Metadiscourse and Counterargument Integration in Student Argumentative Papers

Philip M. McCarthy[1], Noor W. Kaddoura[1], Khawlah Ahmed[1], Rachel Hall Buck[1], Anuja M. Thomas[1], Ayah Al-harthy[1] & Nicholas D. Duran[2]

[1] College of Arts and Sciences, American University of Sharjah, Sharjah, United Arab Emirates

[2] Arizona State University, United States

Correspondence: Philip M. McCarthy, College of Arts and Sciences, American University of Sharjah, Sharjah, United Arab Emirates.

Abstract

Incorporating counterarguments can be challenging for many student-writers, including those for whom English is a second or other language. In this study, we present findings that may lead to improvements in students' understanding of the benefits of integrating counterarguments. In our study, expert readers were presented with excerpts of student academic writing in order to assess the degree to which counterarguments could be identified through student deployment of metadiscoursal features. The results suggest that student-writers deploy metadiscoursal features with insufficient frequency and consistency. The results further suggest that when metadiscoursal features are used in counterarguments, their deployment is largely restricted to the sentence-initial position. The study also considers the pedagogical and technological issues associated with the findings.

1. Introduction

Although argumentative writing may take many forms, one common goal of such writing is to persuade readers towards a given claim through the integration and presentation of contrasting points of view (Browning, Boylan, Burton, DeVries, & Kurtz, 2018; Rothery, 1996; Schleppegrell, 2004; Strongman, 2013). In this persuasive style of argumentative writing, the author's main position (often called the *thesis*) is illustrated through "supporting evidence"; however, to demonstrate a full picture of expertise to the audience, the author must also consider and present the opposing viewpoints, often referred to as "counterarguments" (McCarthy & Ahmed, in press; Nussbaum & Kardash, 2005; Van Eemeren, Grootendorst, & Henkemans, 1996; Yeh, 2002).

Within this framework, A "supporting paragraph" is one in which writers are providing evidence in support of their thesis statement or claim in order to try to convince the reader to accept their position. An example of such a claim may read "First and foremost, there are many benefits if artists are free from restrictions." Meanwhile, the function of a counter-argument is similar in that writers use it to convince the audience to accept their claim; however, writers use the counter-argument in order to present an opposing viewpoint. An example of such a claim in a counter-argument paragraph may read "Looking from another perspective, there are certain drawbacks if artists are restricted."

Despite such commonplace explanations of what constitutes argumentative writing, the actual creation of a quality paper remains a challenge for student-writers (Kuteeva, 2011; Pessoa, Mitchell, & Miller, 2017), including second language writers (Hirvela, 2017; Johns, 1993). For example, students often severely limit the inclusion of counterarguments (or opposing viewpoints) and also misuse language required to signal their counterargument intent (Knudson, 1992; Leitao, 2003; Stapleton, 2001). Consequently, argumentative papers run the risk of appearing one-sided, and this one-sidedness can be detrimental because the actual integration of counterarguments often serves to enhance the paper's overall persuasiveness and purpose (Kuhn, 1991; Leitao, 2003; Walton, 2007). Such challenges often occur because students are more likely to be lacking in writing experience, broader knowledge of the given subject matter, prototypical language use, and any number of associated writing issues.

In this study, we address the importance of counterargument integration in student argumentative papers. We begin by demonstrating empirical research highlighting the challenges of effective counterargument integration, including issues of metadiscourse and intercultural rhetoric. We then highlight current approaches for improving counterargument integration, including a discussion of template usage and specificity in assignment details. Following this discussion, we posit unresolved questions that may lead to further improvements in students' understanding of the benefits of integrating counterarguments. Based on these questions, we then present an empirical study of student-writers' counterargument structure. More specifically, the study assesses the degree to which expert judges can reliably gauge student attempts to integrate counterarguments. We recognize that this study cannot address all the issues of counterarguments; however, it does provide important findings and a pathway to much needed future research on counterargument integration.

2. Literature Review

A lack of counterarguments in student-writing appears to be widespread, regardless of the first language of the student (Kamimura, 1996; Wolfersberger, 2003). For example, Rusfandi (2015) investigated the argument/ counterargument structure in Indonesian EFL learners' essays. The results showed that the majority of the essays in both Indonesian and English were generally one-sided and lacked counterarguments. Similarly, Qin and Karabacak (2010) addressed argumentative writing in Chinese L2 English speakers. The results revealed that support claims were far more common than counterargument claims. Such empirical evidence leads us to the following discussion as to why student-writers face challenges in effectively integrating counterarguments in argumentative writing.

2.1 The Challenge for Counterargument Integration

It is useful to begin by distinguishing the *function* of a counterargument (i.e., what it does) from its *purpose* (i.e., what it seeks to achieve). Accordingly, and as described by McCarthy and Ahmed (in press) and Liu and Lim (2016), while a counterargument may *function* as a presentation of information that challenges the writer's thesis, we should not necessarily see such "challenges" as entailing the opposite or even a conflicting view of the thesis. Instead, counterarguments may be viewed as implying a representative and balanced presentation that reflects existing and presumed issues for the thesis.

In contrast to the *function*, the *purpose* of counterarguments is to demonstrate that 1) the audience's concerns have been fully explored and objectively detailed, while at the same time, 2) the information provided in the counterarguments, in and of itself, should *not* be viewed as being in opposition to the position of the writer. Such an understanding reveals the dialogic essence of the difference between the supporting arguments of the writer's thesis and the complementary but contrasting counterarguments. That is, the success of the writer's thesis depends upon the presentation of both supporting and counter-supporting information, and while this counter-supporting information must receive an objective presentation, it must also receive suitable wording to demonstrate that the information reflects the point of view of a potential audience.

The dialogic relationship described above can be framed within the appraisal theory of Martin and White (2005). Accordingly, the theory posits that a writer needs to engage with readers to establish a balance in the perspective and attitude of the various presentations of information (also see Miller, Mitchell, & Pessoa, 2014). Thus, through selective language choices (or *moves*), the writer's stance is conveyed to the audience as either a positive or negative assessment of the position at hand. For example, the use of the phrase "at least three studies" may be inferred as more positive than an alternative such as "a few studies."

