes, they proceeded to the first coding round. The lead coder, the researcher,

distributed a set of three postings, chosen randomly from the week one responses, and the team coded the responses independently according to the first draft of the codebook.

To assess the degree to which a set of texts were consistently coded by different coders, the researcher calculated inter-coder reliability using κ , Cohen's kappa (Cohen, 1960). κ was chosen because it corrects for chance agreement between the coders by preventing the inflation of reliability scores (Hruschka et al., 2004). Research suggests that the criteria for identifying almost perfect or excellent agreement should have κ of 0.81- 1.00 (Landis & Koch, 1977) and 0.75-1.00 (Cicchetti, 1994). To ensure a very high level of reliability, the researcher chose a strict cutoff of $\kappa \geq 0.90$ (Miles & Huberman, 1994).

The inter-coder reliability for the first coding round had $\kappa = 0.75$. Because this κ was below the criteria set, the inter-coder reliability was found to be insufficient and the team went back and discussed the codes in more detail to modify the codebook and clarify the codes. One of the issues found was that the researcher was coding the same sentence with multiple codes while the other rater was generally assigning only one code per sentence. Further clarification was also needed regarding what was coded "reasons" and what was coded as "evidence." The team decided that, in explaining why participants held particular opinions, sentences which contained phrases such as "I think," "I feel," "in my opinion," and statements along those lines would be coded as "reasons to justify your position." The "evidence that supports your reasons" code would only be assigned to statements which specifically identified evidence for the participants' beliefs, regardless of whether it is scientifically appropriate evidence. For example, many participants cited "miracles" as evidence for why they thought assisted

suicide should not be made legal. Because their reasons for holding a particular opinion were based on a belief in God, mentions of miracles were coded as “evidence.”

After the codebook was modified accordingly, the coding process was repeated again. This time, the team coded two sets of six randomly chosen questions each, conducting no training or code modification in between tests. The researcher assessed inter-rater reliability and found that the first and second set had $\kappa = 0.88$ and 1.0 , respectively. Because κ for the second set and the average of κ for the first and second set are both above the acceptable inter-rater reliability threshold of $\kappa = 0.90$, training was concluded. Once sufficient inter-coder reliability was achieved, the entire set of responses was coded following the final draft of the codebook. The researcher split up the responses from weeks two and three, with each coder doing every other question, blinded to the condition. The coding was completed within three days so the systematic inter-coder reliability checks did not need to be conducted throughout the coding process.

Scoring. Once all the posts for weeks two and three were coded, the team worked together to assign development and balance scores to each post, including the participants’ original post and the replies. The team also worked together to assign explanatory discourse scores. One explanatory discourse score was assigned to each group for each week and these scores excluded all original posts from analysis.

Development scale. The development scale (see Appendix F) takes into account three aspects: lines of argumentation, originality and evidence (Golanics & Nussbaum, 2008). The team focused mainly on lines of argument and evidence (counting both “reasons” and “evidence” codes as evidence) and excluded originality. Originality was

not an important factor in these debates because controversial topics were discussed and the focus of the research is on developing complete arguments rather than coming up with novel information. As a result, the discussion board postings were not restricted in viewing and all classmates could see each other's postings at all times.

The development scale was interpreted in the following way. The team divided the development scale into four categories: lines of argumentation, statement of opinion, reasons and evidence (see Table 5). The assignment of scores was primarily based on the number of reasons and pieces of evidence that a particular text contained. To assign the scores, the team examined the coding and counted up the reasons and evidence within the post. To receive a score of six points, the post had to have five to six lines of argumentation, four reasons, three reasons and one piece of evidence or two reasons and two pieces of evidence. To receive a score of five points, the post had to have five lines of argumentation, three reasons or two reasons and one piece of evidence. Because the development scale is a measure of development, a post had to have at least one reason to get a score of more than three points.

Table 5

Interpretation of Development Scale

	6 points	5 points	4 points	3 points	2 points	1 points
Lines of argumentation	6 – 5	5	4	3	2	1
Statement of opinion	1	1	1	1	1	-
Reasons	2, 3, or 4	2 or 3	1 or 2	1	-	-
Evidence	2, 1, or 0	1 or 0	1 or 0	-	-	-

Balance scale. The balance scale (see Appendix G) evaluates how well a text considers opposing views. The team interpreted and applied the balance scale in the following way (see Table 6). Posts which received five points contained two reasons for a particular opinion (a code under the argument category), one reason as to why others would say that opinion was wrong (a code under the counterargument category), a statement of why that reason was wrong, and a solution. To receive a balance score of four or higher, a post had to include a solution, as well as one reason for and one reason against a particular opinion. To get a balance score of a three, a post had to include one reason for, one reasons against and either why the reason against a particular opinion was wrong or a solution. Posts without codes under the counterargument category received a score of one out of five.

Table 6

Interpretation of Balance Scale

	5 points	4 points	3 points	2 points	1 point
Reasons for	2	1	1 or 1	1	1
Reasons against	1	1	1 or 1	1	-
Why those reasons are wrong	1	1	0 or 1	-	-
Solution	1	1	1 or 0	-	-

Explanatory discourse scale. The explanatory discourse scale (see Appendix H) is a scale that evaluates the interaction of the group as a whole. Because it is a score for

interaction, one overall score was assigned to each group. Unlike development and balance scores, which assigned scores to each post regardless of whether it was an original post or a reply, the team assigned explanatory discourse scores to only reply posts (replies that participants posted to other participants' posts). Prior to assigning explanatory discourse scores as a team, the researcher first went through all the postings and excluded all original posts from the scoring. These posts were excluded because the scale explicitly focuses on participants' replies.

In order to assign an appropriate explanatory discourse score to the group, the team went through the posts and assigned scores to individual posts. The explanatory discourse scale is a measure of whether members' posts are explanatory (critical, but flexible and willing to concede) and to what degree the group as a whole is explanatory (Golanics & Nussbaum, 2008). According to the scale, posts might also be cumulative (all agree/ built on each other's ideas) or disputational (all opposed each other's ideas).

The researcher applied the scale in the following way. Each post was assigned an (E), (C) or (D), depending on whether it was explanatory (E), cumulative (C) or disputational (D). Posts which were critical yet flexible and/or showed that the participant was listening to what others' were saying and willing to concede were given an (E), which was worth one point. Posts which only opposed others' ideas or only agreed with others were scored a (C) or a (D), respectively. Both (C)'s and (D)'s were worth zero points because they did not demonstrate explanatory discourse. The number of each group's (E) scores were then counted up and divided by the total number of possible scores, allowing the team to find out what percentage of the posts within each group were explanatory.

To assign each group an overall score for explanatory discourse, the team then applied the explanatory discourse scale. The scale says that a score of four means that two thirds of the members of a group are explanatory. Therefore, if a group had a score around .667 for a particular week, meaning 66.7% of its members' posts were explanatory, then the group was assigned a score of four for explanatory discourse. To get a five, the group had to have an explanatory percentage of 83.35 or above. To get a score of three, the group had to have explanatory percentage of around 33.33 (see Table 7). Unlike development and balance scores, explanatory discourse scores were assigned for each week, one for week two and one for week three. Thus, each group had two explanatory discourse scores. These scores were then averaged to get an overall explanatory discourse score for each group.

Table 7

Interpretation of Explanatory Discourse Scale

	5 points	4 points	3 points	2 points	1 point
% of explanatory posts within each group	100	66.67	33.33	16.75	n/a
(% range)	(83.35 – 100)	(50.05 – 83.34)	(16.75 – 50.04)	n/a	n/a

Qualitative Data Coding

Qualitative analysis was performed on all responses to the qualitative question, *in no more than a paragraph, please describe your experience of the activity*, which participants received after they participated in three debates. All responses were included in the coding and analysis of emergent themes except responses posted by four

participants who were enrolled in multiple classes and participated in the study multiple times.

