MTR 04B0000074

MITRE TECHNICAL REPORT

An Empirical Evaluation of Structured Argumentation Using the Toulmin Argument Formalism

October 2004

Brant A. Cheikes
Paul E. Lehner
Mark F. Taylor
Leonard Adelman[†]

Sponsor: MITRE Sponsored Research **Project No.:** 51MSR114-A4
Dept. No.: G062

Center for Integrated Intelligence Systems
Bedford, Massachusetts

[†] George Mason University

Abstract

Structured argumentation tools are software-based cognitive aids intended to help information analysts more rigorously develop and communicate the reasoning behind their conclusions. Some of these tools employ Toulmin’s argument formalism, but there has been no controlled research demonstrating the formalism’s effectiveness in supporting argument evaluation or communication. An experiment was conducted to address this need by assessing whether the use of the Toulmin formalism positively impacted participants’ ability to evaluate and communicate the arguments presented in two articles, each approximately 2,000 words in length. The results were mixed, with the formalism having a positive impact for only one of the two articles. In general, participants found it difficult to generate Toulmin structures, and their structures varied greatly even though they started with the same content. Consequently, one should be cautious of the claimed value of structured argumentation tools employing the Toulmin formalism without future empirical research demonstrating its value.

Table of Contents

Introduction

Structured argumentation tools (e.g., see Moulin, Irandoust, Belanger, & Desbordes, 2002; Sillince & Saeedi, 1999 for reviews) are software-based cognitive aids intended to help information analysts carry out the mentally demanding aspects of their work – like organizing and weighing evidence, and drawing and communicating sound conclusions – more systematically and rigorously. Some tools have been proposed or designed to allow analysts to express the details of their reasoning using Toulmin argument structures (Toulmin, 1958). Although research in informal logic and critical thinking indicates that Toulmin's argument formalism does not adequately describe how people typically argue with one another (e.g., see Van Eemeren, Grootendorst, & Henkemans, 1996), it is quite possible that the formalism improves argument evaluation and communication when used prescriptively, as in structured argumentation tools. However, we have been unable to find any controlled experiments testing whether the use of Toulmin argument structures aid the evaluation of argument soundness or facilitate argument communication. A literature search conducted in July 2003 found only one experiment since 1990 (Wallace, 1992) evaluating the training value of Toulmin argument structures, and this experiment found no difference with the control group.

This paper describes an experiment to test whether the use of Toulmin argument structures can (a) affect the evaluation of argument soundness, and (b) facilitate argument communication. The paper is divided into five parts. The first part describes Toulmin argument structures. The second part describes the experimental procedures. The third part presents the hypotheses and measures for testing them. The fourth part presents the results. The fifth part discusses their implications.

Toulmin Argument Structures

Toulmin argument structures have six components:

1. **Claim –** this is the expressed opinion or conclusion that the arguer wants accepted by the audience;

2. **Grounds** (for the claim) – this is the evidence or data for the arguer's claim;

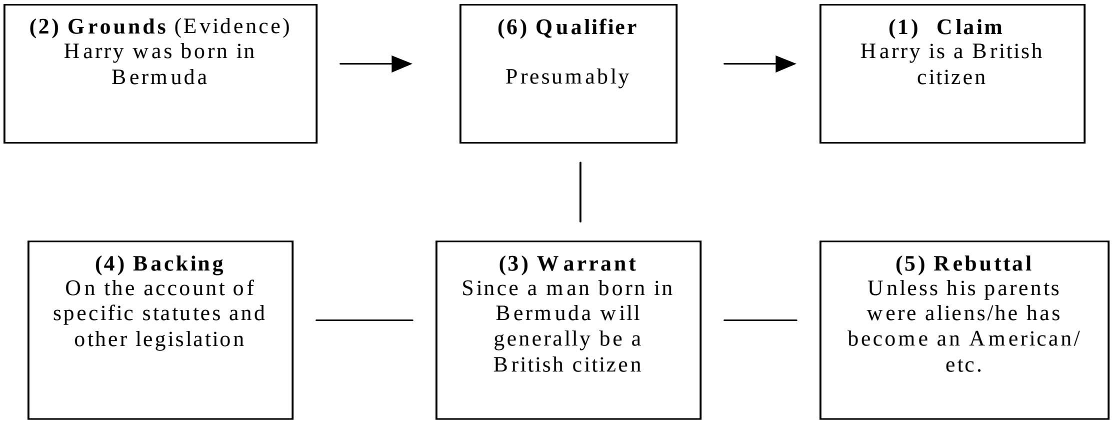
3. **Warrant –** this is the arguer's reasoning (e.g., rule or principle) for connecting the data to the claim;