The heteroglossic nature of counterargument integration may well clarify the writer's attitude and serve to make the audience feel included; however, the approach is also facilitative of the greater goal of actually persuading the reader. As Lancaster (2016) argues "… students who consider multiple positions, acknowledge opposing views in their texts (actual or potential), and make concessions … are better poised to develop sophisticated arguments" (p. 444). Thus, in the case of counterarguments, the writer's presentation must be conveyed as anticipating and recognizing the legitimacy of the audience's challenges; while at the same time, the conveyance must also demonstrate respectful distancing from the probative value of such information (see Lancaster, 2016).

2.2 Metadiscourse and Paragraph Structure

Metadiscourse is an essential ingredient of writing in general, and counterarguments in particular (see Adel, 2006; Hyland and Tse, 2004; Hyland, 2018). According to Hyland and Tse (2004), metadiscourse is "the range of devices writers use to explicitly organize their texts, engage readers, and signal their attitudes to both their material and their audience" (p. 157). Such a definition means that metadiscourse is related to (but distinct from) what Wray (2002) refers to as *formulaic language:* the latter referring more generally to any chunk of language

that is repeated often within a text type. This distinction means that formulaic language is more akin to the *frozen expressions* described in McNamara, Graesser, McCarthy, and Cai (2014) and the *typicals* described in McCarthy, Watanabe, and Lamkin (2012). This distinction also means that metadiscourse can be more closely associated with a critical aspect of appraisal theory. Specifically, through metadiscourse, the writer is required to engage with the audience through selective language in order to ensure that the appropriate presentation of information conveys the desired perspective and attitude.

The role of metadiscourse is typically made manifest through short sentence initial phrases that assist in organizing the text, engaging the audience, and signaling the writer's disposition. For example, in argumentative writing, metadiscourse can be used to signal transitions to alternative evidence: *on the other hand, however, by contrast*, as well as serving to introduce an opposing viewpoint: *some people argue*, *an alternative view* (Adel, 2006; Amiryousefi & Rasekh, 2010; Crismore, Markkanen, & Steffensen, 1993; Hyland, 2018; Vande Koppel, 1985). However, while the role of metadiscourse is critical to effectively shaping the writer's intent, an abundance of studies have demonstrated that many student-writers struggle with its effective incorporation (e.g., Anwardeen, Luyee, Gabriel, & Kaljahi, 2013; Lee & Deakin, 2016; Li & Wharton, 2012; Mohamed & Rashid, 2019; Noorian & Biria, 2010; Rusfandi, 2015; Rustipa, 2014; Tan & Eng, 2014).

The difficulty of metadiscourse integration certainly extends to argumentative writing. For example, Bychkovska and Lee (2017) compare argumentative papers written by L1 English and L1 Chinese undergraduate students, finding that Chinese students misused metadiscourse bundles such as "on the other hand." Similarly, Heng and Tan (2010) compared first year Malaysian students' persuasive writing, finding that the Malaysian students appear to underuse hedges. Meanwhile, Ho and Li (2018) found that writers of both high-rated essays and low-rated essays demonstrated inconsistency in their application of metadiscourse.

It is useful at this point to note that the importance of metadiscourse can extend to the facilitation of the structure and function of the paragraph. To be sure, a discussion of paragraph structure and/or sentence function may have its distractors; however, publications as recent as Graham (2018) and McCarthy and Ahmed (in press) would suggest that teaching the paragraph in terms of its structure and function is still very much alive in writing pedagogy. Accordingly then, we can say that a paragraph initial sentence (sometimes referred to as a *topic sentence*) and a paragraph ending sentence (sometimes referred to as a *topic closer*) are often taught as encasing the supporting evidence presented in the body of the paragraph (Weaver et al., 2016). Within this framework, and as described by McCarthy and Ahmed (in press), the function of each sentence may be made evident through metadiscourse features. Thus, the phrase *more specifically* can serve to elucidate a topic sentence, while the phrase *for example* may serve to provide concrete evidence of the topic. It is important to note here that such a structure is not meant as a straitjacket within which to imprison either the writer or the paragraph; instead, the metadiscourse simply services the intent of a sentence, should the writer decide that such a function is warranted at one particular place in the paragraph.

In the case of counterarguments, the key sentence is likely to be the first sentence as it is in this position that the perspective (or attitude) towards the topic is often most conveniently signaled to the audience (McCarthy et al., 2008; Smith, 2008). Thus, for example, a phrase such as "Some people may argue …" would appear to signal the attitude of the audience rather than that of the writer. While such an explanation describes a useful component of a "first sentence," it is important to recognize that such a claim does not entail that a "first sentence" should be conflated with a "topic sentence." That is, the facilitative inclusion of metadiscourse within a first sentence does not mean that any such sentence must subsequently be deemed a *topic sentence*.

With the above conflation in mind, we accept that the advice of many textbooks is that a topic sentence is typically (but not always) at the beginning of a paragraph since it "states the main idea of a paragraph, limiting the topic to one specific area to be discussed through the use of controlling ideas" (Miller & Pessoa, 2016, p. 850). However, the position of a topic sentence and this particular rule has been questioned because classroom assignments and "real-world" assignments do not always align. Such an observation was started decades ago, famously by Braddock (1974), who analyzed 25 expository essays and found that "only 13%...began with a topic sentence, and only 3% ended with a topic sentence" (p. 301). More recently, researchers have reiterated that many readers expect to see the topic sentence as the first sentence of a paragraph, and that such positioning may facilitate comprehension (McNamara & Kintsch, 1996); however, as McGee (2020) emphasizes, this same expectation may also run the risk of the paragraph being superficially processed. In McGee's own analysis of 10 argumentative essays, evidence was shown of varied use of topic sentences and justifiable reasons were included as to why a student may not want to use topic sentences at all. Thus, it is important that student-writers become aware of their choices so that they can suitably and consistently introduce various viewpoints at appropriate

times and in appropriate locations. By doing so, writers can acknowledge the complexity of issues while maintaining clarity of their over-arching position (see Miller et al., 2014; Ryshina-Pankova, 2014).