Coding. After excluding the repeat participants' responses, the researcher organized all discussion board responses to the qualitative question into a chart and went through each post individually to pick out the relevant concepts and themes. The researcher used the chart to identify recurring ideas, language and patterns of belief that connected people and settings (Marshall & Rossman, 1995). Examining the responses to the qualitative question, the researcher found that many responses contained similar phrases, observations and ideas and made a list of these items. Furthermore, the researcher also kept track of the unexpected observations and ideas that the respondents posted and included these in her notes.

After examining the list of identified phrases and observations, the researcher organized these items into fourteen codes, also known as categories (see Table 8).

Table 8

Qualitative Data Codebook

Codes
Open environment to express opinion
Did not feel environment was open
Topics were touchy/controversial
Liked that the topics were controversial
Interesting to see others' perspectives
Topics made me think about my views
Discussions allowed me to learn from others
Allowed me to consider other people's perspective
Allowed me to consider why others might think my views were wrong
Gave me opportunity to discuss issues that are not often mentioned in class
Discussions similar to class discussions
Encouraged me to do some research
Wish more people participated
Enjoy opportunity for extra credit

Examples of codes are “interesting to see others’ perspectives,” “topics were touchy/controversial,” “topics made me think about my views,” and “did not feel the environment was open.” Later in the analysis, the researcher organized the codes according to treatment group, counted instances of each code, and came up with emergent themes.

CHAPTER III

ANALYSIS AND RESULTS

Quantitative Data

Quality of argumentation, using development and balance scores, was measured in four groups of goal instructions: minimal, moderate, substantial and no goal instructions. SPSS statistical software was used to analyze the data. Data was analyzed using one-way MANOVA with four-levels of independent variables and two dependent variables (development and balance) in the analysis. MANOVA is a significance test of group differences, which reduces the experimental-wise level of Type 1 error and takes into account the inter-correlations among the dependent variables (Hinkle, Wiersma, & Jurs, 2003). The last dependent variable, explanatory discourse, was not included in this analysis because there was only one score assigned to each group. As a result, this data was analyzed using one-way ANOVA, a special case of multiple regression, which focuses on differences across groups rather than on the prediction of one variable (Cohen, 1968; Keith, 2006).

MANOVA. Initial statistics (see Table 9) suggested that participants in different goal instructions groups have similar development and balance scores. One-way MANOVA statistics, found in Table 10, represents a calculation for multivariate significance. Because the study had more than two treatment groups, Wilks' Lambda (λ) outcome (Mayers, 2013) was used. Results indicated that there were statistically significant differences in development and balance scores across different types of goal instructions, $F(6, 540) = 3.207, p = .004 (p < 0.05)$; Wilk's $\lambda = 0.932$, partial $\eta^2 = 0.034$.

Table 9

MANOVA Descriptive Statistics

	Goal Instructions	Mean	Std. Deviation	N
Development	No	3.4328	1.58806	67
	Minimal	3.8776	1.88892	49
	Moderate	4.0122	1.88885	82
	Substantial	3.5455	2.06821	77
	Total	3.7164	1.87947	275
Balance	No	1.4478	1.13195	67
	Minimal	2.5510	1.83781	49
	Moderate	2.1220	1.65843	82
	Substantial	2.1299	1.74982	77
	Total	2.0364	1.64299	275

Homogeneity of Variance. Prior to moving on to conducting post hoc tests, the researcher checked the homogeneity of variance assumption to make sure that that the findings of one-way MANOVA analysis were valid. In particular, she examined the results of the Levene's test, which tests the null hypothesis that the error variance of the dependent variable is equal across groups (see Table 11). Because Sig. values (p) for both development and balance scores were less than alpha of .05 ($p < .05$), the researcher rejected the null hypothesis and concluded that there are significant differences in between-group variances.

ANOVA. Due to problems with homogeneity of variance for both development and balance scores, the researcher conducted independent one-way ANOVA analyses with Brown-Forsythe and Welch's F adjustments. The researcher chose to use the Welch statistic because it is more conservative and powerful than the Brown-Forsythe statistic (Mayers, 2013).

Table 10

MANOVA Results

	Effect	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.794	521.115 ^a	2.000	270.000	.000	.794
	Wilks' Lambda	.206	521.115 ^a	2.000	270.000	.000	.794
	Hotelling's Trace	3.860	521.115 ^a	2.000	270.000	.000	.794
	Roy's Largest Root	3.860	521.115 ^a	2.000	270.000	.000	.794
Goal Instructions	Pillai's Trace	.068	3.197	6.000	542.000	.004	.034
	Wilks' Lambda	.932	3.207 ^a	6.000	540.000	.004	.034
	Hotelling's Trace	.072	3.217	6.000	538.000	.004	.035
	Roy's Largest Root	.058	5.273 ^b	3.000	271.000	.002	.055

Note. Alpha = .05 ^a Exact statistic ^b The statistic is an upper bound on F that yields a lower bound on the significance level.

Table 11

Levene's test for equality of variances

	F	df1	df2	Sig.
Development	4.903	3	271	.002
Balance	17.149	3	271	.000

The unadjusted one-way ANOVA outcomes for development, $F(3, 271) = 1.527, p = .208$, and for balance, $F(3, 271) = 4.820, p = .003$ (see Table 12). Table 13 shows the revised outcome, adjusted by Welch and Brown-Forsythe's F statistics. Using the Welch statistic, the researcher found that $F(3, 140.525) = 1.650, p = .181 (p > .05)$. Because the alpha level was set at .05, the researcher rejected the null hypothesis and concluded that the adjusted F ratio was not significant, i.e., there was no significant difference in development scores across goal instructions types. Similarly, the researcher found the adjusted F ratio of balance scores using the Welch statistic: $F(3, 137.285) = 6.475, p = .000 (p < .05)$, concluding that there were significant differences in balance scores across goal instructions types.

Table 12
Unadjusted ANOVA- Development and Balance

		Sum of Squares	df	Mean Square	F	Sig.
Balance	Between Groups	37.465	3	12.488	4.820	.003
	Within Groups	702.171	271	2.591		
	Total	739.636	274			
Development	Between Groups	16.085	3	5.362	1.527	.208
	Within Groups	951.792	271	3.512		
	Total	967.876	274			

Post-Hoc Test. Because the adjusted F ratio of balance scores was found to be significant with the Welch statistic, the researcher compared the group means using the

Games-Howell post-hoc test. This pairwise comparison test was chosen because it is the appropriate test to use when the equal variances assumption has been violated (Mayers, 2013).

Table 13

Adjusted outcome for homogeneity of variance

		Statistic ^a	df1	df2	Sig.
Balance	Welch	6.475	3	137.285	.000
	Brown-Forsythe	4.755	3	218.093	.003
Development	Welch	1.650	3	140.525	.181
	Brown-Forsythe	1.543	3	248.929	.204

Note. ^a Asymptotically F distributed.

Games-Howell pairwise comparison tests revealed that there were significant differences in the balance scores between minimal and no goal instructions (mean difference = 1.103; 95 % CI = .323, 1.883; $p < .05$; $d = 0.723$; $r = 0.340$), moderate and no goal instructions (mean difference = .6742; 95 % CI = .078, 1.271; $p < .05$; $d = 0.475$; $r = 0.231$), and substantial and no goal instructions (mean difference = .6821; 95 % CI = .051, 1.314; $p < .05$; $d = 0.463$; $r = 0.225$) (see Table 14). The Cohen's effect size values ($d = 0.723, 0.475, 0.463$), suggested medium to high practical significance in the balance scores between minimal, moderate and substantial goal instructions, respectively, and the control group (Cohen, 1988). The tests did not reveal significant differences in balance scores between the other groups (minimal and moderate; minimal and substantial; moderate and substantial; $p < .05$).