4. **Backing –** further facts or reasoning used to support or legitimate the warrant;

5. **Rebuttal –** this represents circumstances or conditions that undermine the argument; it represents any reservations or “exceptions to the rule” that undermines the reasoning expressed in the warrant or the backing for it;

6. **Qualifier –** an adverbial phrase indicating the strength of the claim, such as using the phrases certainly, presumably, probably, possibly, etc.

Here is a simple example by Toulmin presented to illustrate all six parts of an argument.


(2) Grounds (Evidence)
Harry was born in Bermuda
(6) Qualifier
Presumably
(1) Claim
Harry is a British citizen
(4) Backing
On the account of specific statutes and other legislation
(3) Warrant
Since a man born in Bermuda will generally be a British citizen
(5) Rebuttal
Unless his parents were aliens/he has become an American/ etc.

This argument claims that Harry is a British citizen because he was born in Bermuda. This claim is presumably true since people born in Bermuda are generally British citizens (the warrant) because there are statutes and other legislation substantiating this rule (the warrant’s backing). As the rebuttal points out, however, there are exceptions to this rule (the warrant), such as when a person born in Bermuda has parents of another nationality or if that person becomes a naturalized American citizen.

It is quite possible for an argument to lack one or more of the components of Toulmin’s argument structure, or to have deficiencies in them. Indeed, weaker arguments often have significant holes, for example, in the grounds supporting the claim or in the backing supporting the warrant (general rule) or in considering and countering obvious rebuttals. Both arguments used in our experiment had significant deficiencies, most noticeably in failing to consider and counter obvious rebuttals to the warrant and, in turn, the claim.

2 Experimental Method

This section describes the participants, procedures, and materials used in the experiment.

Participants: 24 employees in a research and development corporation volunteered, and 22 participated in the experiment. All were members of a corporate intelligence-analysis community of interest (COI), and responded to an e-mail solicitation sent to that COI. Potential volunteers were told that they would be participating in an experiment intended to assess the costs and benefits of structured argumentation for intelligence analysis, that it would conducted in two sessions (Parts A and B) via e-mail taking a total of 3-4 hours of their time, and that a charge number would be provided to cover their time. All 22 participants completed Part A, but only 20 completed Part B, which requested biographical data. All 20 participants had completed college: 4 had a B.S. degree, 12 had a M.S. degree, and 4 had a Ph.D. degree. Ten of the 20 participants had intelligence analysis experience. Experience ranged from less than a year to 26 years: 2 had 2 years (or less), 4 between 8 and 14 years (inclusive), and 4 had 15 years or more.

Procedures: The entire experiment was conducted via e-mail, and all data was collected within a six-week time period. There were six steps. Steps 1 through 3 represented Part A of the experiment, which tested whether Toulmin argument structures impacted participants' evaluation of argument soundness. Steps 4 through 6 represented Part B, which tested whether the structures facilitated communication of the essential elements of the argument. Participants were told to e-mail their responses back to the experimenter after each step, although they could keep a copy of it for future reference as needed. Lastly, participants were asked to record the start and end times for Steps 1 through 5.

