Argumentative Writing Behavior of Graduate EFL Learners

Esmaeel Abdollahzadeh[1,2] · Mohammad Amini Farsani[3] ·
Maryam Beikmohammadi[4]

Abstract This study analyzed the argumentative writing behavior of Iranian graduate learners of English as Foreign Language in their English essays. Further, the correlations between the use of argument elements and overall writing quality as well as soundness of produced arguments were investigated. To this end, 150 essays were analyzed. The sample essays were found to be predominantly deductive in terms of rhetorical pattern. Moreover, they mainly utilized 'data' and 'claim' most frequently with secondary elements of argument (i.e., counterargument claim, counterargument data, rebuttal claim, and rebuttal data) as the least produced elements. Overall writing quality co-varied significantly positively with the uses of claims, data, counterargument claims, counterargument data, rebuttal claims, and rebuttal data. Essays rated high in terms of overall writing quality were further rated for soundness and relevance of the arguments. The results demonstrate that even for advanced language learners good surface structure cannot necessarily guarantee well thought-out logical structure. The pedagogical implications for writing instruction and research are discussed.

Keywords Argumentative writing · Toulmin · Argument structure · Argument elements · Soundness

✉ Esmaeel Abdollahzadeh
eabdol@gmail.com

Mohammad Amini Farsani
mohammad_farsani@yahoo.com

[1]   Iran University of Science and Technology, Tehran, Iran

[2]   University of Exeter, Exeter, UK

[3]   Kharazmi University, Karaj, Iran

[4]   University of Tehran, Tehran, Iran

1 Introduction

Argumentation skill is defined as "a verbal and social activity of reason aimed at increasing (or decreasing) the acceptability of a controversial standpoint for the listener or reader by putting forward a constellation of propositions intended to justify (or refute) the standpoint before a rational judge" (van Eemeren et al. 1996, p. 5). The need to develop argumentation skills through education has been increasingly recognized as essential in academic contexts (Németh and Kormos 2001; Rapanta et al. 2013; Toulmin 1958, 2003; Wolfe 2011). The last two decades witnessed a growing interest in argumentative reading and writing at university level (Helms-Park and Stapleton 2003; Newell et al. 2011; Varghese and Abraham 1998). Both readers' ability to identify, express, and evaluate the underlying structure of argument such as claims, data, and warrants, as well as writers' ability to analyze, compose, and judge an academically sound argument represent a key component of academic success at both undergraduate and graduate levels (Muller Mirza and Perret-Clermont 2009; Newell et al. 2011).

The importance of argumentative writing is further stressed given the recent surge of English as Foreign Language (EFL) learners, including Iranian EFL learners, applying for graduate studies in English-medium universities. Their English argumentative writing abilities are often tested through some internationally recognized tests such as TOEFL, IELTS, and GRE (Educational Testing Service 2009). In the same vein, graduate student writers are highly expected to produce research papers in which they critically engage with the literature and evaluate the status quo on their topic of interest. Poor argumentative skills have been attributed to learners' unpreparedness, unawareness, and limited skills. Unfamiliarity with the typical structure of English argumentative writing may result in developing inadequate and poorly reasoned papers in English (Lunsford 2002; Varghese and Abraham 1998; Wingate 2012).

The work of the British Philosopher, Toulmin (2003, p. vii), in describing the structure of a basic argument has been of central importance in English as a first language contexts (Ong and Zhang 2010), though it has received little attention in L2 settings. His interpretative framework has been widely used in accounting for "the various elements marking the progress of an argument in English argumentative writing" (Qin and Karabacak 2010, p. 455) and in teaching the elements of an argument as well as measuring learners' knowledge of argumentation (Chambliss 1995). Toulmin's (2003, p. 90) 'model of argumentation' consists of six linguistically and semantically interrelated components: claim (a debatable assertion), data (the evidence to support the claim), warrant (assumptions, beliefs and principles of the author), qualifier (the degree of reliance on conclusions arising from arguments), backing (for strengthening the warrant), and rebuttal (the circumstances under which the claim would not be true).

In recent years, Toulminian model of argumentation and its variations has motivated a series of studies which comes under two strands, one pertaining to the assessment of the structural components (Nussbaum and Schraw 2007; Qin and Karabacak 2010) and the other specifically focusing on the soundness of the

produced arguments (Stapleton and Wu 2015; Wolfe et al. 2009). The former involves description and analysis of the surface structure of arguments. Within this perspective, the strength of an argument is based on the presence or absence of specific combinations of Toulmin components. For example, written scripts employing more elements like counterarguments and rebuttals are assumed to be of better overall quality and more convincing than those with less of these elements (Qin and Karabacak 2010). More specifically, the quality of an argument was rated on the basis of presence or absence of possible opposing views, overall language use, overall argument effectiveness, and overall structure. Examining these features, though helpful, may not be particularly relevant for assessing argument soundness as it hinges too much on structural elements of argumentation at the expense of quality of logic and evidence (Crammond 1998; Nussbaum and Kardash 2005; Simon 2008). In other words, this holistic scoring heeds more the strength of writing not its deficiencies (Weigle 2002; White 1985).

The soundness of written arguments has been less acknowledged in the literature (Clark and Sampson 2007; Qin and Karabacak 2010; Rusfandi 2015; Simon 2008; Stapleton and Wu 2015) and thus warrants further investigation. The above literature underscores the need to investigate argumentation elements across various learning and cultural situations, and the extent to which the use of these elements is recognized in EFL students' writings in their educational effort to learn-to-argue. More specifically, there is paucity of empirical research into the quality of reasoning in learners' written arguments (Kuhn and Reiser 2005; Sampson and Clark 2008; Stapleton and Wu 2015; Zeidler 1997). As Stapleton and Wu (2015) argue "the surface structure, or shell of the argument, may appear appropriate or even exemplary, but the actual substance could still be exceedingly weak" (p. 12). Accordingly, a strong incentive for this study is to examine the relationship between the structure of the arguments(in terms of Toulmin's argumentative elements) produced by learners on the one hand, and the soundness of their arguments on the other (Qin and Karabacak 2010; Rapanta et al. 2013). The findings from this study can provide context-specific implications on the impact of university education on the development of logic and reasoning (as one of the main competencies in higher education) among advanced EFL learners of English. The findings can also provide input on how to help L2 graduates develop sound arguments. It is hoped that the findings provide evidence on EFL learners' critical thinking practices and how they integrate structure and meaning in their reasoning.

1.1 Research Questions

In this project, we initially describe the rhetorical organization of the argumentative essays (Research question1) produced by Iranian graduate EFL learners in terms of how they position their arguments. Based on the location of writers' point of view, argumentative papers can be categorized in four ways (Kubota 1998). They may be *deductive* (when a point of view is presented at the beginning of a paper), *inductive* (when a point of view is presented at the end of a paper), *both* (when both deductive and inductive points of view are presented), and *off* (when a writer fails to address a clear position in a paper).