Table 14

Pairwise Comparison Tests – Balance

(I) Goal Instructions Group	(J) Goal Instructions Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
No	Minimal	-1.1033*	.2967	.002	-1.883	-.323
	Moderate	-.6742*	.2295	.020	-1.271	-.078
	Substantial	-.6821*	.2427	.029	-1.314	-.051
Minimal	No	1.1033*	.2967	.002	.323	1.883
	Moderate	.4291	.3201	.540	-.408	1.267
	Substantial	.4212	.3297	.579	-.440	1.283
Moderate	No	.6742*	.2295	.020	.078	1.271
	Minimal	-.4291	.3201	.540	-1.267	.408
	Substantial	-.0079	.2708	1.000	-.711	.695
Substantial	No	.6821*	.2427	.029	.051	1.314
	Minimal	-.4212	.3297	.579	-1.283	.440
	Moderate	.0079	.2708	1.000	-.695	.711

Note. * The mean difference is significant at the 0.05 level.

ANOVA – Explanatory Discourse. The researcher also conducted one-way ANOVA analysis of explanatory discourse scores. Each group had two scores, one overall score for week two and one for week three. Each group's mean was between 3 and 4 (see Table 15). One-way ANOVA results, $F(3, 4) = 6.00, p = 0.058 (p > .05)$, allowed the researcher to conclude that there were near significant differences in explanatory discourse scores across goal instructions groups (see Table 16). Because no significant differences were found, the researcher did not conduct any post-hoc tests on explanatory discourse scores.

Table 15

Explanatory Discourse Descriptives

Goal Instructions Group	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
No	2	3.0000	.00000	.00000	3.0000	3.0000	3.00	3.00
Minimal	2	4.0000	.00000	.00000	4.0000	4.0000	4.00	4.00
Moderate	2	3.7500	.35355	.25000	.5734	6.9266	3.50	4.00
Substantial	2	3.7500	.35355	.25000	.5734	6.9266	3.50	4.00
Total	8	3.6250	.44320	.15670	3.2545	3.9955	3.00	4.00

Table 16

ANOVA - Explanatory Discourse

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.125	3	.375	6.000	.058
Within Groups	.250	4	.063		
Total	1.375	7			

Qualitative Data

After coming up with a list of fourteen codes using the responses to the open-ended question about the discussion activity, the researcher organized the codes

according to treatment group, or type of goal instructions, and counted instances of each code (see Table 17). Examples of codes that appeared often within each goal instructions group were “open environment to express opinion” in the control group, “topics were touchy/controversial” in the minimal group, “discussions allowed me to learn from others” in the moderate group, and “interesting to see others’ perspectives” in the substantial group.

Table 17

Codes by Treatment Group (Type of Goal Instructions)

Codes	No	Minimal	Moderate	Substantial
Open environment to express opinion	7	5	5	3
Did not feel environment was open	0	1	1	0
Topics were touchy/controversial	4	5	3	5
Liked that the topics were controversial	0	2	0	1
Interesting to see others’ perspectives	5	3	8	10
Topics made me think about my views	3	2	6	2
Discussions allowed me to learn from others	1	1	8	4
Allowed me to consider other people’s perspective	1	0	4	0
Allowed me to consider why others might think my views were wrong	0	1	1	0
Gave me opportunity to discuss issues that are not often mentioned in class	1	1	1	3
Discussions similar to class discussions	0	0	1	1
Encouraged me to do some research	0	1	1	1
Wish more people participated	0	1	0	1
Enjoy opportunity for extra credit	0	0	0	5

After examining the codes, the researcher looked for looked for emergent themes, which gave the data greater depth in meaning (O'Connor & Gibson, 2003). A total of four themes emerged: “discussion board environment,” “views about topics,” “thinking and learning as a result of discussion,” and “relationship to classwork” (see Table 18).

Table 18

Codes organized into Themes

Themes	Codes
Discussion board environment	Open environment to express opinion Did not feel environment was open
Views about topics	Topics were touchy/controversial Liked that the topics were controversial Interesting to see others' perspectives Topics made me think about my views
Thinking and learning as a result of discussion	Discussions allowed me to learn from others Allowed me to consider other people's perspective Allowed me to consider why others might think my views were wrong
Relationship to classwork	Gave me opportunity to discuss issues that are not often mentioned in class Discussions similar to class discussions
Uncategorized	Wish more people participated Encouraged me to do some research Enjoy opportunity for extra credit

The following is an example of a response, in the control group, for the first theme, “discussion board environment”:

I found the three questions asked for this extra credit opportunity were a little controversial. These topics are ones that have been in the news and ones that bring out a lot of emotions in people. I feel that, even though there was a lot of different views and opinions, the class as a whole was very respectful when replying to other class members.

An excerpt from a response found in the minimal goal instructions group for the second theme, “views about topics”:

I enjoyed this extra credit project. I thought the questions were perfect for receiving opposing views. It made me seriously think objectively by citing views that may have been different from my own. It was good to see that some people did have different views than my own.

An example of a response found in the moderate goal instructions group for the third theme, “thinking and learning as a result of discussion”:

This was an interesting extra credit exercise. The discussion questions did make one think outside of the box. To me replying to two other classmates allowed me to view their opinions without being judgmental to their thoughts. I respected what they stated even though I did not agree. Sometimes reading others opinions allows us to view topics in a different light. The discussion questions allowed me to look within myself on writing the answers. It also allowed me the opportunity to look up certain topics on the internet. Those topics I looked up would otherwise most likely have never been looked at. This exercise permitted me to open my own mindset up to look at certain things in a different light. The questions especially the first one was looked at differently among classmates. It

shows that people can interrupt questions in a different way of thinking as far as what is being asked. Young people do have different opinions than the older generation. By reading what they wrote, it allowed me to be respectful and yes to wonder how they can think some of the things they do. You stated be honest!

An example of a response found in the substantial goal instructions group for the fourth theme, “relationship to classwork”:

I actually really enjoyed taking part in this study. I appreciated being able to read and respond to other people's posts. I feel like, at least at my age, there are not many opportunities given for me to take part in debate in such controversial topics. What I enjoyed the most was knowing that those who were responding seemed to be very interested and informed on the different topics. Thank you for the opportunity!

The analysis of the codes and themes revealed that participants across all goal instructions groups noted that the discussion board environment was open, it was interesting to see others' perspectives, the topics were touchy and/or controversial, and the discussions allowed them to learn from others. Most participants' statements were very general and unspecific and no respondents made any statements about the specificity of instructions in the debate prompt. Instead, most participants focused their responses on the actual debate topics. As a result, there seemed to be few differences across the different treatment groups with respect to codes and themes.

The power of qualitative analysis is that it aims to minimize leading the participants in any one particular direction. Thus, the researcher also interpreted the

qualitative findings by examining the differences in respondents' focus on the topics between those who received limited and those who received extended goal instructions.

This approach revealed that the most popular codes in control and minimal goal instructions groups were "open environment to express opinion" and "interesting to see others' perspectives" while the most popular codes in moderate and substantial goal instructions groups were "interesting to see others' perspectives" and "discussions allowed me to learn from others." These findings suggested that respondents who received few instructions focused their discussions on the environment itself while respondents who received many instructions focused their discussions on the impact that topics had on them.

The validity of qualitative analysis findings is higher when they are confirmed by more than one instrument (O'Connor & Gibson, 2003). The researcher triangulated the data from different methods by examining both the qualitative and quantitative results. The major finding of the qualitative data analysis was that the codes and themes emerging from the open-ended response question showed that there were few differences across goal instruction groups. This finding was confirmed by the results of the quantitative portion of the study, which found no significant differences in development and explanatory discourse scores across all goal instruction groups and no significant differences in balance scores across minimal, moderate and substantial goal instructions groups.