Step 1 (Initial argument soundness rating): Participants read an article of approximately three typed, single-spaced pages in length. Twelve of the participants read an article (1,694 words long) which argued the claim that "In the US, better health care also can be cheaper health care." Ten participants read an article (1,966 words long) arguing that "The Saudi Monarchy will not survive the next ten years." (We randomly assigned articles to participants.) Then they indicated how much they agreed with the statement that the article presented a sound argument in support of its claim (reiterated in the statement) by marking a 5-point, Likert scale going from Strongly Agree (1) to Strongly Disagree (5). Participants then had an opportunity to make any additional comments before e-mailing their responses back to the experimenter. On average, it took the 22 participants 20.5 minutes to read the articles, with the "Saudi" article taking nearly 10 minutes less to read than the "Health" article ($p = 0.1$, 2-tailed).

Step 2 (Tutorial): Participants read a six-page tutorial describing Toulmin argument structures. The tutorial began by providing a general introduction and the "Harry is a British citizen" example presented above. Then the tutorial gave two increasingly complex examples. For both examples, the information was first presented in a paragraph and then participants were asked to fill in a blank template containing the labeled boxes for each of the six components of the Toulmin structure. After completing the template, the participants could go to the next page to see our answers. All participants completed both templates and we assume (but cannot verify) that they did so before seeing our answers. On average, it took participants 23.5 minutes to complete the tutorial.

Step 3 (Argument structuring and subsequent soundness rating): Participants generated a Toulmin argument structure representing the article they read in Step 1. They received a Toulmin structure template containing the article's stated claim, and blank spaces for the other five components of the structure. They were thus free to write whatever they thought the article offered for the grounds, warrant, backing, qualifier, and rebuttal of its stated claim. After completing their structure, participants again rated the soundness of the argument using the same statement and scale presented in Step 1. In addition, they indicated their agreement with the statement that the Toulmin argument structure was easy to generate for the article they read, again using a 5-point, Likert scale going from Strongly Agree (1) to Strongly Disagree (5). On average, it took the 22 participants 43.8 minutes to complete Step 3. There was no difference statistically in the mean time to structure the two articles.

Step 4 (Rating soundness and understandability of structure for unread article): Participants now received a randomly-assigned argument structure created for the article that they did not read in Step 1. (We randomly determined which two of the 12 participants who read and structured the "Health" article in Step 1 were dropped from Step 4 since only ten participants read and structured the "Saudi" article in Step 1.) After examining the Toulmin structure, participants were asked to rate the soundness of the argument presented in the structure using the 5-point, Likert scale used in Steps 1 and 3. In addition, they were asked to respond to the statement that the Toulmin argument structure was easy to understand, again using a 5-point, Likert scale going from Strongly Agree (1) to Strongly Disagree (5). Lastly, participants had an opportunity to write comments. On average, it took the 20 participants 9.9 minutes to complete Step 4. There was no difference statistically in the mean time to read the structures for the two articles.

Step 5 (Argument soundness rating after reading article): Participants were now asked to read the article for which they had only received the argument structure in Step 4. After reading the article, participants rated the soundness of the article's argument on the 5-point, Likert scale. In addition, they were asked to indicate their agreement with the statement "The argument structure accurately reflected the argument in the article," again using the 5-point Likert scale. Finally, they were given an opportunity to write any comments they wanted to make. On average, it took the 20 participants 28.15 minutes to complete Step 5, and there was again no difference between articles.

Step 6 (Biographical questionnaire): Participants completed a confidential, biographical questionnaire asking about their educational and intelligence analysis background (summarized above), the number of courses they had taken on structured argumentation (explicitly identifying logic and philosophy), how many years they had been performing analysis (with a particular interest in intelligence analysis), and how much prior knowledge they had about the topics discussed in the two articles. The results presented below were not affected by the responses to any of these questions and, therefore, the biographical information is not considered further below.