Chen (2001a, 2008) reported that Chinese EFL students used inductive patterns in discourse when writing English papers. However, Hirose (2003), as well as Kobayashi and Rinnert (2008) showed that Japanese EFL students used deductive patterns in discourse when writing English argumentative papers. Further, Rashidi and Dastkhezr (2009) found that Iranian intermediate EFL students tended to present the main ideas at the beginning of their papers, so did undergraduate Turkish EFL learners in Uysal (2008). Overall, except for Chinese learners, these studies provide some evidence for a dominantly deductive pattern in the argumentative discourse of Asian EFL students.

Second, drawing on Toulmin's model of argumentation, we attempt to analyze the typical structure of these essays (Question 2 below). The results can help us identify the elements of the developed arguments (e.g., claims, grounds, and warrants), and thus statistically analyze the relationship between the use of these components and the overall quality of the arguments. Finally, we analyze the highly structured (in terms overall quality) essays in terms of the degree of soundness (i.e. acceptability and relevance) of their arguments. More specifically, the following questions will be investigated:

1. What is the overall organizational structure of English argumentative essays written by Iranian graduate EFL learners?
2. Is there any statistically significant relationship between the use of argumentation elements as evidenced by the revised Toulmin model and the overall quality of English argumentative essays?
3. To what extent well-structured arguments are qualitatively sound in their reasoning?

2 Review of the Related Literature

2.1 Toulminian-Inspired Studies on Argumentative Writing

Second language argumentative writing research addresses a wide range of issues such as learners' perceptions of argumentative writing and the associated instruction they have received (Wingate 2012); comparison of organizational structures of argumentative writing papers across different languages (Kubota 1998; Uysal 2008); the effect of task complexity on students' argumentative writing performance (Ong and Zhang 2010); the relationship between learners' academic achievement and their performance in argumentative writing skills (Preiss et al. 2013; Stapleton 2001); instructional components designed to enhance argumentative writing quality (Varghese and Abraham 1998); and the investigation of how different kinds of arguments are situated in academic contexts (Wolfe 2011).

In the last few decades, an extensive body of research has accumulated in the field of argumentative writing inspired by the works of Toulmin (1958, 2003, p. vii). One set of studies sought to use Toulmin's model as an instructional and methodological instrument to teach argumentative writing both in L1 and L2

contexts (Bacha 2010; Butler and Britt 2011; Lunsford 2002; Varghese and Abraham 1998; Wingate 2012). In these studies, the effectiveness of an explicit instructional approach, the type and quality of instruction learners received, and creating argumentative writing tutorial environments in classrooms have been explored.

A second set of studies has particularly focused on the application of Toulmin's theoretical framework in analyzing argumentative writing (e.g., Németh and Kormos 2001; Nussbaum and Kardash 2005; Wolfe 2011). In these studies, mostly using small sample sizes, the frequency of use of Toulmin's elements across various levels of expertise, the contribution of those elements to the overall quality of argumentative writing, and the influence of goal specification on the use of elements of argument structures have been investigated. It is noteworthy, however, that these studies have been mostly conducted in L1 contexts and very few in L2 contexts (e.g., Qin and Karabacak 2010).

A third set of studies has focused on the soundness of arguments in terms of acceptability, relevance, and adequacy. They have proposed and used some criteria for measuring the soundness of arguments (Hughes and Lavery 2008; Means and Voss 1996). These criteria constitute acceptability (an assertion which is logical to accept as true), relevance (a reason which supports a conclusion), and adequacy (all premises provide enough support to explain a belief in the conclusion). These studies have provided an account of the nature and quality of student-generated written arguments, different schemes and frameworks for evaluating the quality of reasoning, and some deficiencies in operationalizing the quality of reasoning (Kelly and Takao 2002; Lawson 2003; Sandoval 2003; Schwarz et al. 2003; Zohar and Nemet 2002). This study draws on the last two strands of research in argumentation, i.e. examining the structural components of argumentation, as well as the quality and/or soundness of the produced arguments in graduate EFL students' writings.

2.2 The Overall Quality of Argumentative Papers

Very few studies have investigated proficient EFL learners' argumentative writing structure in terms of how a position in an argument is put forward, what types of reasons are specified to support the position, and whether any opposing point of view is offered and refuted (Qin and Karabacak 2010). Answering these questions will inform "… designing of instructional materials and planning of classroom activities for L2 argumentative writing instruction" (Qin and Karabacak 2010, p. 455). Hence, the second question of this study addresses the structural elements of the developed arguments in view of the revised Toulmin model by examining advanced EFL writers' essays in terms of using argument elements. Then, it examines the relationship between the use of these elements and the overall quality of the developed arguments (see 4.2 for more details).

2.3 Soundness of Arguments

Toulminian-inspired studies have been criticized for overemphasizing "structural elements of argumentation at the expense of quality of logic and evidence"

(Stapleton and Wu 2015, p. 13). Several frameworks have been developed to deal with the issue of quality of reasoning in terms of logic and evidence (i.e., argument soundness). For example, Schwarz et al. (2003) assessed quality of arguments in terms of the overall number of reasons, argument types, the acceptability of an argument based on the logical structure, number of reasons supporting counterarguments, and types of reasons.

Other studies focused on domain-specific criteria for evaluating the quality of arguments. For example, Zohar and Nemet (2002) noted that good arguments 'include true, reliable, and multiple justifications' (p. 40). In this way, arguers are likely to generate a simple argument by constructing a claim with at least a single relevant justification. Their framework; however, did not address content issues such as the adequacy, usefulness, and accuracy of a claim. Similarly, Takao and Kelly (2003) devised an analytic scheme to analyze argument quality based on the relative epistemic status of the propositions. Initially a researcher needs to identify the propositions in an argument and then categorize them based on epistemic level. Sandoval and Millwood (2005) measured students' arguments in terms of field-dependent criteria. They focus specifically on conceptual and epistemological quality of students' arguments. Although this framework provides 'the highest mechanical specificity in terms of content quality', it offers less explicit focus on the structure of arguments (Sampson and Clark 2008). Erduran et al. (2004) developed an analytical framework for assessing the quality of arguments in terms of argument complexity level operationalized in terms of the presence and nature of rebuttals.

Surveying the above literature on the overall quality of written arguments reveals several gaps in research on argumentation in L2 writing. First, considerable attention has been given to 'the shell of the argument' (see Qin and Karabacak 2010), and thus more rigorously designed empirical studies to verify the quality of the student-generated arguments is needed. Stapleton and Wu (2015) maintain that "both surface structure and substance need to be considered when assessing the overall quality of an argument essay" (p. 14). They discovered patterns of inadequacies in the reasoning and substance of the highly well-structured essays. Further, Sampson and Clark (2008) emphasize the need to examine the connection between structural components and quality features such as relevance, sufficiency, and accuracy of their content. Hence, the third question of this study explores the soundness of the well-structured arguments.