Summary of Findings

In the examination of the effects of goal instructions with different degrees of complexity on the quality of argumentation, as measured by development, balance and

explanatory discourse scores, the researcher found that there were significant differences in balance scores between the goal instructions groups and the control condition (no goal instructions) ($F(3, 137.285) = 6.475, p < .05$), but not among the goal instruction groups themselves (minimal, moderate and substantial). In particular, pairwise comparison tests showed that there were significant differences between minimal goal instructions and the control condition ($M = 1.103, 95\% \text{ CI } [.323, 1.883], p < .05; d = 0.723; r = 0.340$), moderate goal instructions and the control condition ($M = .6742, 95\% \text{ CI } [.078, 1.271], p < .05; d = 0.475; r = 0.231$), and substantial goal instructions and the control condition ($M = .6821, 95\% \text{ CI } [.051, 1.314], p < .05; d = 0.463; r = 0.225$). Results also showed that there were no significant differences across treatment groups in development ($F(3, 140.53) = 1.65, p > .05$) and explanatory discourse scores ($F(3, 4) = 6.00, p > .05$). However, explanatory discourse scores were very close to significant ($p = 0.058$), showing that the goal instructions did have some positive effect.

Themes emerging from the qualitative portion the study also showed few differences across goal instruction groups. Most participants noted that the topics were controversial, that they felt comfortable expressing their opinions, and that they found it interesting to read other student's perspectives on these controversial topics. Most participants' opinions were very general and unspecific and no respondents made any statements about the specificity of instructions in the debate prompt. Instead, most participants focused the responses on the actual debate topics.

Qualitative analysis also revealed that the most popular codes in control and minimal goal instructions groups were "open environment to express opinion" and "interesting to see others' perspectives" while the most popular codes in moderate and

substantial goal instructions groups were “interesting to see others’ perspectives” and “discussions allowed me to learn from others.” These findings suggest that participants who received limited goal instructions focused mainly on the discussion board environment while participants who received extended goal instructions focused their discussions on the impact that topics had on them.

CHAPTER IV

DISCUSSION AND CONCLUSIONS

The process of argumentation alters students' underlying beliefs and allows them to identify misconceptions (Baker, 1999). Argumentation is central to learning because it facilitates conceptual change, the process of incorporating new perspectives on a particular issue into a conceptual framework (Asterhan & Schwarz, 2007; Baker, 1999; Nussbaum & Sinatra, 2003; Wiley & Voss, 1999). When incorporated into learning environments, argumentation enhances students' epistemic and conceptual understanding of a topic (Duschl & Osborne, 2002; Wiley & Voss, 1999). However, argumentation is only conducive to learning when the quality of argumentation is high, i.e., students construct strong arguments (Kuhn, 1991; Means & Voss, 1996). One way that students can be prompted to construct better arguments might be through goal instructions (Jonassen, 1999; Collins, Brown, & Newman, 1989; Woods, Bruner, & Ross, 1976). This study investigated the use of goal instructions as a scaffold to support the development of strong arguments in an asynchronous online discussion.

Research Questions

In this study, the researcher compared the effects of different goal instruction conditions (none, minimal, moderate, and substantial) on argumentation development, balance and explanatory discourse scores in an effort to answer the study's central research question: what is the effect of goal instructions with different degrees of specificity on learners' argumentation achievement? More specifically, the study aimed to answer the following research questions: (a) What are the effects of minimal,

moderate, substantial and no goal instructions on participants' development scores? (b) What are the effects of minimal, moderate, substantial and no goal instructions on participants' balance scores? (c) What are the effects of minimal, moderate, substantial and no goal instructions on participants' explanatory discourse scores? (d) What are the participants' experiences of minimal, moderate, substantial, and no goal instructions?

The study found significant differences in the balance scores between minimal, moderate, and substantial goal instructions and the control condition (no goal instructions), but no significant differences among the goal instructions conditions themselves. These findings indicate that goal instructions are effective at creating more balanced responses, which present both sides of an issue. Furthermore, a comparison of effect sizes suggests that minimal goal instructions are more effective than moderate and substantial goal instructions in encouraging participants to present both sides of an issue.

While the study did not find significant differences in explanatory discourse scores among the groups, the differences were very close to being significant. This suggests that goal instructions did have some positive effect on helping participants consider other people's positions in a constructive way and build on each other's ideas.

This study's balance and explanatory discourse findings are supported by the findings of a previous study, which found that learners who were instructed to produce counterclaims and rebuttals generated more counterclaims, rebuttals, and reasons that supported their rebuttals (Nussbaum & Kardash, 2005). They also confirm findings from another study that found that the goal to generate as many reasons as possible resulted in more balanced postings where both sides of an issue were evaluated and explored almost evenly (Nussbaum, 2005). One of the reasons why goal instructions may be effective

scaffolds for balanced arguments is their ability to activate collaborative argumentation schema. This schema in turn helps learners identify patterns of locating and correcting problems within the argument (Nussbaum, 2002). Given the findings of this study, it is also possible to conclude that goal instructions do not need to be particularly specific or detailed to be effective in facilitating balance in discourse.

The study did not find significant differences in the development scores among any of the groups. Findings revealed that minimal, moderate, substantial and no goal instructions did not significantly help participants develop better arguments or create postings with more reasons and evidence in support of their positions. These findings are supported by a previous study, which found that goal instructions did not have a positive effect on argument development and explanatory discourse for low-issue knowledge participants, when prior knowledge was controlled (Golanics & Nussbaum, 2008).

This study's findings that goal instructions did not have a positive effect on argumentation development and explanatory discourse are also supported by literature on cognitive development. Previous research suggests that cognitive development typically takes a minimum of a year to promote through deliberate psychological education (Brendel, Kolbert, & Foster, 2002).

The development scale was interpreted in such a way that it focused mainly on lines of argument and evidence, counting both "reasons" and "evidence" codes as evidence, and excluded originality. Originality was not considered to be an important factor in evaluating development because participants were able to see each other's initial discussion postings. As a result, many participants who posted their initial posts later in the week were able to read and engage with the responses that others have already made

on the topic. This decision to exclude originality from the development score may have impacted the outcome of the study, which found no significant differences in development scores across goal instruction groups.

Some findings of this study varied from the findings of other studies on goal instructions, which found that more specific goal instructions resulted in better argumentation (Golanics & Nussbaum, 2008; Nussbaum, 2005, Nussbaum & Kardash, 2005). However, the findings of this study may be attributed to participants' decision to ignore the majority of the goal instructions. In this study, quality of argumentation was measured by development, balance and explanatory discourse scores and depended on the coding of the responses according to aspects of quality argumentation. The goal instructions were designed in such a way that if participants in the substantial goal instruction groups answered all of the questions in the goal instructions then their responses would have all the appropriate codes of a quality argument, i.e., 'reasons,' 'evidence,' 'counterarguments,' 'reasons why others are wrong,' etc. The researcher expected participants in the moderate and substantial goal instructions groups to have much higher development and balance scores than participants in the other groups. However, very few participants in the moderate and substantial goal instructions groups posted complete responses. Only two responses in the moderate condition and one response in the substantial condition were assigned codes under the entire counterargument category, and no participants stated what evidence those who disagreed with them had for their reasons or provided the assumption. As a result, it appears that many participants did not read, did not understand or chose to ignore the majority of the

goal instructions. This finding further supports the idea that the goal instructions need to be simpler.