2.3 Intercultural Rhetoric

Intercultural rhetoric is "the study of written discourse between and among individuals with different cultural backgrounds" (Connor, 2011, p. 1). Previously referred to as Contrastive Rhetoric, Intercultural Rhetoric acknowledges that a writer's native language, culture, and educational experience are likely to have an effect on their textual production. Such a recognition extends to accessing knowledge pertinent to the goals of the paper, interpreting perceptions of cultural norms, and engaging the empathetic skills associated with audience awareness (Ackerman, 1988). It is in line with such issues that we can consider some of the challenges for L2 students seeking to integrate counterarguments.

In terms of assessing available knowledge, L2 student-writers may experience difficulties attributable to relatively low target language proficiency. As Cummins (2000) argues, non-native speakers of English may not have immediate access to terms and phrases that help construct their written work. Moreover, knowledge that is available may contrast or conflict with L1 knowledge. It is also often the case that non-native speakers may have only studied advanced writing in the context of the target language. As such, they may not have available corresponding L1 knowledge, a situation that makes more challenging the evaluation or validation of presumed knowledge.

The role of culture in issues of counterargument integration may be particularly evident for issues such as *myside bias*. That is, in an English language setting, it is all too easy to filter the world through Western values, even assuming that Western values are themselves homogenous (Tannen, 1998; Tannen, 1999). As such, it is important to remember that not all cultures are likely to generate the same feelings about highlighting opposing views, nor are they likely to feel equally at ease with criticizing well-established positions (McCarthy et al., 2007; Rusfandi, 2015).

One further issue to consider for L2 student-writers is the empathetic processes that may facilitate students' appreciation of an audience. Where such processes are not fully engaged or considered, students may be less likely to fully appreciate the concerns of those whom they seek to persuade. Indeed, research shows that many students may hold beliefs that argumentative writing is combative in nature, suggesting that students can struggle to understand practices within disciplinary writing communities (see Buck, 2020; Ivanic, 1998).

2.4 Improving Counterargument Integration

The research described above demonstrates that student-writers may lack counterargument incorporation for a variety of reasons. Moreover, such a problem may be exacerbated by student-writers' difficulties with metadiscourse awareness. The research also suggests that the causes of these issues may be attributable as much to skill-levels as to cultural issues. Evidence such as this has led to a wide range of guidance for counterargument integration. In all such cases, it emerges that writing pedagogy and explicit assignment requirements are fundamental if student-writing is to benefit from the presence of counterarguments.

At the broader level of this guidance, Nussbaum and Schraw (2007) propose three strategies for facilitating counterargument integration. The authors argue that the first and most common strategy is *refutation*. This strategy aims to either diminish (rebut) or disprove (refute) the counterargument that is challenging the writer's original claim. The second suggested strategy is *synthesizing*. The term *synthesizing* suggests a compromise that combines the merits of both viewpoints. The third strategy is *weighing*. This strategy is used to posit that while both sides have merit, one side emerges as being the stronger.

Meanwhile, Graff and Birkenstein (2014) suggest that student-writers utilize appropriate *templates*. For example, the authors recommend sentence starters such as "A number of sociologists have recently suggested that X's work has several fundamental problems" or "It has become common today to dismiss X's contribution to the field of sociology." Such sentence starters can be modified to suit the needs of a particular argument or audience perspective.

Such templated (or formulaic) language is in line with the frozen expressions of McNamara et al. (2014) and the guidance of McCarthy and Ahmed (in press). However, some may argue that such formulaic structures homogenize writing, stifle student creative development, and may not even be representative of established authors. Such objections extend to the formula for paragraphs and the use of topic sentences. After-all, paragraphs are much more than rules (Gibbons, 2019; McKnight, 2021) and teaching formulaic rules can both include and exclude writers (McKnight, 2020).

Such caution certainly has merit, but while there is no doubt that language (in general) and writing (in particular) is creative in nature, research has shown that even the most proficient language users tend to repeat particular structures (Lewis, 1997; Wray, 2000). For example, research has shown that formulaic language constitutes as much as 52.3% of written discourse (Erman & Warren, 2000). As such, it is not a surprise that respected sites such as the Harvard College Writing Center argue that formulaic writing can facilitate counterargument integration.

The frequent use of formulaic writing in both structure and phrases suggests that it is both unavoidable and, presumably, beneficial. The practice may be particularly useful for L2 students because although these learners may produce grammatically correct sentences, their work is often judged to be unnatural because of the lack of formulaicity (Lewis, 1997; Li & Schmitt, 2009). On the whole then, the evidence suggests that the use of formulaic writing may be a useful tactic for enhancing student academic writing skills, especially for L2 learners (AlHassan & Wood, 2015).

Students' reluctance to integrate counterarguments may also be attributed to a lack of explicit assignment requirements. This position is demonstrated by Nussbaum and Schraw (2007), who focused on two different treatments of counterargument instruction in a classroom setting. The first treatment provided the students with a graphic organizer for how to construct a counterargument; the second treatment involved explicit instruction for argumentative writing. The results revealed that both treatments promoted the integration of counterarguments in student-writing.

In a related study, Nussbaum and Kardash (2005) focused on the effects of goal instruction in generating counterarguments in student-writing. In their study, one group of the students received a prompt that explicitly stated the requirement for including counterarguments (goal instruction) while the other group received a bare prompt. The results revealed significantly higher rates of counterarguments and rebuttals among the students who received the prompt with goal instruction.

Instruction and requirements also appear to be relevant to L2 students. For example, Cheng (2010) developed a three-stage socio-cognitive approach to counterargument development. The first stage aims to help the students distinguish text types related to argumentative genres. This stage presents the students with a series of questions that draw their attention to contextual variables in a given text. Some of the questions included *what purpose does the text serve? what's the writer's purpose?* and *who are the readers?* The second stage focuses on the process of constructing the argument. This process aims to help students develop and strengthen their arguments. The final stage teaches students to adapt their texts with regard to their audience and their language use. This stage highlights the aspect of audience accommodation, which is said to be crucial in producing persuasive writing (Fontaine, 1988; Rafoth, 1985; Roen & Willey, 1988). By this approach, the results of Cheng's experiment revealed higher rates in the use of counterarguments.