3 Method

3.1 Participants

250 male and female Iranian graduate learners of TEFL (Teaching English as a Foreign Language) took part in the study. TEFL is considered a social science, and these graduate learners were taking different courses related to learning and teaching of English. They had all obtained a BA degree in English language and literature or in English translation and had passed the Iranian national matriculation exam for entering university. This national exam measures the participants' English

Table 1 Distribution of the study participants (N = 150)

Category	Sub-categories	N	Percent
Gender	Male	58	39
	Female	92	61
Age	22–27	106	71
	28–43	44	29

language proficiency as well as their knowledge of issues in TEFL (i.e. language teaching/learning principles, language testing, and educational linguistics). The participating volunteers came from eleven state universities of reputation across the country. Admission to these universities is more competitive than to other private universities. Thus, the entrants are all assumed to be highly proficient in English. Their formal writing experience is basically limited to two obligatory undergraduate courses, i.e. 'Principles of Writing' and 'Essay Writing'. In their graduate program, they all take 'Academic Writing' in order to help them develop academically sound texts (e.g., term papers, review papers, and dissertations). In most of their modules, these graduate students are encouraged to position their arguments and evaluate the current literature in their assignments and engage with their audience. Although developing sound arguments is highly appreciated, argumentation as a skill may not be explicitly taught in their curriculum. All the state universities across the country with MA in TEFL departments were targeted and contacted. 11 of them eventually agreed to participate in the study. The number of the participants ranged from 7 to 28 in these universities, and their age ranged from 23 to 43. The instructors in these departments agreed to cooperate and get the consent of their students to participate in the study. Participation was voluntary, and they were notified about the purpose of the study, and that they could withdraw from the study anytime they wanted. Consequently, the final participants were 150 as some of them did not complete the writing task as requested and some withdrew from the study (see Table 1).

3.2 Writing Task

The participants were asked to write an argumentative essay in English on a social issue. To select the appropriate topic, we referred to the online database "Opposing Viewpoint Resource Center" published by Thomson Higher Education (http://gale. cengage.com/Opposing Viewpoints). This database is a repository of different controversial topics. The researchers chose 11 topics which seemed appropriate for the purposes of the study. To gain a reasonable justification for choosing the topic, 14 experienced writing instructors were called upon to rate the selected topics on a five-point Likert scale questionnaire, ranging from 1 (the least interesting) to 5 (the most interesting). Based on the instructors' ratings, the topic 'Iran poses a serious threat to the United States vs. Iran does not pose a serious threat to the United States' was finally selected. From among 14 raters, nine of them selected this topic as appropriate for the purposes of the study.

It was also assumed that the participants have adequate background knowledge on the issue given the massive amount of exposure to the media and public debate

about the issue. However, the chosen topic might be emotionally charged for the respondents as all of them came from Iran. To control for this bias, the wording of the topic was reversed for half of the participants. This so-called split-ballot technique is utilized with the expectation that two alternative phrasings of the same topic may yield a more valid picture than will a single phrasing (Revilla and Saris 2013). The writing task involved simple and clearly-worded instructions on how to do the task. It required learners to develop well-organized arguments explaining and supporting their views, and making their position clear on the issue. This was followed by the essay prompt and three blank pages appended.

3.3 Procedure

Prior to data collection, an informed consent letter was given to the instructors and students to participate in the study. The instructors were briefed on the purposes of the study and the data collection procedures. Students were reassured that all the data will be treated confidentially and used for research purposes only, and they could withdraw from the study if they wanted to. All the respondents were given a writing package with simple and consistent instructions on how to do the task. The allotted time (50 min) for writing the essay was decided based on piloting the topic with a small sample of participants similar to the target group. They were asked to develop a balanced argument of at least 400 words in a session based on their background knowledge and personal experience on the selected topic.

The final participants were briefed on avoiding using biased, emotionally-charged or sketchy arguments. They were asked to present opposing views on the issue and come up with their own clear points of view. A uniform procedure for data collection was followed across all the eleven universities. All the instructors were given a script explaining explicitly how to administer the task, and were also contacted and briefed on potential questions they might have about the script and administration procedures. The collected papers were rated by two experienced writing instructors both holistically and analytically.

4 Data Analysis

4.1 The Rhetorical Organization of Argumentative Essays

Results of the Chi square goodness-of-fit test for total frequencies revealed that there were significant differences in the rhetorical pattern of the written arguments ($X^2 = 45.3$, $df = 3$, $p = .00$). Results for question one on the type of the developed arguments showed that about 82% (n = 123) of the participants (see Fig. 1) state their positions clearly either at the beginning (deductive) or at the end (inductive of the essay. Almost half of the essays (49%, n = 73,) presented a clear point of view deductively supported by reasons and pieces of evidence. Around 20% (n = 28) of the essays were Off-type, i.e., failed to adopt a clear point of view and the stance on the topic was neutral. They stated the pros and cons of the topic without taking a stance. A small portion of the essays (18%, n = 27) had an *Inductive* rhetorical
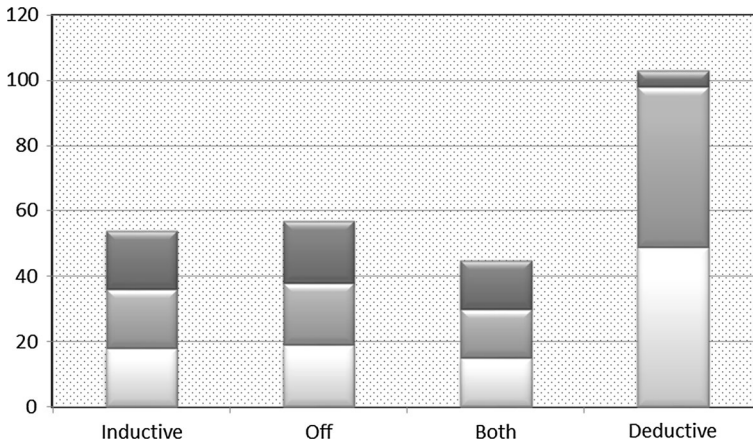
Fig. 1 The percentage of English argumentative essay types

organization which summarized the writer's thesis at the end preceded by supporting reasons and pieces of evidence. Moreover, a smaller percentage was of *Both* type (15%, n = 22,) in which the writer stated the topic of interest at the outset and the writers' stance was maintained until the very end of the essays. In sum, the essay types were *Deductive* (49%), *Off* (19%), *Inductive* (18%), and *Both Inductive and Deductive* (15%) in terms of the overall rhetorical organization.