The researcher also did not find differences in the emergent themes across goal instructions groups. The majority of the responses to the open-ended qualitative question focused exclusively on the controversial aspects of the debate topics, rather than on the context of the goal instructions. The researcher expected participants in the substantial goal instructions condition to point out something about the complexity of the prompt, for example, that the prompt asked them to do too much in one discussion board posting, required them to think too much about the topic, etc. However, no participants made any such statements, providing more evidence for one of main findings of the quantitative analysis portion: that the majority of the participants possibly ignored the vast majority of the goal instructions. The researcher suspects that very few participants, if any, read and/or considered answering the questions which came after the debate topic question. Findings might have been more informative if the qualitative question was much more explicit by instructing participants to address the nature of the goal instructions directly instead of leaving it more open-ended.

Quantitative analysis also revealed that “open environment to express opinion” and “interesting to see others’ perspectives” were the most popular codes in control and minimal goal instructions groups and “interesting to see others’ perspectives” and “discussions allowed me to learn from others” were the most popular codes in moderate and substantial goal instructions groups. These findings suggested that participants who received limited instructions focused their discussions on the environment itself while participants who received extended instructions focused their discussions on the impact

that debates had on them. Therefore, it is likely that more extended instructions made an impact on encouraging participants to think about their views and consider other people's perspectives.

Conclusion

Different types of goal instructions presented in this study aimed at encouraging participants to engage in better argumentation without explicitly teaching them how to construct good arguments. Results showed that goal instructions are effective in creating more balanced responses, which consider both sides of a controversial topic, and that any level of specificity in the goal instructions is effective in creating balanced replies in comparison to no goal instructions. Results also suggest that minimal goal instructions might be more effective in creating balance replies than moderate and substantial goal instructions and that goal instructions have some positive effect on encouraging students to explore opposing perspectives. This study's findings are supported by prior studies, which also found goal instructions to be effective in helping participants evaluate and explore both sides of an issue (Nussbaum, 2005; Nussbaum & Kardash, 2005). Finally, the effectiveness of goal instructions in facilitating balance may be attributed to their ability to activate participants' collaborative argumentation schema (Nussbaum, 2002). The results of this study suggest that teachers who use goal instructions in distance courses to provide their students with a platform to engage in quality argumentation should focus their energies on providing simple and concise instructions in addition to the debate prompts.

Limitations

One of the main limitations of this study is how quality of argumentation is

evaluated. The researcher chose to use the development, balance and explanatory discourse scales used previously because it presented a more holistic approach to evaluating quality of argumentation (Golanics & Nussbaum, 2008). However, using these scales is only one of the many ways of evaluating argumentation quality, and there is currently uncertainty in how argumentation quality should be judged (Erduran, 2008; Nussbaum, 2011; Scheuer et al., 2010). Some studies focus on certain features of sound arguments such as relevant and acceptable reasons (Means & Voss, 1996), others on a general evaluator criteria, such as lack of supporting evidence (Kuhn, Kenyon & Reiser, 2006), and others propose a framework for evaluating arguments on two dimensions: conceptual quality and levels of opposition (Clark & Sampson, 2008). Furthermore, some studies rely entirely on quantity of arguments, counterarguments, rebuttals, and evidence, ignoring the content of the arguments (Nussbaum, Winsor, Aqui & Poliquin, 2007) while others focus entirely on one aspect of argumentation: the counterargument (Jonassen & Cho, 2011; Kuhn, Goh, Iordanou, & Shaenfield, 2008).

Another important limitation of this study is that originality, or the extent to which participants' initial posts contained original arguments that were not brought up by any other students, was not considered an important factor in evaluating development. Originality was one of the evaluators used to apply the development scale and this study's decision to leave the discussion groups open and interpret the scale without it likely had an impact on how those responses were coded and scored. Furthermore, this decision impacted the outcome of the study, which found no significant differences in development scores across goal instruction groups.

Finally, another limitation of this study is lack of direct incentive. While participants were provided with extra credit points to encourage their participation in the study, they were not incentivized directly to provide complete and thorough responses. Participants who participated in all three weeks of the study were provided with the same number of extra credit points regardless of the number and the thoroughness of their responses. This decision might have had an adverse effect on the quality of argumentation, especially for participants in the moderate and substantial groups.

Future Research

The findings of this study suggest that the majority of the moderate and substantial goal instructions may have been dismissed by many of the participants. This conclusion suggests that further research is needed to determine why the instructions were ignored and how goals should be presented to encourage participants to follow all the instructions. It is possible that the controversial nature of the debates made the participants eager to relay their opinions and to engage in the discussion, causing them to ignore, overlook or dismiss the instructions in the prompts. Another possibility is that the instructions were too complicated and/or participants did not want to engage in the actual activity of considering other people's reasons, evidence, and/or positions.

One possible avenue for further research is to use less controversial debate topics and another is to evaluate the quality of argumentation using one or a couple of other approaches discussed above. Evaluating argumentation according to different measures such as quantity of reasons, evidence, rebuttals, etc. or focusing exclusively on counter argumentation as a measure of a quality argument will give goal instruction research a more comprehensive understanding of their effectiveness.

Future research might also consider conducting a similar study, but this time closing the discussion boards for a few days in the beginning of the study, until all participants posted their initial replies. This approach of using closed moderated discussion boards will allow researchers to determine the extent of participants' original thinking and give a better understanding of the impact of originality on participants' development scores across goal instructions groups.

Future research might also look into alternative presentation options. Goal instructions in this study were presented as a large block of text following the debate prompt. This manner of presentation might have given participants in moderate and substantial groups the permission to ignore most of the instructions. To promote higher response rates or adherence to directions, future research might consider presenting complex goal instructions in a different way within the discussion board itself. For instance, researchers might break up the sentences into individual bulleted or numbered parts or even leave large blanks in between the each goal instruction. These methods of presentation might encourage participants to fill in answers to all questions posted, thus prompting them to respond to all of the instructions.

Finally, future research into different types of goal instructions might look into assigning extra credit based on thoroughness of their responses. This approach in actually grading participants' responses might encourage them to respond to all questions in the goal instructions instead of ignoring the more complex and extended goal instructions.

REFERENCES

- Alexopoulou, E., & Driver, R. (1996). Small-group discussion in physics: Peer interaction modes in pairs and fours. *Journal of Research in Science Teaching*, 33(10), 1099-1114.
- Anderson R.C., Nguyen-Jahiel K., McNurlen B., Archodidou A., Kim S., Reznitskaya A., Tillmans M. & Gilbert L. (2001). The snowball phenomenon: spread of ways of talking and ways of thinking across groups of children. *Cognition and Instruction* 19, 1–46.
- Andriessen J. (2006). Collaboration in computer conferencing. In A.M. O'Donnell, C.E. Hmelo-Silver & G. Erkens (Eds.), *Collaborative Learning, Reasoning, and Technology* (pp. 197-230). Mahwah, NJ: Erlbaum.
- Andriessen J., Baker M. & Suthers D. (2003) Argumentation, computer support, and the educational context of confronting cognitions. In *Arguing to Learn: Confronting Cognitions in Computer-Supported Collaborative Learning Environments* (eds J. Andriessen, M. Baker, & D. Suthers), pp. 1–25. Kluwer, Boston.
- Asterhan, C. S. C., & Schwarz, B. B. (2007). The effects of monological and dialogical argumentation on concept learning in evolutionary theory. *Journal of Educational Psychology*, 99(3), 626–639.
- Baker, M. (1999). Argumentation and constructivist interaction. In J. Andriessen & P. Coirier (Eds.), *Foundations of argumentative text processing* (pp. 179-202).

Amsterdam: University of Amsterdam.