The strategies and guidance presented here are useful at many levels. In the first place, they tend to acknowledge that counterarguments are a necessary component of argumentative papers. A more important aspect is that they demonstrate that counterarguments do not have to weaken the author's position. Indeed, through such approaches, the author's position can be significantly strengthened.

3. Purpose of the Current Study

The evidence above suggests that counterarguments (and their associated metadiscourse) are critical to many types of persuasive papers. However, the evidence also suggests that student-writers face challenges with counterarguments and, subsequently, the structure often lacks sufficient integration. These challenges appear to include the selection of metadiscoursal wording and guidance as to where and how often in the paragraph such wording needs to be placed. Considerable research will be required to establish sufficient empirical evidence to address all the issues that counterargument integration faces, and it would be hubris to suggest that one paper could solve it all. Nevertheless, the current study serves as a reasonable point of departure for addressing, at least, some of the elements of facilitating improved counterargument integration.

The current study focuses on student-writers' counterargument paragraphs as compared to student-writers' support paragraphs. That is, the focus of the study is at the level of the paragraph rather than the essay as a whole. Such a focus allows us to better understand how student-writers integrate the assumed points of view of the audience, and how readers may interpret the efficacy of such integration. As such, for the purposes of this study, we consider the paragraph as a *discrete structure*, ensuring that the assessments made are not the result of the influence of the paper's topic, title, thesis, and surrounding text. Using these discrete paragraphs for both support and counter argument paragraphs, the study assesses the degree to which judges (i.e., trained instructors) can

recognize student attempts to integrate counterarguments through identifying the linguistic components and paragraph structures that students have used. Thus, the issue at hand is to assess when, where, and how students are introducing the perspectives of their audience. Through such an assessment, we may gain better insight as to student practices so as to better facilitate teaching practices.

Specifically then, the study centers on three assessments concerning counterargument structure. Our first assessment acknowledges that judges should be able to distinguish and identify most discrete paragraph types (i.e., "support" from "counterargument"). However, the focus for this assessment is less on any statistical significance of the distinction and more on the degree of error or divergence between the judges. Accordingly, we predicted that inconsistent deployment of metadiscourse would lead to widespread divergence in judgement evaluations. Such divergence may reveal where, why, and to what degree the misapplication or non-application of metadiscourse is a factor in counterargument integration.

Our second assessment considers the importance of the initial sentence of the paragraph. Although no claim is made as to topic sentence placement per se, the frequent emphasis on paragraph initial sentences means that we can predict that the most salient metadiscourse is likely to be present in this paragraph initial position. As such, by assessing judges' accuracy after removing first sentences, we can assess the degree to which metadiscourse in the remaining paragraph structure conveys the attitudinal position of the student-writer.

Our third assessment evaluates the degree to which the judges' accuracy is dependent on the function of the paragraph under analysis. We predict that accuracy will be weaker for counterargument paragraphs because these structures rely more heavily on the accurate application of appropriate metadiscourse. That is, support paragraphs are presumably the default (or unmarked) structure, meaning that counter arguments without appropriate metadiscourse would be more likely to be judged incorrectly for function.

In sum, the current study features three hypotheses:

1) An inconsistent deployment of metadiscourse will lead to a divergence in judgement evaluations.
2) The most salient metadiscourse will be present in the paragraph initial position.
3) Accuracy of assessments will be weaker for counterargument paragraphs than for support paragraphs.

4. Material and Methods

4.1 Corpus

Although there are many public corpora of student essays, including the Michigan Corpus of Upper-Level Student Papers (MISCUSP), the British Academic Written English (BAWE), the Corpus of Ohio Learner and Teacher English (COLTE), and the Malaysian Corpus of Students' Argumentative Writing (MCSAW), for this experiment, we used the Argument Annotated Essays Corpus, a well-established, but lesser utilized corpus of 90 student essays (Afantenos, Peldszus, & Stede, 2018; Budzynska & Villata, 2017; Lippi & Torroni, 2016). The essays for the corpus were collected from *essayforum*, which is an online active community offering writing feedback for persuasive essays written by novice students. In total, the corpus includes 1,673 sentences, comprising 34,917 words. The essays in this corpus cover various controversial topics pertaining to issues of society and politics, providing students with the opportunity to include opposing viewpoints (for full details of the corpus, see Stab & Gurevych, 2014).

4.2 Method

For the purpose of our experiment, one of the authors of this study examined, identified, and divided all sections of all essays into one of the following categories: *introduction*, *body*, or *conclusion*. This action was undertaken because, while support and counterarguments may well be referred to in introductions and conclusions, fully developed support and counterargument paragraphs are more likely to occur in the body section. Following this categorization, the same author then identified and labelled each paragraph within the section according to its function: either "supporting" or "counter-arguing." As their names imply, supporting paragraphs provide evidence *for* the writer's thesis, while counter-arguing paragraphs provide evidence *against* the thesis.

In order to narrow the focus of the study further, the same author removed any paragraphs that included a refutation of the counterargument. As noted earlier, refutations are a strategy that may be used within counterarguments; however, their presence being relatively rare and their structure being distinct, they were not included in the present study. This process resulted in the identification of 117 supporting arguments and 35 counterarguments; in itself, the much higher number of supporting paragraphs is an early indication of the relative lack of counterargument production in student-writing. It is important to emphasize at this point that the categorization and labelling were achieved *with* the careful consideration of the paper's topic, title, thesis, and/or surrounding paragraphs. This information would *not* be available in the assessments described below.