4.2 Argument Elements and Overall Argument Quality

The second question of the study involved two stages: structural analysis of the arguments, and examining the correlation between the use of argument elements and overall argument quality. To examine the essays structurally, elements of arguments (i.e. claim, counterclaim, and rebuttals and their associated supporting reasons such as rebuttal claim/data) were identified and their frequencies were calculated. We drew on Qin and Karbacak's (2010) rubric for identifying these elements which was originally based on Toulmin (2003), Nussabaum and Kardash (2005), Nussabaum and Shraw (2007; see "Appendix 1" for definitions and examples from the corpus). This rubric has been shown to be reliable in identifying argument elements (Qin and Karabacak 2010). Inter-rater reliability of rated essays for claim, data, counter-argument claim, counter-argument data, rebuttal claim, and rebuttal data was .91, .96, .86, .84, .85, and .87 respectively, and the overall inter-rater reliability was .87. In case of any discrepancy in the identification of argumentative elements between the raters, data were negotiated until a consensus was achieved. We also drew upon the semantic structure and linguistic elements in the produced texts following Qin and Karabacak (2010). Claims, for example, were identified through elements such as *in my opinion*, *I believe*, and *I think*; datathrough prepositional phrases such as *for that reason* and subordinators such as *because*; counterargument and rebuttal through certain phrases such as *Some people claim that…. However;* It is *said that….. but; even though; despite;* and *although*.

Table 2 Frequency of use of argument elements

Argument elements	N	Mean	SD	Rank
Data	150	2.29	1.051	1
Claim	150	1.30	.565	2
CAclaim	150	.49	.528	3
Reclaim	150	.34	.475	4
CAdata	150	.16	.368	5
Redata	150	.08	.272	6

CA claim Counterargument claim, *CA data* counterargument data, *Reclaim* rebuttal claim, *Redata* rebuttal data

The results show that graduate EFL writers tend to use all the elements of argument structure in their writing (see Table 2). As can be seen, the highest mean scores relate to the fundamental elements of Toulmin model, i.e. 'data' and 'claims'. The lowest means relate to secondary elements (i.e. counterargument claims, counterargument data, rebuttal claim, and rebuttal data). Rank-ordered frequency of use of the elements is as follows: counterargument claim > rebuttal claim > counterargument data > rebuttal data.

The Chi Square test on the relationship between the use of primary (Mean = 3.58; SD = 1.36) and secondary (Mean = 1.06; SD = 1.19) elements indicated that the primary ($\chi 2 = 150.840$, $df = 8$, $p = .00$), and secondary elements ($\chi 2 = 92.497$ $df = 4$, $p = .00$) tend to be systematically related, i.e. the frequency of use of the two elements is significantly correlated. Further, writers tend to use significantly more primary than secondary elements.

Similarly, the Chi Square test of independence indicated that counterargument claim and rebuttal claim ($\chi 2 = 66.21$, $df = 3$, $p = .00$), and counterargument data and rebuttal data, ($\chi 2 = 1.47$; $df = 2$, $p = .00$) tend to be systematically related as well, i.e. the frequency of use of the two and their corresponding elements is significantly correlated in advanced EFL writing. Moreover, these writers employed significantly more claims than data (82% counterargument claim vs. 24% counterargument data).

To answer the second part of question 2, overall argument quality of the essays was examined following Nussbaum and Schraw (2007), and Qin and Karabacak (2010). This involved grading the essays holistically by two raters in terms of 'the overall argument effectiveness', 'the presence or absence of the possible opposing views', 'overall structure', and 'overall language use'. These criteria served as general indicators of an effective argument. The raters were notified not to bias towards any of these three dimensions of quality when rating the papers (Qin and Karabacak 2010). To ensure consistency of ratings, two raters, scored 20 randomly-selected essays using the rubric. Then, the authors discussed vague points of the rubric until consensual agreement was reached. The inter-rater reliability was found to be .88.

Table 3 Descriptive statistics for the overall quality of the essays

			Mean	SD	Min	Max
Overall	Quality	Score	2.98	1.058	1	5

Table 4 Correlation between the uses of argument elements and overall writing quality

	Claim	Data	CAclaim	CAdata	Reclaim	Redata
Holistic writing scores	.29**	.47**	.246**	.19*	.28**	.19*

* significant at .05 level

** significant at .01 level

Pearson Product-Moment Correlation coefficient revealed that the essay scores of overall quality co-varied significantly positively with the uses of the six elements of arguments (see Tables 3, 4).

Among the argument elements, 'data' correlated significantly with the overall writing performance (r = .47). All other elements correlated significantly positively with participants' writing quality: claim (r = .32 $p < .01$), followed by rebuttal claim (r = .28 $p < .01$), counter argument claim (r = .24, $p < .05$), counter argument data (r = .19, $p < .05$.), and rebuttal data (r = .19 $p < .05$). This means that the more these elements are used, the higher the overall quality of the argumentative essays.

4.3 Analysis of Argument Soundness

We found that the above criteria for measuring overall argument quality per se do not take account of the criteria (such as acceptability, relevance, and adequacy) for measuring soundness of arguments (Means and Voss 1996; Rapanta et al. 2013; Rusfandi 2015; Stapleton and Wu 2015). Thus, to answer the third study question, we adopted an integrative analytic approach following Stapleton and Wu (2015) to attend to both the structure and substance of written arguments (see "Appendix 2"). They maintain that "for an argumentative essay to be persuasive, not only must it follow surface structure by including alternative viewpoints and showing their weaknesses, but it must also support claims with good quality reasons that convince others" (p. 22).

The rubric used contains descriptors of the surface structure (i.e. surface elements of arguments described in 4.2), and the quality of supporting reasons demonstrating the magnitude of soundness. It involves two broad levels, and a total of five sublevels with attributed scores. In this rubric, the six elements of argumentation are differentially weighted from a scale of 0 to 5 for claim and scale of 0 to 10 for the two categories of counter-argument claim and rebuttal claim, as well as a scale of 0, 10, 15, 20, and 25 for the three categories of data. Increased scores were given to the categories of data because they demand higher level of critical thinking and argumentation skills (Stapleton and Wu 2015).

To analyze the reasoning quality of the arguments, the following procedures for selection of scripts were taken: First, following Nussbaum and Kardash (2005) and Nussbaum and Schraw (2007), we selected only those scripts that received high scores in the second phase of analysis explained in 4.2. That is, essays whose overall argument quality was plus one standard deviation above the mean were selected (Crossley et al. 2014; McNamara et al. 2013). The reason for the 'best evidence selection' is to discover whether essays rated high in terms of argument quality in the second phase would still be highly rated in terms of argument relevance and

soundness. This filtering resulted in 40 scripts (out of 150). Then, each script was scored independently based on Stapleton and Wu's (2015) analytic scoring rubric for argumentative soundness. To ensure that student-articulated scripts were assessed reliably, all the 40 scripts were coded independently by two raters, who were briefed to adhere to the rubric. Points of disagreement were sorted out between raters through discussion until consensual agreement was reached. The overall agreement between them was .91. Analysis of the 40 high-rated scripts revealed three argumentative profiles presented below.