It is important to note that we were concerned about the possibility of demand characteristics and pretest sensitization (Cherulnick, 2001) prior to conducting the experiment because the procedure of first asking participants to read an article and rate its logical soundness, and then structure and re-evaluate it after a tutorial, conveys the hypothesis that Toulmin structuring might help evaluate argument soundness. However, we concluded that although possible, it was unlikely that demand characteristics and pretest sensitization would affect our results because (1) participants were told they were participating in a study on structured argumentation, (2) as members of the intelligence-analysis community of interest (COI) they were well aware that the intelligence community was funding the development of structured argumentation tools and methods, and (3) as employees in a company that routinely performs evaluations for intelligence-analysis and other government agencies, participants were just as likely (if not more likely) to be overly critical than acquiescent. Given the limited number of available participants from the COI list and the desire to evaluate the effect of Toulmin structuring on both argument evaluation and communication, we (a) took the risk and used the procedures described above, and (b) did not use a control group receiving no structuring.

Articles

The two articles used in the experiment were titled “The Overtreated American” and “Will the Saudi Monarchy Survive.” The first article was a greatly modified version of a magazine article. The second article was distilled from a series of newspaper articles on the Saudi monarchy. (Participants were not told the source or authors of the articles, both of which were used because of their availability.) The articles given to participants were modified so

they appeared to contain reasonable arguments on the surface, but actually had significant holes. For example, neither article contained any rebuttal information. Of particular importance, the Saudi article was crafted to be aligned with the basic elements of Toulmin argument structures. That is, the Saudi article was written in such a way as to make identification of grounds, warrants, etc., relatively straightforward. In contrast, the health care article was not as closely aligned with the six Toulmin structural elements.

3 Hypotheses and Measures

Five principal hypotheses guided the experiment. The first two hypotheses focused on whether the use of Toulmin argument structures facilitated participants' evaluation of argument soundness, the last three hypotheses focused on whether Toulmin structures facilitated argument communication.

Hypothesis 1: Toulmin structures will positively impact analysis. Developing Toulmin structures after reading each article will cause participants to change their opinions about the strength of the articles' claim because structuring will help them uncover the weaknesses in the articles' arguments. This change was measured by the difference in participants' rating for the logical soundness of the article before and after structuring. Support for Hypothesis 1 would be indicated by significantly lower mean ratings of an argument's logical soundness after developing Toulmin argument structures; that is, positive "after minus before" differences in the soundness ratings given the scale.

Hypothesis 2: Toulmin structures will be easy to generate. Regardless of the hypothesized impact on argument evaluation, prior discussions with information analysts suggested that they will not integrate Toulmin argument structures (or any structured argument formalism or tool) into their routine work practices unless the training requirements are minimal and the structures easy to generate, consistent with earlier research (Adelman, Rook, & Lehner, 1985) with decision support and expert system prototypes. The tutorial ensured that participants received minimal training. We also predicted participants would consider Toulmin argument structures easy to generate, as measured by their responses using the Likert scale described in Step 3 of the procedures.

Hypothesis 3: There will be significant variation in participants' Toulmin structures. We deliberately chose to give participants a template that identified only the article's claim and provided blank boxes for the other elements of a Toulmin structure to be consistent with discussions indicating that structuring methods and associated training need to fit naturally into analysts' working environment for the methods to be adopted. However, we predicted that use of such a free-form template would result in significant differences in how participants represented the argument for the same article.

Hypothesis 4: Toulmin structures would represent the articles' argument. Although we predicted that there would be significant variation in participants' Toulmin structures, we still predicted that, on average, the structures would adequately represent the articles' main argument. This hypothesis was based on the assumption that if Toulmin argument structures were effective for communication, participants would need only the structure, not the article, to assess an argument's logical soundness. This hypothesis was measured in two ways. The first measure was the difference in participants' ratings of the logical soundness of the article based on first receiving just the structure (Step 4) and then reading the article (Step 5). If the structures represented the arguments, then there should be no difference in their ratings.