4.3 Judging the Data

To assess the paragraph function and establish inter-rater reliability, we used an *expert judge* approach (Duran, McCarthy, Graesser, & McNamara, 2007; Graesser, Olney, Haynes, & Chipman, 2005). In this approach, three independent judges with an extensive background in teaching and writing were trained to distinguish the function of paragraphs. The expert judge approach is often appropriate as those most familiar with students' challenges are in the best position to assess materials. The training process involved presenting the judges with examples of "supporting" and "counter-arguing" paragraphs. Following the previously described studies and example terminology, the judges were also informed of certain metadiscourse structures that are indicative of signaling paragraph function. For example, expressions such as *some people may argue*, *on the other hand*, and *despite the fact that* are often used to signal counterarguments. Note that although we refer to these evaluators as *judges*, their role is a surrogate for *teachers*. That is, the three judges have significant and appropriate experience in teaching writing and have received specific training for the task at hand.

Following the training process, and to make the task manageable in terms of time, a total of 58 body paragraphs were randomly selected and given to the judges for assessment. Of the 58 paragraphs, 21 were counterarguments while the remaining were support paragraphs. The judges' assessments were conducted in two stages. First, the judges were given the paragraphs *without* the first sentences. Second, the judges were given the same set of paragraphs, randomly reordered, this time *with* the first sentences included. Such a procedure was necessary to better assess the value of the paragraphs' initial sentences.

5. Results

The results largely supported all three of our predictions. Specifically, the findings suggest 1) that even trained judges are inconsistent in their ability to distinguish the paragraph types of "support" from "counterargument," 2) that most of the deployed attitudinal language indicative of the paragraph function is present in the first sentence of student paragraphs, and 3) that counterargument paragraphs were harder for judges to identify, presumably because of missing appropriate metadiscoursal features.

5.1 Quantitative Analysis

For the purposes of evaluating the judges' assessments, we applied three different approaches: 1) *conservative over-all accuracy* (requiring correct and unanimous three-judge agreement); 2) *paired judgements* (requiring correct agreement by any pair of judges), and 3) *individual assessment* (where accuracy is considered for individual judges). The first assessment, *conservative over-all accuracy,* required all three judges to be both in full agreement with each other and also in agreement with the identified function of the paragraph. Thus, for example, if the function of the paragraph was known to be a counterargument (based on the original investigator's viewing of the whole text) and if, in addition, all three judges identified the paragraph as a counterargument, then the assessment was deemed a *hit*. If any number of judges were not in agreement, or if the judges were not in line with the original investigator's view of the function of the paragraph, then the rating was judged to be a *miss*. The second assessment, *paired judgements*, is similar to *conservative over-all accuracy,* except that it required only two of the three judges to agree with each other and also to agree with the identified function of the paragraph. The final assessment, *individual assessment*, was simply the number of hits/misses that an individual judge recorded in relation to the known function of the paragraph based on the original investigator's viewing the whole text.

The relationship between *hits* and *misses* determines the over-all accuracy of any of the assessment types. To assess this relationship, we used the Fisher's Exact Test to gauge the significance of the result. Accordingly, a Fisher's Exact p-value of less than .05 indicates a significant result, suggesting that the result is unlikely to be attributable to chance.

Our first assessment assumes that trained judges will be able to significantly distinguish the discrete paragraph types of "support" from "counterargument." However, it is the degree of error (rather than success) that may be the more revealing. That is, where and to what degree there is *not* agreement between judges is indicative of the degree to which counterargument integration relies on contextual (and therefore distant) signaling of the function of the argument.

Assessment 1: Whole-paragraph condition. Our first assessment predicted that judges would be able to identify and distinguish "support paragraphs" from "counterargument paragraphs" (see Table 1). Specifically, we predicted that judges would be able to achieve this distinction based on the metadiscourse features that were present in the paragraphs. The results were statistically significant; however, the error rates demonstrate that the judgements were mixed.

Table 1. Whole paragraph condition results across all types of assessments

	Conservative over-all Accuracy	Paired Judgements		Individual Judges	
		min	max	min	max
Hits	44	45	48	48	54
Misses	14	13	10	10	4
Accuracy*	75.86%	77.59%	82.76%	82.76%	93.10%

Note. * indicates $p < .001$

For the first assessment of *conservative over-all accuracy*, judges recorded 44 *hits* and 14 *misses* (75.86% accuracy). The result is significant ($p < .001$) and suggests that judges can distinguish most support paragraphs from counterargument paragraphs. The accuracy of the paired judges was similar, ranging from *min* = 45 out of 58 (77.59%) to *max* = 48 out of 58 (82.76%). Again, the results suggest statistically significant accuracy in distinguishing paragraph function ($p < .001$). Individually, judges were also accurate in most distinctions, with results ranging from *min* = 48 out of 58 (82.76%) to *max* = 54 out of 58 (93.10%). Once, again, the results suggest statistically significant accuracy in distinguishing paragraph function ($p < .001$).

The results provide evidence that trained judges do not need to view the whole text (or have access to the thesis statements) in order to distinguish most student written support paragraphs from counterargument paragraphs. However, the error rate for these results should be noted carefully. In the *conservative over-all accuracy* condition, more than 24% of assessments were inconsistent, and even individually, one trained judge was in error for over 17% of assessments. Such a view provides initial evidence that the intent of the student-writer is often insufficiently made salient in counterarguments.

Assessment 2: No-first-sentence-paragraph condition. Our second assessment considered the importance of the initial sentence of the paragraph. More specifically, we predicted that most (but not all) of the important linguistic features for paragraph function would be in the first sentence. Consequently, if the first sentence of the paragraph were removed, we predicted that judges' accuracy of identification and distinction would be lowered. Recall at this point that judges had not previously seen the first sentences (nor any of the paragraph's sentences) when this assessment was conducted. The results for this assessment were in line with our predictions (see Table 2).

Table 2. No-first-sentence paragraph condition results across all types of assessments

	Con. Accuracy	Paired Judges			Individual Judges		
		Pair 1	Pair 2	Pair 3	Ind. 1	Ind. 2	Ind. 3
Hits	23	28	29	31	**36**	**40**	**41**
Misses	**35**	30	29	27	22	18	17
Accuracy	39.66%	48.28%	50.00%	53.45%	62.07%	68.97%	70.69%

Note. Numbers in bold indicate direction and significance. *Con. Accuracy* refers to Conservative over-all Accuracy.