4.3.1 Profile 1: One-Sided Good Surface Structure But Weak Argument Quality: Failure to Include Counterargument

The first profile represents scripts (n = 20) with one-sided argument in terms of structure as they simply depicted a claim with at least one reason. The arguments were weak in terms of soundness as very few reason(s; 20%, n = 4) provided were acceptable and relevant to the claim. Table 5 presents the frequencies of the data with the corresponding claims in the selected scripts. The quality of the supporting reasons for the claim ranged from *no relevant reason* to *multiple sound reasons*. As shown in Table 5, the majority of the one-sided arguers (60%, n = 12) supplied a claim with one or two reasons for each claim; some reasons were sound and acceptable and some weak and irrelevant though. A small percentage of the arguers (15%, n = 3) provided only one reason for the claim and the reasons provided were weak and irrelevant. However, about 15% (n = 3) of the writers provided multiple reasons most of which were acceptable and free of irrelevancies for the claim. Only a small number (5%, n = 1) of the arguers failed to provide an acceptable reason. Interestingly, only 5% of the arguers succeed in producing very strong data for supporting claims. In sum, 20% (n = 4) of the scripts in this cluster received higher scores for soundness, i.e. used multiple reasons that were relevant, acceptable, and justified.

4.3.2 Profile 2: Two-Sided Good Surface Structure But Weak Argument Quality (No Rebuttals)

The second profile represents a two-sided argument as it includes argument-counterarguments (without rebuttal). Structurally speaking, the scripts containing

Table 5 Frequency of features of one-sided arguments used in the scripts (N = 20)

Quality of reasoning		Frequency	Percent
Level 5	Provide multiple sound reasons/free of irrelevancies	1	5
Level 4	Provide multiple reasons; most of them are acceptable	3	15
Level 3	Provides one to two reasons/some are weak and some are sound	12	60
Level 2	Provides only one reason/weak or irrelevant reason	3	15
Level 1	Provides no relevant reason to support	1	5
Total		20	

Table 6 Frequencies features of two-sided arguments used in the scripts (N = 12)

Argument quality	Claim–data (C–D)		Counter argument claim (CC)–counterargument data (CD)	
	Frequency	Percent	Frequency	Percent
Provide multiple sound reasons/free of irrelevancies	0	0	1	9
Provide multiple reasons; most of them are acceptable	3	25	1	8
Provides one to two reasons/some are weak and some are sound	3	33	3	25
Provides only one reason/weak or irrelevant reason	3	25	4	33
Provides no relevant reason to support	2	17	3	25
Total	11		12	

these features (n = 12) present not only the writers' assertion and its corresponding reasons but also the possible opposing views with their corresponding justification.

Table 6 shows that the arguers supplied a claim (33%, n = 4) and counter-argument claim (25%, n = 3) with one to two reasons for each claim. Almost 25% (n = 3) of the writers provided multiple reasons most of which were acceptable and free of irrelevancies for the claim; over 8% (n = 1) of the arguers provided multiple acceptable and sound reasons for the counter-argument claim. For both claim and counter-argument claim, the arguers provided only one reason and the reason provided was weak and irrelevant (25 and 33.3% respectively). Interestingly, multiple data in both claim and counter-argument claim components had the lowest frequencies: 0 and 8%, respectively. Less than 10% (n = 1) of the arguers succeeded in producing very strong data for supporting counterargument claim(s), i.e. they used multiple reasons that were relevant and acceptable for the claim and counterclaim. None of the arguers succeeded in producing very strong data for supporting claim.

4.3.3 Profile 3: Two-Sided Good Surface Structure But Weak Argument Quality (Rebuttal(s) Included)

This profile represents a high level of structural quality as it consists of all the six elements of argumentation discussed before. However, the overall reasoning quality is still far from satisfactory. As can be seen (Table 7), almost all the claims, counterargument claims, and rebuttal claims were supported with low frequencies of data. The arguers provided only one reason for claim, counter-argument claim, and rebuttal claim and the reason provided was weak and irrelevant. 13% of the arguers (n = 1) provided multiple reasons which were mostly acceptable and free of irrelevancies for the claims and counter-argument claims; the arguers failed to provide sound reasons for the rebuttal claim. This means that the more the better is not always the case in developing sound arguments.

Table 7 Frequencies of features of two-sided arguments used in the scripts (N = 8)

Argument quality	C–D		(CC–CD)		Rebuttal claim–Rebuttal data	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Provide multiple sound reasons/ free of irrelevancies	0	0	0	0	0	0
Provide multiple reasons; most of them are acceptable	1	13	1	13	0	0
Provides one to two reasons/some are weak and some are sound	3	37	1	13	0	0
Provides only one reason/weak or irrelevant reason	3	37	3	37	4	50
Provides no relevant reason to support	1	13	3	37	3	38
Total	8		8		7	

5 Discussion

The graduate learners organized most of their arguments in a deductive fashion. In line with most of the previous research (Hirose 2003; Kobayashi and Rinnert 2008; Rashidi and Dastkhezr 2009), their writing mainly falls into the dominant categories of *Deductive*, *Off*, *Inductive*, and *Both deductive and Inductive* patterns, respectively. The majority of the participants (82%) state their positions clearly either at the beginning (deductively, 49%) or at the end (inductively, 18%) or Both (15%). These advanced learners seem to adhere to English writers' use of a deductive pattern. Their formal training over the years and academic writing experience in English, might lean them to approximate English writers' style in terms of directedness in their essays. It could also be due to the dominant product-based mode of writing instruction in Iran at tertiary level in which more emphasis is based on stating the main claim at the beginning (Abdollahzadeh 2010). The rather minimal employment of the inductive pattern might show the impact of schooling on raising the writers' awareness of this pattern as a non-English pattern (Chen 2008; Husin and Ariffin 2012). The rather low use of *Both* organizational patterns, on the other hand, may be attributed to the writers' poor development of the arguments.

Participants employed all the elements of written argumentation; however, the extent of using basic and secondary components, though interdependent, was significantly different. The majority of the essays included basic elements, namely, the writer's opinion (claim) and supporting evidence (data). These elements are the most preferred ones for learners (De Bernardi and Antolini 1996; Lunsford 2002; Qin and Karabacak 2010; Varghese and Abraham 1998). The current findings reveal a predisposition among Iranian graduate EFL writers not to present much counterargument and rebuttals in their written argumentations, despite the fact that good arguments involve counterarguments and rebuttals to augment writing quality (Nussbaum and Kardash 2005; Wolfe et al. 2009). About half of the argumentative

essays applied some form of rebuttals and counterarguments. A probable explanation might be the cognitive constraints of developing the secondary elements as they are more complex to produce (Coirier et al. 1999; Wolfe et al. 2009), and their lack of experience and awareness of the effectiveness of these elements in argument quality. Further, they might perceive of secondary elements as optional or unnecessary for writing argumentatively. Counterarguments play a vital role in argumentation structure (Toulmin 2003). A significant majority of them failed to represent a critical and reflective positioning towards the topic which could subsequently affect the quality of their arguments in their academic assignments in their attempts to argue to learn and later join their respective academic community. We found that despite good surface structure, many reasons provided by the arguers were weak resonating Sadler's (2004) contention that arguers might not typically "display high-quality written argumentation as defined by an ability to articulate and defend contentious positions" (p. 523).