The second measure was participants’ agreement with the statement “The argument structure accurately reflected the argument in the article,” again using the 5-point Likert scale.

Hypothesis 5. Toulmin argument structures will be easy to understand, as measured by responses agreeing with this statement (Step 4).

4 Results

Hypothesis 1: Toulmin structures will positively impact analysis – Partial support. Figure 1 presents the mean ratings of the soundness of the articles' arguments before and after structuring. The mean rating of the health care article's soundness was significantly lower after structuring according to a repeated-measures t-test [mean "after minus before structuring" difference = 0.6, t(11) = 1.87, $p < 0.05$, 1-tailed]. There was no difference in the mean ratings before and after structuring for the Saudi article [mean "after minus before structuring" difference = -0.2, t(9) = -0.80, $p > 0.05$, 1-tailed]. So structuring did affect analysis for the article that was not as closely aligned with elements in Toulmin argument structures.

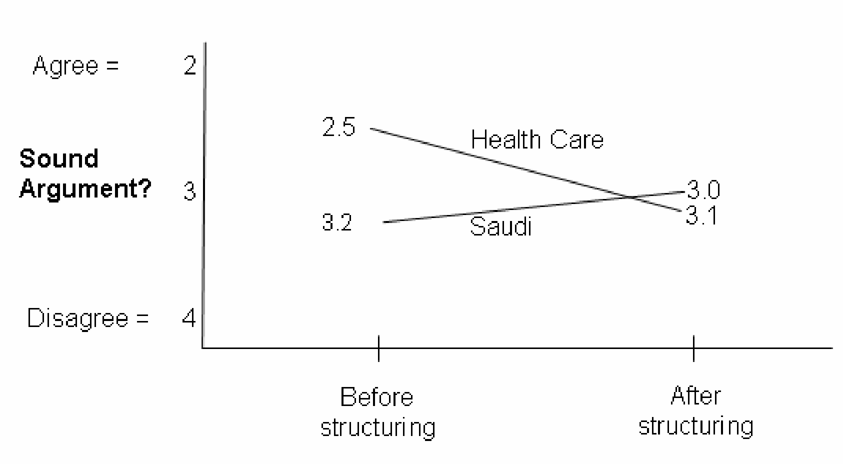


Figure 1: Mean rating of argument soundness before and after structuring.

Hypothesis 2: Toulmin structures will be easy to generate – Not supported. The mean rating for how easy the Toulmin argument structures were to generate for the health care argument was 2.9. It was 3.0 for the Saudi article. Since neither mean value was significantly different than the midpoint (3.0) of the Likert scale, we concluded that the participants did not consider the Toulmin structures easy to generate.

We also correlated participants' ratings for how easy the structures were to generate with (a) the time it took them to read the article and to structure their argument, and (b) the absolute value of the difference in their ratings of the arguments' soundness before and after structuring. None of the former correlations (ease of rating with time measures) were significant, suggesting that ease (or difficulty) of structure generation was measuring a cognitive effort construct, not just time. The latter correlation (ease of structuring with the absolute value of the amount of change in argument soundness ratings) was significant [r =

0.42, n = 22, p = 0.05, 2-tailed], suggesting more cognitive effort was related to a greater change in argument soundness.

Hypothesis 3: There will be significant variation in participants' Toulmin structures – Supported. Table 1 shows how many of each type of elements were generated by participants structuring the Saudi article, as well as the total number of elements and words used in their structures. Although the structures for the Saudi article were more consistent than those for the health care article, one can see that there was substantial variation in the number of grounds, backings, and rebuttals used in the Saudi structures. For example, the number of grounds ranged from 1 to 22, the number of backings from 1 to 10, and the number of rebuttals from 0 to 9. These results suggest that the participants generated very different looking Toulmin argument structures.