As with the *whole-paragraph* results of Assessment 1 (see above), we begin with the evaluation for the assessment of *conservative over-all accuracy*. Here, the judges recorded 23 *hits* and 35 *misses* (39.66% accuracy). In contrast to the findings of Hypothesis 1, the result here is significant for number of *misses* ($p = .041$). For the less conservative rating of *paired judges*, the accuracy ranged from 28 out of 58 (48.28%; $p =$ n.s.), through 29 out of 58 (50.00%; $p =$ n.s.), to 31 out of 58 (53.45%; $p =$ n.s.). For *individual judge* assessment, the results were significant and accurate for *hits*. Here, the results ranged from 36 out of 58 (62.07%; $p = .015$), through 40 out of 58 (68.97%; $p < .001$), to 41 out of 58 (70.69%; $p < .001$).

The findings suggest that when the first sentence of a paragraph is unavailable, judges struggle to distinguish support paragraphs from counterargument paragraphs. Even individually, the inaccuracy could reach almost 38%. These results suggest that while counterargument features may exist beyond the first sentence of a paragraph, such features may be highly subjective.

We can compare the *whole-paragraph* condition (Assessment 1) to the *no-first-sentence paragraph* condition (Assessment 2) by summing the *hits* and *misses* for the assessment categories of *paired judgements* assessment and *individual assessment* (see Table 3). Note that as the *conservative over-all* assessment produces only one

result, there can be no summing. Accordingly, *conservative over-all* assessment for the *whole-paragraph* condition was 44 *hits* and 14 *misses* (75.86%). For the *no-first-sentence paragraph* condition, the accuracy was 23 *hits* and 35 *misses* (39.66%). The greater accuracy for the *whole-paragraph* condition was significant (*p* < .001), indicating that the *conservative over-all* assessment accuracy increases when the first sentence of a paragraph is made available.

Assessing for *paired judges*, the condition of *whole-paragraph* assessments provided a total of 139 *hits* and 35 *misses* (79.88% accuracy). For the condition of *no-first-sentence paragraph*, there was a total of 88 *hits* and 86 *misses* (50.57% accuracy). The greater accuracy for the condition of *whole-paragraph* is again significant (*p* < .001), indicating that the assessment accuracy of *paired judges* also increases when the first sentence of a paragraph is available.

Finally, for *individual assessment*, the *whole-paragraph* condition provided a total of 152 *hits* and 22 *misses* (87.36% accuracy). For the *no-first-sentence paragraph* condition, there was a total of 117 hits and 57 misses (67.24% accuracy). The greater accuracy for the *whole-paragraph* condition is again significant (*p* < .001), indicating that the assessment accuracy of *individual judges* once more increases when the first sentence of a paragraph is available.

The results provide compelling evidence for the importance of the first sentence in paragraphs. That is, the results suggest that the metadiscourse present in the initial sentence of a paragraph provides vital linguistic features for readers. By the same token, the results also indicate that metadiscourse is present beyond the first sentence of a paragraph although these features are likely to be minimal, irregular, subjective, and/or ambiguous.

Assessment 3: Support Paragraphs vs. Counterargument Paragraphs. Our third assessment evaluated the degree to which judges differed in terms of accuracy when identifying *support paragraphs* or *counterargument paragraphs*. We predicted that accuracy would be greater for *support paragraphs* caused by students' lack of familiarity and subsequent inconsistency in applying metadiscourse that is appropriate for counterarguments. The results were in line with our predictions (see Table 3).

Table 3. Whole-paragraph and no-first-sentence paragraph condition comparison across all types of assessments

	Con. Accuracy		Paired Judges		Individual Judges	
	WP	NFSP	WP	NFSP	WP	NFSP
Hits	**44**	23	**139**	88	**152**	117
Missed	**14**	35	**35**	86	**22**	57
Accuracy	**75.86%**	39.66%	**79.88%**	50.57%	**87.36%**	67.24%

Note. Numbers in bold indicate condition is significant at *p* < .001; *Whole-Paragraph* (WP), *No-first-sentence-Paragraph* (NFSP); *Con. Accuracy* refers to Conservative over-all Accuracy.

Beginning with the *support paragraphs*, and for the evaluation of *conservative over-all assessment* of *whole-paragraphs*, judges recorded 33 *hits* and 4 *misses* (89.19% accuracy) (see Table 4). The result is significant (*p* < .001), suggesting that judges accurately identified *support paragraphs*. The accuracy of the *paired judges* was similarly accurate, ranging from *min* = 33 *hits* and 4 *misses* (89.19%) to *max* = 35 *hits* and 2 *misses* (94.59%). Again, the results suggest significant accuracy in identifying *support paragraphs* (*p* < .001). Individually, judges were also accurate in identifying *support paragraphs*, with results ranging from *min* = 35 *hits* and 2 *misses* (94.59%) to *max* = 36 *hits* and 1 *miss* (97.30%). Once again, the results suggest significant accuracy (*p* < .001).

Table 4. Support whole-paragraph results across all types of assessment

	Conservative over-all Accuracy	Paired Judges		Individual Judges	
		min	max	min	max
Hits	33	33	35	35	36
Misses	4	4	2	2	1
Accuracy %	89.19	89.19	94.59	94.59	97.3

Note. All values are significant at *p* < .001

Turning to the *counterargument paragraphs*, and for the evaluation of *conservative over-all assessment* of *whole-paragraphs,* judges recorded 11 *hits* and 10 *misses* (52.38% accuracy) (see Table 5). The result is not

significant (p = n.s.) and suggests that judges are not in full and reliable agreement to accurately and consistently identify counterargument paragraphs. The accuracy of the *paired judges* was similar, ranging from *min* = 11 *hits* and 10 *misses* (52.38%) to *max* = 14 out of 21 (66.67%). The results are not significant for any of the paired raters (p = n.s.). When the results are summed across the three pairs, the interpretation is somewhat improved but do not reach a level of significance (37 *hits* and 26 *misses*). Finally, *individual judges'* results ranged from *min* = 13 *hits* and 8 *misses* (61.90%) to *max* = 19 *hits* and 2 *misses* (90.48%). The results were only significant for one of the judges (p < .001). When these results are summed across the three judges, the results reach a level of significance (46 *hits* and 17 *misses*; p < .001).