- Baker, M. (2003). Computer-mediated argumentative interactions for the co-elaboration of scientific notions. In J. Andriessen, M. Baker & D. Suthers (Eds.), *Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments* (pp. 47-78). Boston, MA: Kluwer Academic.
- Beardseley, M. C. (1950). *Practical Logic*. New York: Prentice-Hall.
- Bell, N., Grossen, M., & Perret-Clermont, A.-N. (1985). Sociocognitive conflict and intellectual growth. In M. W. Berkowitz (Ed.), *Peer conflict and psychological growth* (pp. 283-322). San Francisco: Jossey-Bass.
- Bell, P., & Linn, M. C. (2000). Scientific arguments as learning artifacts: designing for learning from the web with KIE. *International Journal of Science Education*, 22(8), 797-817.
- Brendel, J. M., Kolbert, J.B, & Foster, V.A. (2002). Promoting student cognitive development. *Journal of Adult Development*, 9(3), 217-227.
- Chin, C., & Osborne, J. (2010). Students' questions and discursive interaction: their impact on argumentation during collaborative group discussions in science. *Journal of Research in Science Teaching*, 47(7), 883-908.
- Chinn, C.A. (2006) Learning to argue. In *Collaborative Learning, Reasoning, and Technology* (eds A.M. O'Donnell, C.E. Hmelo-Silver & G. Erkens), pp. 355–383. Erlbaum, Mahwah, NJ.
- Cho, K. L., & Jonassen, D. H. (2002). The effects of argumentation scaffolds on argumentation and problem solving. *Educational Technology Research and Development*, 50(3), 5-22.

- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment* 6, 284–90.
- Clark, D. B., & Sampson, V. (2008). Assessing dialogic argumentation in online environments to relate structure, grounds, and conceptual quality. *Journal of Research in Science Teaching*, 45, 293–321.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychosocial Measurement* 20, 37–46.
- Cohen, J. J. (1968). Weighted Kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70, 213–220.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd Edition). Hillsdale, NJ: Lawrence Earlbaum Associates.
- Collins, A., Brown, J.S. & Newman, S.E. (1989). Cognitive apprenticeship: Teaching the craft of reading, writing and mathematics. In L.B. Resnick (Ed.), *Knowing, learning and instruction: Essays in honor of Robert Glaser* (pp. 453-494). Hillsdale, NJ: Erlbaum.
- Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into Practice*, 39(3), 124-130.
- Creswell, J.W. *Research Design: Qualitative, Quantitative and Mixed Methods Approaches* (3rd Edition). (2009). Thousand Oaks, CA: Sage.
- Dole, J. A., & Sinatra, G.M. (1998). Reconceptualizing change in the cognitive construction of knowledge. *Educational Psychologist*, 33(2/3), 109–128.
- Duschl, R. A., & Osborne, J. (2002). Supporting and promoting argumentation

discourse in science education. *Studies in Science Education*, 38, 39–72.

- Erduran, S. (2008). Methodological foundations in the study of argumentation in the science classroom. In S. Erduran & M. P. Jimenez-Aleixandre (Eds.), *Argumentation in science education: Perspectives from classroom-based research* (pp. 47–69). Dordrecht, The Netherlands: Springer.
- Erkens, G., Andriessen, K., & Peters, N. (2003). Interaction and performance in computer-supported collaborative tasks. In H. van Oostendorp (Ed.), *Cognition in a digital world* (pp. 225-251). Mahwah, NJ: Erlbaum.
- Ferretti R.P., MacArthur C.A. & Dowdy N.S. (2000). The effects of an elaborated goal on the persuasive writing of students with learning disabilities and their normal achieving peers. *Journal of Educational Psychology*, 92, 694–702.
- Golanics, J. D., & Nussbaum, E. M. (2008). Enhancing online collaborative argumentation through question elaboration and goal instructions. *Journal of Computer Assisted Learning*, 24, 167-180.
- Hinkle, D.E., Wiersma, W., & Jurs, S.G. (2003). *Applied statistics for the behavioral sciences*. Boston: Houghton Mifflin.
- Hruschka, D. J., Schwartz, D., Cobb St. John, D., Picone-Decaro, E., Jenkins, R. A., & Carey, J. W. (2004). Reliability in coding open-ended data: lessons learned from HIV behavioral research. *Field Methods* 16 (3), 307-331.
- Inch, E. S., & Warnick, B. (2002). *Critical thinking and communication*. Boston: Allyn & Bacon.
- Järvelä, S., & Häkkinen, P. (2003). The levels of web-based discussions: Using

- perspective- taking theory as an analytical tool. In H. van Oostendorp (Ed.), *Cognition in a digital world* (pp. 77-95). Mahwah, N.J.: Lawrence Erlbaum Associates.
- Jeong, A. C., & Joung, S. (2007). Scaffolding collaborative argumentation in asynchronous discussions with message constraints and message labels. *Computers & Education*, 48(3), 427-445.
- Johnson, D. W., & Johnson, R.T. (1987). *Learning together and alone: Cooperative, competitive and individualistic learning*. Englewood Cliffs, NJ: Prentice Hall.
- Johnson, D. W., & Johnson, R. (2004). Cooperation and the use of technology. In D. H. Jonassen (Ed.), *Research for Educational Communications and Technology* (2 ed., pp. 785-811). Bloomington, IN: Association for Educational Communications & Technology.
- Jonassen, D. H. (1999). Toward a design theory of problem solving. *Educational Technology: Research & Development*, 48(4), 63–85.
- Jonassen, D. H. & Cho, Y. H. (2011). Fostering Argumentation While Solving Engineering Ethics Problems. *Journal of Engineering Education*. 100(4). 680-702.
- Jonassen, D. H., & Kim, B. (2010). Arguing to learn and learning to argue: Design justifications and guidelines. *Educational Technology: Research & Development*, 58(4), 439–457.
- Keith, T. Z. (2006). *Multiple regression and beyond*. Boston: Allyn & Bacon.
- Koschmann, T. D. (1996) Paradigm shifts and instructional technology. In T. Koschmann (Ed.) *CSCL: Theory and Practice of an Emerging Paradigm*, (pp.

- 1-23). Mahwah, NJ: Lawrence Erlbaum.
- Koschmann, T. D. (2003). CSCL, argumentation, and Deweyan inquiry: Argumentation is learning. In J. Andriessen, M. Baker & D. Suthers (Eds.), *Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments* (pp. 261-269). Boston, MA: Kluwer Academic.
- Kuhn, D. (1991). *The skills of argument*. Cambridge, England: Cambridge University Press.
- Kuhn, D., Goh, W., Iordanou, K., & Shaenfield, D. (2008). Arguing on the computer: A microgenetic study of developing argument skills in a computer-supported environment. *Child Development, 79*, 1310–1328. doi:10.1111/j.1467-8624.2008.01190.x
- Kuhn, L., Kenyon, L., & Reiser, B. J. (2006). Fostering scientific argumentation by creating a need for students to attend to each other's claims and evidence. *A paper submitted to the 7th International Conference for the Learning Sciences*, Bloomington, IN.
- Kuhn, D., Shaw, V., & Felton, M. (1997). Effects of Dyadic Interaction on Argumentative Reasoning. *Cognition and Instruction, 15*(3), 287-315.
- Landis, J. R. & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics 33*, 159–74.
- Leitão, S. (2000). The potential of argument in knowledge building. *Human Development, 43*, 332-360.
- Mandl, H., Gruber, H., & Renkl, A. (1996). Inert knowledge: Analyses and remedies. *Educational Psychologist, 31*(2), 115-121.