Table 1: Numeric Descriptions for Saudi Structures

No. Grounds	No. Qualifiers	No. Claims	No. Backing	No. Warrants	No. Rebuttals	No. Elements	No. Words
22	1	1	5	1	9	39	548
5	1	1	5	1	3	16	362
6	1	1	2	1	1	12	176
1	1	1	6	6	1	16	275
13	1	1	1	1	1	18	382
5	1	1	1	1	3	12	98
10	1	1	10	5	2	27	338
12	1	1	4	2	0	20	301
8	1	1	8	8	1	27	1205
5	1	1	1	1	1	10	167

There was a significant correlation between the number of elements and words in the structures [r = 0.62, n =22, p < 0.01, 2-tailed test], and between both measures and the amount of time to create the structures [r = 0.61 and 0.66, respectively, n =22, p < 0.01, 2-tailed]. However, none of these measures even approached a significant correlation with perceived ease of structure generation or the absolute value of the change in the argument soundness ratings before versus after structuring. These results further support the position that perceived ease of structure generation was a cognitive effort measure.

Hypothesis 4: Structures would represent the articles' argument – Partial support. We predicted that if the structures adequately represented the arguments, there should be no difference in participants' ratings of the arguments' soundness when they received the structure and then read the article. Figure 2 shows this prediction was only supported for the Saudi article [mean "after minus before reading article" difference = -0.2, t(9) = -0.32 p > 0.05, 2-tailed]. The soundness ratings were significantly worse for the health care article using just Toulmin structures, that is, before participants read the article [mean "after minus before reading article" mean = -1.2, t(9) = -4.22, p < 0.01, 2-tailed]. These results suggest that Toulmin structures were an effective communication medium for assessing argument soundness only for the article crafted to fit the elements of Toulmin structures. However, participants' mean response to the statement "the structure accurately reflected the argument in the article" was 2.3 for the health care article and 2.4 for the Saudi article. Both means were significantly different than "3" on the Likert scale, [e.g., for the latter, t(9) = -1.96, p < 0.05, 1-tailed], suggesting that participants thought the structures "accurately" represented the articles' arguments.

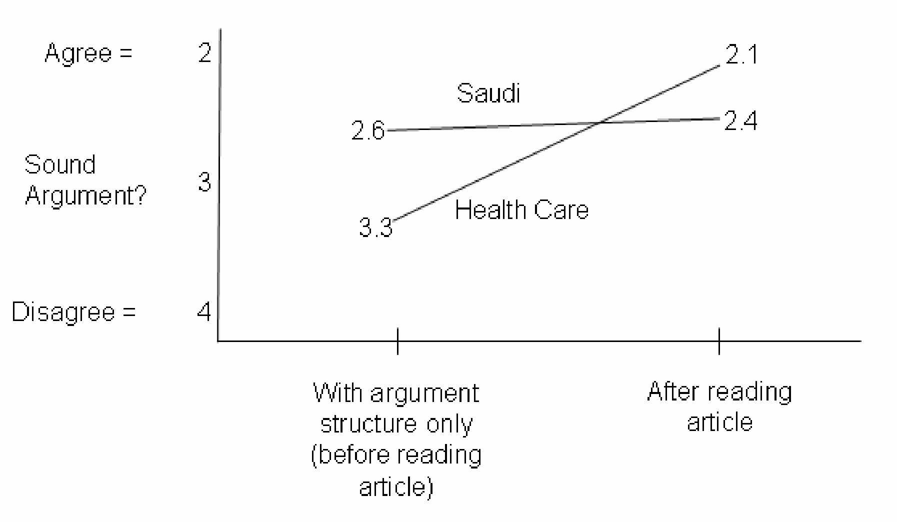


Figure 2: Mean rating of argument soundness with structure only and then article.