Using an integrated assessment framework, we assessed arguments in terms of substance and structure (Sampson and Clark 2008; Stapleton and Wu 2015). The analysis revealed several patterns of argumentative behavior. First, the selected scripts (n = 40) were grouped in terms of the occurrence of double surface structure (cluster 1: claim–data), quadruple surface structure (cluster 2: claim–data–counter-argument claim–counter–argument data), and sextuple (cluster 3: claim–data–counterargument claim–counter–argument data–rebuttal claim–rebuttal data) combinations.

The first argumentative writing profile (cluster 1) showed structurally well-designed essays which were significantly low in terms of argumentation quality. All these scripts were described as one-sided argument, containing claim(s) and one or more reasons. The arguers here tended to support their claim(s) and maintain their position with some reasons (ranging from one to multiple pieces of evidence (Sampson and Clark 2008). However, claim-data argument is the least sophisticated form of an argument (Rusfandi 2015). Despite well-designed surface structure, most of the arguers failed to provide relevant and acceptable reasons for the corresponding claim(s) proving that student-generated arguments often lack substance, including components that are inaccurate and/or irrelevant in terms of quality (Rapanta et al. 2013; Schwarz et al. 2003; Simon 2008; Sampson and Clark 2008).

The second writing profile was also well-constructed structurally but tended to be weak in terms of quality. The arguers in this profile tended to support their claim(s) and counter-argument claims with either data (from one to multiple pieces of evidence), dismissing rebuttals and maintaining their position. Structurally, with increasing argument components, the written arguments become more complex and sophisticated. Despite producing more complex and sophisticated arguments compared to those in the first cluster, the students were not able to adequately align rebuttals with the counterarguments and thus failed to refute them. This profile cannot be considered strong as there was no indication of including rebuttals as essential components of better quality arguments (Erduran et al. 2004). This could be attributed to the complex nature of the argument-counterargument structure in L2 (Qin and Karabacak 2010) as well as "risk avoidance, lack of confidence, and

reformulation difficulties in producing argument-counter-argument claims and supported data" (Kobayashi and Rinnert 2008, p. 35). There is some evidence that arguers who did not provide a counter-argument in their English L2 essays included this when they wrote in L1 (Kobayashi and Rinnert 2008). Further research can help us discover the extent to which transfer of argumentation strategies occurs in L2 argumentative production. The impact of schooling and L1 educational and writing culture could be another factor accounting for L1–L2 differences in argumentative development.

The final argumentative profile embraced rebuttals (claim and data) and was far more effective than those without rebuttals. Nonetheless, although appropriate surface structures were followed, the overall reasoning quality of argumentation was still far from satisfactory. The findings demonstrate that good surface structure cannot necessarily guarantee well thought-out logical structure. Therefore, although they were linguistically competent, a significant majority of the graduate students still could not produce arguments backed by evidence and counter-evidence. Raising the graduate student's consciousness about the value of justifying claims and inquiry into the argument, and attending opposing perspectives could be an important pedagogical practice to pursue both in EFL and L1 contexts (Sadler 2004).

6 Conclusion

The findings provide empirical evidence uncovering the gap between structure and substance in linguistically advanced student argumentative discourse. The reasoning quality of the student-generated arguments was generally weak, though the frequency of use of argument elements was rather high. Forming a good surface structure and sound argument quality is believed to contribute substantially to persuasiveness of argumentative writing (Sampson and Clark 2008; Stapleton and Wu 2015). Assessing written productions of students, then, requires using integrative evaluation criteria in which more weight is given to arguments discussing alternative viewpoints, while taking account of good quality reasons to support claims in an attempt to convince the audience.

Given that enhancing argumentative skills is an ambitious educational goal which needs time and practice (Means and Voss 1996; Sadler 2004), writing programs and writing-across-the curriculum instructors need to create ample opportunities for learners to take part in argumentative practices in which they can evaluate and provide justified explanations for their claims and assertions, attending contradictory claims, and the formation of counterarguments and rebuttals (Sampson and Clark 2008; Rusfandi 2015). This argumentative intervention can develop EFL learners' critical thinking skills as people seem to learn better when they argue (Kuhn 2008), and thus help them understand the epistemic nature of knowledge and participate more effectively in their respective scientific discourse. After all, learners need to appreciate how to produce a cognitively mature argument and contextualized arguments using rebuttals, counterarguments and qualifiers (Schwarz 2009).

We mainly explored the persuasive discourse of advanced EFL learner writers. Students' topical knowledge or lack thereof in the discipline was not investigated in this study. Further, we know that textual practices are contextualized within the genres unique to different disciplinary communities where knowledge-making and knowledge-sharing are central. The implication for disciplinary writing is that understanding ways of knowing in different sciences is informed by the 'tool kit' acquired through learning to argue (Newell et al. 2015). Therefore, students can use this toolkit to understand what knowing means in their discipline, and later argue to learn their way into the discipline's texts and ways of knowing (Carter 2007).

There might be other factors impacting the quality and structure of the students' argumentative essays such as use of metadiscourse, evaluation markers, hedging, argumentative strategies across L1 and L2, as well as norms of suitable argumentation behavior across various disciplines (Sampson and Clark 2008; Uccelli et al. 2013). Examining these issues was outside the scope of this study. Future research can determine the contributory role of these elements in producing sound arguments.

Appendix 1

See Table 8.

Table 8 Definitions and examples of argumentation elements used in student essays

Elements	Definition with examples from the corpus
Claim	An assertion in response to a contentious topic or problem
	Example
	Iran is not an essential threat to the United States
Data	Evidence to support a claim. It can take various forms, such as facts, statistics, anecdotes, research studies, expert opinions, definitions, analogies, and logical explanations
	Example
	Because this group believes that Iran is using the nuclear program in a friendly manner
	This group tries to convince the world that no non-humanistic action ever occurred by Iran
Counterargument Claim	The possible opposing views that can challenge the validity of a writers' claim; these opposing views can be supported by data (Németh and Kormos 2001)
	Example
	However, some people believe that Iran is a threat to the United States
	On the other hand, Iran is trying to show supremacy toward the US
Counterargument Data	Evidence to support a counterargument claim
	Example
	For example, in 1979 an amazing political bang occurred which was considered as a threat to the United States and its friends

Table 8 continued

Elements	Definition with examples from the corpus
Rebuttal Claim	Statements in which the writer responds to a counter-argument by pointing out the possible weakness in the claim, data, or warrant, such as logical fallacies, insufficient support, invalid assumptions, and immoral values
	Example
	But this is one side of the story, and it is not going to end here. I reckon such similar countries with their harsh discourses, are considered great threats to other countries on the earth
Rebuttal Data	Evidence to support a rebuttal claim
	Example
	America and Iran with their own ideologies and great differences in their attitudes and politics can threaten other countries as well

Appendix 2

See Table 9.