Table 5. Counterargument whole-paragraph results across all types of assessment

	Conservative over-all Accuracy	Paired Judges		Individual Judges	
		min	max	min	max
Hits	11	11	14	13	**19**
Misses	10	10	21	8	**2**
Accuracy	52.38%	52.38%	66.67%	61.90%	**90.48%**

Note. Numbers in bold indicate significance at p < .001

The results of Assessment 3 support our predictions. That is, judges are better able to identify *support paragraphs* than *counterargument paragraphs*. For *support paragraphs*, even the most conservative assessment of accuracy suggests that support paragraphs either 1) contain identifiable metadiscourse as to the function of the paragraph or 2) that an absence of metadiscourse generates a default assessment of support. For *counterargument paragraphs*, the results suggest that *counterargument paragraphs* in this data do *not* contain reliable, consistent, and/or easily identifiable metadiscourse as to the function of the paragraph.

5.2 Error Analysis

The inclusion of an error analysis is useful as it allows us a degree of insight into some of the inconsistencies in the judges' ratings. Perhaps more importantly, an error analysis facilitates direction for future research by identifying issues that may form new hypotheses. Note that the error analysis presented here is largely exploratory and qualitative in nature. With this in mind, we consider two of the student paragraph examples (see appendix):

Paragraph 1

That is not to say, however, that advertisements have no downsides. Of course, the advertising expenses lead to a higher product price and some of them express fake information, creating information asymmetry between consumers and companies. Yet, its merits still outweigh these downsides.

Paragraph 2

Finally, even people who are not interested in online game can still be negatively affected by using computer too much. Some social surveys have shown that a few children use computer for studying purpose, most of them are attracted by Facebook, blog, etc. instead. Due to this neglect, they will have a bad result in school because when they can not live without internet, they will have no time for their studying.

Paragraph 1 is a counterargument, and all three judges correctly rated it as such (both with and without its initial sentence). However, Paragraph 2 is a support, and while all three judges rated it correctly with the initial sentence was present, only one of the three judges rated it correctly with the first sentence removed.

To better understand these ratings, we use the terms *signaled* and *tagged*. Signaled refers to explicit (usually, sentence initial) wording that indicates the function of a sentence. For example, *therefore* indicates an upcoming conclusion whereas *in other words* may indicate a clarification. By contrast, tagged wording is typically within the sentence, and though experience may indicate where, when, and why it is used, its function is not necessarily entailed by a prototypical definition. For example, the phrase *a few studies* may imply a sense of negativity whereas the semantically similar *at least three studies* may be viewed more positively.

Armed with this terminology, we can hypothesize that the two student paragraphs have considerable instances of both signaled and tagged language. For example, Paragraph 1 begins with the signals *That is not to say* and *however*, both of which imply a contrary position. The signal *of course* then offers a concession, but only a concession, whereas the phrase *some of them* is ostensibly a positive, but hardly a resounding approval. The

writer then signals with *yet*, indicating a counter, and the marked *still* of the phrase *its merits still outweigh these downsides* firmly indicate the writer's position.

For an analysis of Paragraph 2, we begin (as in the study) without the aid of the first sentence. Under this condition, two of the three judges viewed the paragraph as a counterargument. The tagged language of the paragraph may explain why this judgment was made: the word "*some*" in "*Some social surveys*," and the "*a few*" of the phrase "*a few children*" is not a ringing endorsement. Even the "*most of*" in the phrase "*most of them*" seems arbitrary and nondescript. The subsequent choices of "*instead*," "*due to* (*this*)," and the "*no*" of "*no time*" are not particularly negative, but an absence of clearly positive choices may strengthen the initial available evidence and lead to the judgement that the paragraph was a counterargument.

Once the initial sentence is restored to the second example paragraph, the author's actual position is made clear by the word choices of "*Finally, even people who are …*," which provide a prominent signal and a supporting tagged word choice. To be sure, the word "*finally*" at the beginning of a paragraph is not explicit support; however, it does strongly imply the end of a substantial list, and writers tend to make extensive lists about what they support rather than what they do not. The supporting tagged language choice of *even* is a concession, but it strongly indicates that the writer believes the stated opinion to be robust.

Examples such as these (and further examples provided in the Appendix) suggest that the paragraph's function is made manifest by a combination of signaled and tagged language, with signals being the stronger variety, especially if placed in the paragraph initial position.

6. Discussion

In this study, we assessed how judges identified and distinguished support paragraphs from counterargument paragraphs. Based on a corpus of 58 paragraphs, the judges made two assessments: initially evaluating paragraphs with the first sentences removed, then with the paragraphs as a whole.

The results indicate that judges struggled to consistently and accurately identify many counterargument paragraphs. The result seems to be explained by a greater requirement for counterargument paragraphs to demonstrate consistent and recognizable metadiscourse. That is, counterargument paragraphs would appear to be the *marked* counterpart of the supporting paragraph. As such, the result largely supports the findings of Knudson (1992), Leitao (2003), and Stapleton (2001).

The results also support such studies as McNamara and Kintsch (1996), in that they demonstrate that the first sentence in a paragraph is a significant indicator of the paragraph's function (presumably more so in the case of counterarguments). The error rates of judges' identifications further suggests that student-writers appear to be relatively weak in conveying the language necessary for the recognition of counterarguments. As such, the result can be compared with such studies as McGee (2020), in that the function of certain paragraphs may determine the degree of need for explicit reader guidance. More specifically, as counterarguments represent a change in perspective (from the writer's perspective to that of the audience), a greater presence of explicit metadiscourse would likely be beneficial.