- Marshall, C. & Rossman, G. (1995). *Designing Qualitative Research*. Sage Publications. Thousand Oaks, California.
- Mayers, A. (2013). *Introduction to Statistics and SPSS in Psychology*. Prentice Hall. Upper Saddle River, New Jersey.
- Means, M. L., & Voss, J. F. (1996). Who reasons well? Two studies of informal reasoning among children of different grade, ability, and knowledge levels. *Cognition and Instruction, 14*(2),139-178.
- Mercer, N. (1996). The quality of talk in children's collaborative activity in the classroom. *Learning and Instruction 6*, 359-377.
- Merriam, S. B. (1998). *Qualitative research and case study applications in education*. San Francisco, CA: Jossey-Bass.
- Miles, M. B. & Huberman, A. M. (1994). *Qualitative data analysis*. Thousand Oaks, CA: Sage.
- Morrison, G. R., & Anglin, G. J. (2006). An instructional design approach for effective shovelware: Modifying materials for distance education. *Quarterly Review of Distance Education, 7*(1), 63-74.
- Newton, P., Driver, R., & Osborne, J. (1999). The Place of Argumentation in the Pedagogy of School Science. *International Journal of Science Education, 21*(5), 553-576.
- Nussbaum, E. M. (2002). How introverts versus extroverts approach small-group argumentative discussions. *The Elementary School Journal, 102*(3), 183-197.
- Nussbaum, E. M. (2005). The effect of goal instructions and need for cognition on

- interactive argumentation. *Contemporary Educational Psychology*. 30(3), 286-313.
- Nussbaum, E.M. (2011). Argumentation, dialogue theory, and probability modeling: alternative frameworks for argumentation research in education. *Educational Psychologist*, 46(2), 84-106.
- Nussbaum, E.M. & Kardash, C.M. (2005). The effects of goal instruction and text on the generation of counterarguments. *Journal of Educational Psychology* 97, 157– 169.
- Nussbaum, E. M., & Schraw, G. (2007). Promoting argument-counterargument integration in students' writing. *Journal of Experimental Education*, 76(1), 59-92.
- Nussbaum, E. M., & Sinatra, G. M. (2003). Argument and conceptual engagement. *Contemporary Educational Psychology*, 28(3), 384-395.
- Nussbaum E.M., Winsor D.L., Aqiu Y.M. & Poliqiun A.M. (2006) *Putting the pieces together: on-line argumentation vee diagrams enhance thinking during discussion*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.
- O'Connor, H & Gibson, N. (2003). A step-by-step guide to qualitative data analysis. *Pimatisiwin: A Journal of Aboriginal & Indigenous Community Health* 1(1), 63-78.
- Oh S. & Jonassen D.H. (2007) Scaffolding online argumentation during problem solving. *Journal of Computer Assisted Learning* 23, 95–110.
- Page-Voth V. & Graham S. (1999) Effects of goal setting and strategy use on the

writing performance and self-efficacy of students with writing and learning problems. *Journal of Educational Psychology* 91, 230–240.

Piaget, J. (1977). *The Development of Thought: Equilibration of Cognitive Structures*, New York: Viking.

Pilkington R. & Walker A. (2003) Using CMC to develop argumentation skills in children with a 'literacy deficit'. In *Arguing to Learn: Confronting Cognitions in Computer-Supported Collaborative Learning Environments* (eds J. Andriessen, M. Baker & D. Suthers), pp. 144–175. Kluwer, Boston.

Sappo, T., & Mononen-Aaltonen, M. (1998). Developing dialogic communication culture in media education: integrating dialogism and technology. *Media Education Publication*, 7.

Scheuer, O., Loll, F., Pinkwart, N., & McLaren, B. M. (2010). Computer supported argumentation: A review of the state of the art. *International Journal of Computer-Supported Collaborative Learning*, 5, 43–102.

Schwarz, B. B., Neuman, Y., & Biezuner, S. (2000). Two Wrongs May Make a Right ...If They Argue Together! *Cognition and Instruction*, 18(4), 461-494.

Siegel, D.J. (1995). Memory, trauma, and psychotherapy: A cognitive science view. *Journal of Psychotherapy Practice and Research*, 4, 93– 122.

Slavin, R. E. (1996). Research on cooperative learning and achievement: What we know, what we need to know. *Contemporary Educational Psychology*, 21(1), 43-69.

Soper, D.S. (2013). A-priori Sample Size Calculator for Multiple Regression [Software]. Available from <http://www.danielsoper.com/statcalc>

- Veerman A.L. (2003) Constructive discussions through electronic dialogue. In *Arguing to Learn: Confronting Cognitions in Computer-Supported Collaborative Learning Environments* (eds J. Andriessen, M. Baker & D. Suthers), pp. 117–143. Kluwer, Boston.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard.
- Vygotsky, L. S. (1981). The development of higher forms of attention in childhood. In J. V. Wertsch (Ed.), *The concept of activity in Soviet psychology*. Armonk, N.Y.: Sharpe.
- Wiley, J., & Voss, J. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology, 91*, 301–311.
- Wittrock, M. C. (1992). Generative Learning Processes of the Brain. *Education Psychologist, 27*(4), 531-541.
- Woods, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*, 89-100.

APPENDIX A

INFORMED CONSENT DOCUMENT

PROJECT TITLE: Using goal instructions to improve the quality of argumentation in asynchronous online discussions

INTRODUCTION

The purposes of this form are to give you information that may affect your decision whether to say YES or NO to participation in this research, and to record the consent of those who say YES. Other than your normal class attendance in the specified classroom location on record in the Registrar's Office, you will not be required to travel to any location beyond your usual traversal of your university to access the Blackboard course management system. You may participate in this study using any computer or mobile device that is Internet accessible.

RESEARCHERS

Dr. Amy Adcock, Ed.D.

Darden College of Education

STEM Education and Professional Studies

Yekaterina Prudchenko, Ph. D. Student

Darden College of Education

STEM Education and Professional Studies

DESCRIPTION OF RESEARCH STUDY

The study will extend current research into goal instructions that facilitate argumentation.

The primary purpose of the study is to compare the effectiveness of goal instructions with different degrees of specificity on the quality of argumentation in asynchronous online discussions.

EXCLUSIONARY CRITERIA

To the best of your knowledge, you should be at least 18 years old, and you should not have been enrolled at this institution or any other institution of higher education (e.g., community college, junior college, virtual college, another university, etc.) for the equivalence of one academic year (i.e., two semesters, three trimesters, four quarters, or the equivalence of 30 semester credit hours).

RISKS AND BENEFITS

RISKS: There is no known emotional, psychological, physical risk involved in this study. However, as with any research, there is some possibility that you may be subject to risks that have not yet been identified.

BENEFITS: The main benefit to you for participating in this study is extra credit points, which will be determined by the course instructor. Furthermore, the results of the study

may help the researchers recommend appropriate interventions that might assist your university in developing a program to aid in developing learning strategy sessions that can contribute to the academic success of first-year university students.

COSTS AND PAYMENTS

The researchers want your decision about participating in this study to be absolutely voluntary. Other than the extra credit points awarded at the discretion of the course instructor, the researchers are unable to give you any payment for participating in this study.

NEW INFORMATION

If the researchers find new information during this study that would reasonably change your decision about participating, then they will make you aware of it.

CONFIDENTIALITY

All information obtained about you in this study is strictly confidential unless the law requires disclosure. The results of this study may be used in reports, presentations and publications, but the researchers will not identify you.

WITHDRAWAL PRIVILEGE

It is OK for you to say NO. Even if you say YES now, you are free to say NO later, and walk away or withdraw from the study— at any time. Your decision will not affect your relationship with Old Dominion University, or otherwise cause a loss of benefits to which

you might otherwise be entitled. The researchers reserve the right to withdraw your participation in this study, at any time, if they observe potential problems with your continued participation. If you choose not to participate, you can write a short essay arguing your position on the question for extra credit.

COMPENSATION FOR ILLNESS AND INJURY

If you say YES, then your consent in this document does not waive any of your legal rights. However, in the event of injury or illness arising from this study, neither Old Dominion University nor the researchers are able to give you any money, insurance coverage, free medical care, or any other compensation for such injury. In the event that you suffer injury as a result of participation in this research project, you may contact: Dr. Amy Adcock, Responsible Principal Investigator, at (757) 683-5491; Yekaterina Prudchenko, Investigator at (310) 499-3488; Dr. Nina Brown, Chair of the Darden College of Education Human Subjects Review Committee, at (757) 683-3245); or, Dr. George Maihafer, the current IRB Chair, at 757-683 6028 at Old Dominion University, who will be glad to review the matter with you.