Table 9 Analytic scoring rubric for argumentative writing (adapted from Stapleton and Wu 2015)

Appendix B: Analytic scoring rubric for argumentative writing (Adapted from Stapleton & Wu, 2015)

1. Claim(s) (5%)	Score: 5			Score: 0	
	States point(s) of view			Does not state point(s) of view	
2. Data (25%)	Score: 25	Score: 20	Score: 15	Score: 10	Score: 0
	a. Provides multiple reasons for the claim(s), and b. All reasons are sound/acceptable and free of irrelevancies	a. Provides multiple reasons for the claim(s), and b. Most reasons are sound/acceptable and free of irrelevancies, but one or two are weak	a. Provides one to two reasons for the claim(s), and b. Some reasons are sound/acceptable, but some are weak or irrelevant	a. Provides only one reason for the claim(s), or b. The reason provided is weak or irrelevant	a. No reasons are provided for the claim(s); or b. None of the reasons are relevant to/support the claim(s)
3. Counterargument Claim(s)/Alternative Point(s) of View (10%)	Score: 10			Score: 0	
	Provides counterargument claim(s)/alternative view(s)			Does not provide counterargument claim(s)/alternative view(s)	
4. Counterargument Data/Supporting Reasons for Alternative Point(s) of View (25%)	Score: 25	Score: 20	Score: 15	Score: 10	Score: 0
	a. Provides multiple reasons for the counterargument claim(s)/alternative view(s), and b. All counterarguments/reasons for the alternative view(s) are sound/acceptable and free of irrelevancies	a. Provides multiple reasons for the counterargument claim(s)/alternative view(s), and b. Most counterarguments/reasons for the alternative view(s) are sound/acceptable and free of irrelevancies, but one or two are weak	a. Provides one to two reasons for the counterargument claim(s) /alternative view(s), and b. Some counterarguments/reasons for the alternative view(s) are sound/acceptable, but some are weak or irrelevant	a. Provides only one reason for the counterargument claim(s)/alternative view(s), or b. The counterargument/reason for the alternative view is weak or irrelevant	a. No reasons are provided for the counterargument claim(s)/alternative view(s); or b. None of the reasons are relevant to/support the counterargument claim(s)/alternative view(s)
5. Rebuttal Claim(s) (10%)	Score: 10			Score: 0	
	Provides rebuttal claim(s)			Does not provide rebuttal claim(s)	
6. Rebuttal Data* (25%)	Score: 25	Score: 20	Score: 15	Score: 10	Score: 0
	a. Refutes/points out the weaknesses of all the counterarguments, and b. All rebuttals are sound/acceptable c. The reasoning quality of all the rebuttals are stronger than that of the counterarguments	a. Refutes/points out the weaknesses of all the counterarguments, and b. Most rebuttals are sound/acceptable, but one or two are weak c. The reasoning quality of most rebuttals are stronger than that of the counterarguments, while one or two are equal to that of the counterarguments	a. Refutes/points out the weaknesses of all the counterarguments, and b. Some rebuttals are sound/acceptable, but some are weak c. The reasoning quality of some rebuttals are stronger than that of the counterarguments, while some are weaker than that of the counterarguments	a. Refutes/points out the weaknesses of some counterarguments, or b. Few of the rebuttals are sound/acceptable; most of them are weak, or c. The reasoning quality of most rebuttals are weaker than that of the counterarguments	a. No rebuttals are provided; or b. None of the rebuttals can refute the counterarguments

References

Abdollahzadeh, E. 2010. Undergraduate Iranian EFL learner's use of writing strategies. *Writing and Pedagogy* 2(1): 65–90.

Bacha, N. 2010. Teaching the academic argument in a university EFL environment. *Journal of English for Academic Purposes* 93: 229–241.

Butler, J.A., and M.A. Britt. 2011. Investigating instruction for improving revision of argumentative essays. *Written Communication* 28(1): 70–96.

Butt, D., R. Fahey, S. Feez, S. Spinks, and C. Yallop. 2000. *Using functional grammar: An explorer's guide*. Sydney: Macquarie University.

Carter, M. 2007. Ways of knowing, doing, and writing in the disciplines. *College Composition and Communication* 58: 385–418.

Chambliss, M.J. 1995. Text cues and strategies successful readers use to construct the gist of lengthy written arguments. *Reading Research Quarterly* 30(4): 778–807.

Chen, J. 2008. An investigation into the preference for discourse patterns in the Chinese EFL learning context. *International Journal of Applied Linguistics* 182: 188–211.

Chen, J.P. 2001a. Markedness in intercultural discourse: A study of Chinese EFL students' discourse patterns. PhD thesis, in ERIC, RIE June 2001 searching codes: Chen, Jianping/ED448593.

Clark, D.B., and V. Sampson. 2007. Personally-seeded discussions to scaffold online argumentation. *International Journal of Science Education* 29: 253–277.

Coirier, P., J.E.B. Andriessen, and L. Chanquoy. 1999. From planning to translating: The specificity of argumentative writing. In *Foundations of argumentative text processing*, ed. P. Coirier, and J. Andriessen, 1–28. Amsterdam: Amsterdam University Press.

Crammond, J.G. 1998. The uses and complexity of argument structures in expert and student persuasive writing. *Written Communication* 152: 230–268.

Crossley, A.S., R. Roscoe, and S.D. McNamara. 2014. What is successful writing? An investigation into the multiple ways writers can write successful essays. *Written Communication* 31(2): 184–214.

De Bernardi, B., and E. Antolini. 1996. Structural differences in the production of written arguments. *Argumentation* 10: 175–196.

van Eemeren, F.H., R. Grootendorst, A.F. Snoeck Henkemans, J.A. Blair, R.H. Johnson, E.C.W. Krabbe, C. Plantin, D.N. Walton, C.A. Willard, J. Woods, and D. Zarefsky. 1996. *Fundamentals of argumentation theory. A handbook of historical backgrounds and contemporary developments*. Mahwah, NJ: Lawrence Erlbaum.

Erduran, S., S. Simon, and J. Osborne. 2004. Tapping into argumentation: developments in the application of Toulmin's argument pattern for studying science discourse. *Science Education* 88: 915–933.

Helms-Park, R., and P. Stapleton. 2003. Questioning the importance of individualized voice in undergraduate L2 argumentative writing: An empirical study with pedagogical implications. *Journal of Second Language Writing* 12(3): 245–265.

Hirose, K. 2003. Comparing L1 and L2 organizational patterns in the argumentative writing of Japanese EFL students. *Journal of Second Language Writing* 12: 181–209.

Hughes, W., and J. Lavery. 2008. *Critical thinking: An introduction to the basic skills*, 5th ed. Ontario: Broadview Press.

Husin, M.S., and K. Ariffin. 2012. The rhetorical organization of English argumentative essays by Malay ESL students: The placement of thesis statement. *The Journal of Asia TEFL* 9(1): 147–169.

Kelly, G.J., and A. Takao. 2002. Epistemic levels in argument: An analysis of university oceanography students' use of evidence in writing. *Science Education* 86(3): 314–342.