Hypothesis 5. Toulmin argument structures will be easy to understand – Partial support. On average, participants thought the structures for the Saudi article were easy to understand [mean = 1.6, which was significantly lower than 3.0, t(9) = -4.59, p < 0.005, 1-tailed test]. In contrast, participants did not think the structures for the health care article were easy to understand [mean = 2.8, t(9) = -0.53, p > 0.05, 1-tailed test]. The difference in the mean understandability ratings for the structures representing the two articles (i.e., 1.6 versus 2.8) approached the traditional 0.05 significance level using a two-tailed test [t(18) = -1.89, p < 0.075, 2-tailed since we did not predict *a priori* that the structures would be easier to understand for one article than another].

After seeing the results for Hypothesis 4 (significantly worse mean soundness ratings for the healthcare, but not Saudi structures) and Hypothesis 5 (that the healthcare structures were harder to understand than the Saudi structures), we thought there might be a negative correlation between (a) the difference in soundness ratings, and (b) how easy the structures were to understand. That is, the worse the argument soundness ratings for the structures than the article (larger negative differences), the greater participants' difficulty in understanding the structures (larger, positive numbers on rating scale). That is what we found [r = -0.68, n = 20, p < 0.01, 2-tailed test]. The negative correlations were more pronounced for the health care (r = -0.85) than Saudi (r = -0.28) structures.

5 Discussion

As noted in the Introduction, there is little evidence demonstrating the value of Toulmin argument structures even though tools are being developed to implement them (e.g., see Moulin, Irandoust, Belanger, & Desbordes, 2002; Sillince & Saeedi, 1999 for reviews). We addressed the need for such evidence by performing a controlled experiment to test whether a structured analytic method using Toulmin argument templates (blank spaces for entering Toulmin argument components) could (a) positively impact the evaluation of argument soundness, and (b) facilitate argument communication. We focused on communication, as well as evaluation, and provided our participants with only minimal training in developing Toulmin structures, because discussions with information analysts indicated that structured analysis methods would need to fit naturally into analysts' working environment to be used voluntarily, consistent with earlier research (Adelman, Rook, & Lehner, 1985) with decision support and expert system prototypes.

We found partial support for the value of Toulmin argument structures. For example, in Part A of the experiment we found Toulmin structures helped participants evaluate the logical soundness of the article about the US health care system, but not the article about the Saudi monarchy. Both articles had substantial flaws in terms of Toulmin argument structures, and the mean soundness ratings for both articles left ample room for improvement on the response scale. The principal difference in how the articles were constructed was that the Saudi article was more closely aligned with the elements of Toulmin structures than the health care article. These results suggest that Toulmin structures may help people critically evaluate articles (or reports) whose arguments are not well structured, in terms of Toulmin structural elements, but further controlled research is required before that conclusion could be reached with confidence. However, it is noteworthy that Toulmin structuring did have a significant and positive impact despite minimal training, that is, despite a weak experimental manipulation.

There also was partial support for the value of Toulmin structures for communication. On the positive side, participants thought the structures accurately reflected the argument for both articles. In addition, there was no difference in their mean argument soundness ratings for the Saudi structures or article (Part B), suggesting that, on average, participants thought the Saudi structures adequately reflected the argument in the Saudi article. Lastly, participants found the Saudi structures easy to understand. On the negative side, however, participants' structures varied greatly, even though they started with the same content. In addition, participants did not find the structures easy to generate, and in the case of the health care structures, did not find them easy to understand. Lastly, the mean argument soundness rating was significantly worse when participants examined the health care structures before reading the article.

We think the lower mean soundness rating for the health care structures (before reading the article) was a direct function of participants’ difficulty in understanding the Toulmin structures, for the lower soundness ratings were related directly to participants’ difficulty in understanding the Toulmin structures. On the other hand, it is possible the lower mean ratings for the health care structures meant the structures helped participants see the weaknesses in the argument—weaknesses that were not as apparent when they later read the article. This alternative interpretation is, however, harder to accept because participants (a) read the article right after examining the structures, so there was minimal time for forgetting argument weaknesses represented in the structure, and (b) did not find the health care structures easy to understand. We think participants said the health care structures accurately reflected the article’s argument because, after reading the article, they better understood the argument and, in hindsight, how the structures represented it. Of course, future research needs to more definitively address these two possibilities.