VOLUNTARY CONSENT

By signing this form, you are saying several things. You are saying that you have read this form or have had it read to you, that you are satisfied that you understand this form, the research study, and its risks and benefits. The researchers should have answered any questions you may have had about the research. If you have any questions later on, then the researchers should be able to answer them:

Amy Adcock, Ed.D. – (757) 683-5491

Yekaterina Prudchenko – (310) 499-3488

7930 Willoughby Ave Apt 7

West Hollywood, CA 90046

If at any time you feel pressured to participate, or if you have any questions about your rights or this form, then you should call Dr. Ted Remley, Chair of the Darden College of Education Human Subjects Review Committee at (757) 683-3326 or Dr. George Maihafer, the current IRB Chair, at (757) 683 6028, or the Old Dominion University Office of Research, at (757) 683 3460.

And importantly, by clicking Next below, you are telling the researcher YES, that you agree to participate in this study. The researcher should give you a copy of this form for your records.

APPENDIX B
DEMOGRAPHIC SURVEY

1. Enter your University ID Number (UIN) to make sure you get credit for participating in this study. _____
2. Gender (Select): M F
3. How old are you? _____
4. What is your major? _____
5. What year are you in school? (Select)
 - a. Freshman
 - b. Sophomore
 - c. Junior
 - d. Senior

APPENDIX C

WEEK 1 TREATMENTS

Treatment	Debate Question
No Goal Instructions	Should hospitals be mandated to provide birth control? Explain.
Minimal Goal Instructions	Should hospitals be mandated to provide birth control? Provide as many reasons as you can to justify your position. Then provide as many reasons as you can as to why others might disagree with you, and why those reasons are wrong.
Moderate Goal Instructions	Should hospitals be mandated to provide birth control? Provide as many reasons as you can to justify your position citing evidence that supports your reasons. Provide as many reasons as you can as to why others might disagree with you, with evidence that supports those reasons. Then explain why those reasons are wrong.
Substantial Goal Instructions	Should hospitals be mandated to provide birth control? Provide as many reasons as you can to justify your position citing evidence that supports your reasons. What are your assumptions? What are the implications of these assumptions? Provide as many reasons as you can as to why others might disagree with you, with evidence that supports those reasons. What are their assumptions? What are the implications of those assumptions? Then explain why those reasons are wrong.

APPENDIX D

WEEK 2 TREATMENTS

Treatment	Debate Question
No Goal Instructions	Should doctor assisted suicides be legal or illegal for terminally-ill patients? Explain.
Minimal Goal Instructions	Should doctor assisted suicides be legal or illegal for terminally-ill patients? Provide as many reasons as you can to justify your position. Then provide as many reasons as you can as to why others might disagree with you, and why those reasons are wrong.
Moderate Goal Instructions	Should doctor assisted suicides be legal or illegal for terminally-ill patients? Provide as many reasons as you can to justify your position citing evidence that supports your reasons. Provide as many reasons as you can as to why others might disagree with you, with evidence that supports those reasons. Then explain why those reasons are wrong.
Substantial Goal Instructions	Should doctor assisted suicides be legal or illegal for terminally-ill patients? Provide as many reasons as you can to justify your position citing evidence that supports your reasons. What are your assumptions? What are the implications of these assumptions? Provide as many reasons as you can as to why others might disagree with you, with evidence that supports those reasons. What are their assumptions? What are the implications of those assumptions? Then explain why those reasons are wrong.

APPENDIX E

WEEK 3 TREATMENTS

Treatment	Debate Question
No Goal Instructions	Should recreational use of marijuana be legal or illegal for adults? Explain.
Minimal Goal Instructions	Should recreational use of marijuana be legal or illegal for adults? Provide as many reasons as you can to justify your position. Then provide as many reasons as you can as to why others might disagree with you, and why those reasons are wrong.
Moderate Goal Instructions	Should recreational use of marijuana be legal or illegal for adults? Provide as many reasons as you can to justify your position citing evidence that supports your reasons. Provide as many reasons as you can as to why others might disagree with you, with evidence that supports those reasons. Then explain why those reasons are wrong.
Substantial Goal Instructions	Should recreational use of marijuana be legal or illegal for adults? Provide as many reasons as you can to justify your position citing evidence that supports your reasons. What are your assumptions? What are the implications of these assumptions? Provide as many reasons as you can as to why others might disagree with you, with evidence that supports those reasons. What are their assumptions? What are the implications of those assumptions? Then explain why those reasons are wrong.

APPENDIX F
DEVELOPMENT SCALE

6 points	Posts with five to six lines of argumentation, most of which are original, i.e. not brought up by another participant, with evidentiary support
5 points	Posts with five to six lines of argumentation, four of which are original and half of which are supported with evidence
4 points	Posts with three to four lines of argumentation, three of which are original and half of which are supported with evidence
3 points	Posts with three to four lines of argumentation, one of which is original and one of which is supported with evidence
2 points	Posts with one to two lines of argumentation, none of which are original and one or two of which are supported with evidence
1 point	Posts with one to two lines of argumentation, all unclear

Note. Development score assesses how well arguments are developed and gives higher scores for originality and evidentiary support (Golanics & Nussbaum, 2008).

APPENDIX G
BALANCE SCALE

5 points	Posts that propose solutions and 'it depends' arguments
4 points	Posts that propose small solutions/it depends arguments, or explore both sides of the issue to some degree, or where there is a shift in perspective
3 points	Posts that make some concessions to other students' arguments and built upon opposing viewpoints
2 points	Posts that are mainly one sided but rebut the opposing side
1 point	Posts that show no consideration of an opposing view point

Note. Balance score assesses how well arguments present both sides of an issue (Golanics & Nussbaum, 2008).

APPENDIX H

EXPLANATORY DISCOURSE SCALE

- 5 points All members are critical but flexible, willing to concede, i.e. exploratory
- 4 points Two thirds of the members are explanatory
- 3 points One thirds of the members are explanatory
- 2 points All members are either cumulative (all agree/build on each other's ideas) or disputational (all opposed each other's ideas)
- 1 point Group members mainly repeat one another's comments.

Note. Explanatory discourse scores assesses how well participants interact with one another and consider each member's perspectives in a constructive way (Golanics & Nussbaum, 2008).

VITA

Yekaterina Prudchenko

7930 Willoughby Ave Apt.7

West Hollywood, CA 90046

Education

PhD Candidate, Education: Instructional Design and Technology

Old Dominion University; GPA: 3.98; January 2010 – December 2013

M.A. Interdisciplinary Studies: English and Writing

Western New Mexico University; GPA: 4.0; December 2012

M.S. Applied Mathematics (equivalency)

California State University Long Beach; May 2008

B.A. Mathematics

University of Southern California; August 2005

Publications

Reid, A., Prudchenko, Y. (2013). Online behavior of the social media student. In D'Agustino, Steven (Eds.), *Immersive Environments, Augmented Realities and Virtual Worlds: Assessing Future Trends in Education* (pp.161-176). Fordham University, New York: IGI Global.

Lee, J., Prudchenko, K., Ruiz, E. J., & Boekema, C. (2004). Plausibility of antiferromagnetism in and around the vortex cores of Bi2212 and Tl2223. *Journal of Applied Physics* 95(11) 6906-6908. DOI: 10.1063/1.1669371.

Prudchenko, K., Boekema, C. (2004). Magnetism in superconducting Tl2223. *Journal of Young Investigators* 11 (1).