Kobayashi, H., and C. Rinnert. 2008. Task response and text construction across L1 and L2 writing. *Journal of Second Language Writing* 17: 7–29.

Kubota, R. 1998. An investigation of L1–L2 transfer in writing among Japanese university students: Implications for contrastive rhetoric. *Journal of Second Language Writing* 7: 69–100.

Kuhn, D. 2008. *Education for thinking*. Cambridge, MA: Harvard University Press.

Kuhn, L., and Reiser, B. 2005. Students constructing and defending evidence-based scientific explanations. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Dallas, TX.

Lawson, A. 2003. The nature and development of hypothetico-predictive argumentation with implications for science teaching. *International Journal of Science Education* 25(11): 1387–1408.

Lunsford, K.J. 2002. Contextualizing Toulmin's model in the writing classroom: A case study. *Written Communication* 19(1): 109–174.

McNamara, D.S., S.A. Crossley, and R. Roscoe. 2013. Natural language processing in an intelligent writing strategy tutoring system. *Behavior Research Methods* 45(2): 499–515.

Means, M.L., and J.F. Voss. 1996. Who reasons well? Two studies of informal reasoning among children of different grade, ability, and knowledge levels. *Cognition and Instruction* 14: 139–178.

Muller Mirza, N., and A. Perret-Clermont (eds.). 2009. *Argumentation and education: Theoretical foundations and practices*. New York: Springer.

Németh, N., and J. Kormos. 2001. Pragmatic aspects of task-performance: the case of argumentation. *Language Teaching Research* 5: 213–240.

Newell, G.E., R. Beach, J. Smith, and J. VanderHeide. 2011. Teaching and learning argumentative reading and writing: A review of research. *Reading Research Quarterly* 46(3): 273–304.

Newell, G.E., D. Bloome, and A. Hirvela. 2015. *Teaching and learning argumentative writing in high school English Language Arts classrooms*. New York: Routledge/Taylor & Francis Group.

Nussbaum, E.M., and C.M. Kardash. 2005. The effects of goal instructions and texts on the generation of counterarguments during writing. *Journal of Educational Psychology* 97(2): 157–169.

Nussbaum, E.M., and G. Schraw. 2007. Promoting argument - counterargument integration in students' writing. *The Journal of Experimental Education* 76: 59–92.

Ong, J., and L.J. Zhang. 2010. Effects of task complexity on the fluency and lexical complexity in EFL students' argumentative writing. *Journal of Second Language Writing* 19(3): 218–233.

Preiss, D.D., J.C. Castillo, E.L. Grigorenko, and J. Manzi. 2013. Argumentative writing and academic achievement: A longitudinal study. *Learning and Individual Differences* 28(3): 204–211.

Qin, J., and E. Karabacak. 2010. The analysis of Toulmin elements in Chinese EFL university argumentative writing. *System* 38(3): 444–456.

Rapanta, C., M. Garcia-Mila, and S. Gilabert. 2013. What is meant by argumentative competence? An integrative review of methods of analysis and assessment in education. *Review of Education research* 83(4): 483–520.

Rashidi, N., and Z.A. Dastkhezr. 2009. A comparison of English and Persian organizational patterns in the argumentative writing of Iranian EFL learners. *Journal of International Language English* 8: 131–152.

Revilla, M., and W.E. Saris. 2013. The split-ballot multitrait-multimethod approach: Implementation and problems. *Structural Equation Modelling: A Multidisciplinary Journal* 20(1): 27–46.

Rusfandi, R. 2015. Argument-counterargument structure in Indonesian EFL learners' English argumentative essays: A dialogic concept of writing. *RELC Journal* 46: 181–197.

Sadler, T. 2004. Informal reasoning regarding socioscientific issues: A critical review of the research. *Journal of Research in Science Teaching* 41: 513–536.

Sampson, V., and D.B. Clark. 2008. Assessment of the ways students generate arguments in science education: Current perspectives and recommendations for future directions. *Science Education* 92: 447–472.

Sandoval, W.A. 2003. Conceptual and epistemic aspects of students' scientific explanations. *Journal of the Learning Science* 12(1): 5–51.

Sandoval, W.A., and K. Millwood. 2005. The quality of students' use of evidence in written scientific explanations. *Cognition and Instruction* 23: 23–55.

Schwarz, B.B. 2009. Argumentation and learning. In *Argumentation and education: Theoretical foundations and practices*, ed. N. Muller Mirza, and A.-N. Perret-Clermont, 91–126. Berlin: Springer-Verlag.

Schwarz, B.B., Y. Neuman, J. Gil, and M. Ilya. 2003. Construction of collective and individual knowledge in argumentative activity. *Journal of the Learning Sciences* 12: 219–256.

Simon, S. 2008. Using Toulmin's argument pattern in the evaluation of argumentation in school science. *International Journal of Research & Method in Education* 31: 277–289.

Stapleton, P. 2001. Assessing critical thinking in the writing of Japanese university students: Insights about assumptions and content familiarity. *Written Communication* 18: 506–548.

Stapleton, P., and Y. Wu. 2015. Assessing the quality of arguments in students' persuasive writing: A case study analyzing the relationship between surface structure and substance. *Journal of English for Academic Purposes* 17: 12–23.

Takao, A.Y., and G.J. Kelly. 2003. Assessment of evidence in university students' scientific writing. *Science & Education* 12(4): 341–363.

Toulmin, S. 1958. *The uses of argument*. Cambridge: Cambridge University Press.

Toulmin, S. 2003. *The uses of argument*, vol. 2. Cambridge, UK: Cambridge University Press.

Uccelli, P., C. Dobbs, and J. Scott. 2013. Mastering academic language: Organization and stance in the persuasive writing of high school students. *Written Communication* 30(1): 36–62.

Uysal, H. 2008. Tracing the culture behind writing: rhetorical patterns and bidirectional transfer in L1 and L2 essays of Turkish writers in relation to educational context. *Journal of Second Language Writing* 17(3): 183–207.

Varghese, S.A., and S.A. Abraham. 1998. Undergraduates arguing a case. *Journal of Second Language Writing* 7: 287–306.

Weigle, S.C. 2002. *Assessing writing*. Cambridge, UK: Cambridge University Press.

White, E.M. 1985. *Teaching and assessing writing*. San Francisco: Jossey-Bass.

Wingate, U. 2012. Argument! Helping students understand what essay writing is about. *Journal of English for Academic Purposes* 11(1): 145–154.

Wolf, C.R. 2011. Argumentation across the. *Written Communication* 28(2): 193–219.

Wolfe, C.R., M.A. Britt, and J.A. Butler. 2009. Argumentation schema and the side bias in written argumentation. *Written Communication* 26: 183–209.

Zeidler, D.L. 1997. The central role of fallacious thanking in science education. *Science Education* 81: 483–496.

Zohar, A., and F. Nemet. 2002. Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching* 39(1): 35–62.