We make three closing points. First, there was no correlation between participants’ ratings for how easy the structures were to generate and either (a) the time it took them to read the article and to structure their argument or (b) the number of elements or words in the structures. In contrast, perceived ease of structure generation was correlated significantly with the absolute value of the difference in their ratings of the arguments’ soundness before and after structuring. These findings suggest that perceived ease of structure generation was measuring a cognitive effort construct, not simply time, and that more cognitive effort was related to a greater change in ratings of argument soundness

Second, as discussed in the Method section, we were concerned about demand characteristics and pretest sensitization (Cherulnick, 2001) prior to conducting the experiment because the procedure of first asking participants to read an article and rate its logical soundness, and then structure and re-evaluate it after a tutorial, clearly conveyed the hypothesis that structuring might help them evaluate the article’s soundness. However, we considered it unlikely that demand characteristics and pretest sensitization would determine our results because (1) participants were told that they were participating in a study on structured argumentation, (2) as members of the intelligence-analysis community of interest they were well aware that the intelligence community was funding the development of structured argumentation tools and methods, and (3) as employees in a company that routinely performs evaluations for intelligence-analysis and other government agencies, they were just as likely (if not more likely) to be overly critical than acquiescent. Although we can not rule out possible demand characteristics among some participants, the lack of positive effects when evaluating the Saudi structures in Part A, the negative communication effects for the health care structures in Part B, and the correlations demonstrating a cognitive effort construct (as reported above) strongly suggest that, overall, participants expressed their opinions and were not swayed by the experiment’s procedural structure.

Third, we deliberately gave participants minimal training and used free-form templates because prior research indicated structured argumentation methods need to fit naturally into analysts' working environment to be used voluntarily. Our results showing partial support for the hypotheses that Toulmin structuring would improve argument evaluation and facilitate communication suggest that tools using free-form Toulmin templates, deployed with minimal training, will probably not be adequate. Substantial training and prescriptive templates might yield different results. We think this would be unpopular, and presupposes that templates could be prescribed to express analysts' judgment. Nevertheless, it might be worth a try, for even with minimal training, participants were better able to evaluate the soundness of the less well-constructed (health care) argument. The potential evaluation benefit of structured argumentation may be worth the cost of more training and prescribed templates, even if the structures themselves do not facilitate communication. However, before committing to developing tools utilizing this approach, we should do the research and obtain the hard evidence for it first.

References

Adelman, L., Rook, F., & Lehner, P.E. (1985). Users and R&D specialists evaluation of decision support systems: Development of a questionnaire and empirical results. *IEEE Transaction on Systems, Man, and Cybernetics, 15*, 334-342.

Cherulnik, P.D. (2001). *Methods for Behavioral Research*. Thousand Oaks, CA: Sage.

Moulin, B., Irandoust, H., Belanger, M., & Desbordes, G. (2002). Explanation and argumentation capabilities: Toward the creation of more persuasive agents. *Artificial Intelligence Review, 17,* 169-222.

Sillince, J.A.A., & Saeedi, M.H. (1999). Computer-mediated communication: Problems and potential of argumentation support systems, *Decision Support Systems, 26,* 287-306.

Toulmin, S.E. (1958). *The Uses of Argument*. Cambridge: Cambridge University Press.

Van Eemeren, F.H., Grootendorst, R., & Henkemans, F.S. (1996). *Fundamentals of Argumentation Theory*. Mahwah, NJ: Erlbaum.

Wallace, S.P. (1992). A study of argumentative/persuasive writing related to a model of critical thinking in grades nine and eleven (ninth-grade, eleventh grade). *Dissertation Abstracts International, 53,* 3098.