Taken as a whole, the results support the research of studies such as Martin and White (2005) and Lancaster (2016). That is, the results demonstrate that writers need to adjust the presentation of their perspectives and attitude in order to best convey the intent of the information presented. Above all, the results also support and extend the many studies that demonstrate the challenges faced by student-writers in incorporating counterarguments through appropriate metadiscourse. On this point, we should recall that the corpus itself yielded just 35 identifiable counterargument paragraphs, a mere 23% of the total paragraphs identified. This evidence alone, supporting Kamimura (1996) and Wolfersberger, (2003), would seem to support greater explicit instruction of counterargument integration (as recommended by Graff & Birkenstein, 2014; Nussbaum & Kardash, 2005; Nussbaum & Schraw, 2007).

The current study has a number of implications for writing pedagogy. First, the findings indicate that instructors may benefit from greater explicit assignment requirements for the integration of counterarguments. With the benefit of such instruction, including explicit instruction as to appropriate metadiscourse and the effects of using such phrases on an intended audience, students may be better able to integrate counterarguments, thereby increasing the likelihood of a successful completion of a persuasive paper. Instruction may also include examples and exercises that identify and distinguish rebuttals, refutations, synthesizing, and weighing. It would also include advice and instruction as to the problem of myside bias and discussions that illuminate audience awareness. In addition, material and approaches that are sufficiently culturally sensitive should be included to avoid combative perceptions while enhancing probative and critical analysis.

Along with the benefits to teaching, the current study also has important implications for software development. Systems such as Auto-Peer (McCarthy et al., 2021) analyze student papers and provide extensive assessment, feedback, and explanations. The current research is useful because it indicates that metadiscoursal choices may successfully identify counterarguments. As such, an algorithm for counterargument identification can be deployed. Provided with such an algorithm, automated writing systems can better inform student-writers as to the presence and deployment of counterarguments in their papers.

7. Limitations and Future Research

The current study is limited in terms of the data assessed and the breadth of the ratings approach. However, despite these limitations, the findings are sufficiently useful to direct future research. Specifically, such research may include 1) a broader taxonomy of metadiscoursal language that is indicative of the distinctions between support and counterargument paragraphs, 2) greater analysis of counterargument structures, including refuting, rebutting, synthesizing, and weighing, and 3) development and analysis of teaching material. In addition, future studies need to 1) compare students' and instructors' analyses of paragraph function and 2) compare high-achieving student-writers' analyses of counterarguments with those of lower-achieving student-writers'.

8. Conclusion

The findings of the current study provide evidence in line with previous research demonstrating that students are in need of greater explicit instruction with regard to counterargument integration. While considerable future research lies ahead, the current study is a positive step towards a better understanding of counterargument integration and provides useful findings that may serve to facilitate students and teachers in producing balanced, effective, and persuasive arguments.

References

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition. *Journal of Experimental Psychology: General, 117*(3), 288-318. https://doi.org/10.1037/0096-3445.117.3.288

Adel, A. (2006). *Metadiscourse in L1 and L2 English*. John Benjamins Publishing Company. https://doi.org/10.1075/scl.24

Afantenos, S., Peldszus, A., & Stede, M. (2018). Comparing decoding mechanisms for parsing argumentative structures. *Argument & Computation, 9*(3), 177-192. https://doi.org/10.3233/AAC-180033

AlHassan, L., & Wood, D. (2015). The effectiveness of focused instruction of formulaic sequences in augmenting L2 learners' academic writing skills: A quantitative research study. *Journal of English for Academic Purposes, 17*, 51-62. https://doi.org/10.1016/j.jeap.2015.02.001

Amiryousefi, M., & Rasekh, A. E. (2010). Metadiscourse: Definitions, issues and its implications for English teachers. *English Language Teaching, 3*(4), 159-167. https://doi.org/10.5539/elt.v3n4p159

Anwardeen, N. H., Luyee, E. O., Gabriel, J. I., & Kaljahi, S. A. R. (2013). An analysis: The usage of metadiscourse in argumentative writing by Malaysian tertiary level of students. *English Language Teaching, 6*(9), 83-96. https://doi.org/10.5539/elt.v6n9p83

Braddock, R. (1974). The frequency and placement of topic sentences in expository prose. *Research in the Teaching of English, 8*(3), 287-302. Retrieved from https://www.jstor.org/stable/40170600

Browning, E., Boylan, E., Burton, K., DeVries, K., & Kurtz, J. (2018). *Let's get writing!* (A. Moser, Ed.). Virginia Western Educational Foundation. Retrieved from https://open.umn.edu/opentextbooks/textbooks/725

Buck, R. H. (2020). Undergraduate students' expectations of persuasion and argumentation: A pilot study at a Middle Eastern university. In N. Kenny et al. (Eds.), *English for specific purposes instruction and research: Current practices, challenges and innovations* (pp. 27-52). Palgrave-MacMillan. https://doi.org/10.1007/978-3-030-32914-3_3

Budzynska, K., & Villata, S. (2017). Processing natural language argumentation. In P. Baroni, D. Gabbay, M. Giacomin, & L. van der Torre, L. (Eds.), *Handbook of Formal Argumentation* (pp. 576-625). College Publications.

Bychkovska, T., & Lee, J. J. (2017). At the same time: Lexical bundles in L1 and L2 university argumentative writing. *Journal of English for Academic Purposes, 30*, 38-52. https://doi.org/10.1016/j.jeap.2017.10.008

Cheng, F. W. (2010). A socio-cognitive modeling approach to teaching English argumentation. *The Asian ESP Journal, 6*(1), 120-146.

Connor, U. (2011). *Intercultural rhetoric in the writing classroom*. University of Michigan Press. https://doi.org/10.3998/mpub.3488851

Crismore, A., Markkanen, R., & Steffensen, M. (1993). Metadiscourse in persuasive writing: A study of texts written by American and Finnish university students. *Written Communication, 10*(1), 39-71. https://doi.org/10.1177/0741088393010001002

Cummins, J. (2000). Language proficiency in academic contexts. *Language, power and pedagogy: Bilingual children in the crossfire* (pp. 57-85). Multilingual Matters. https://doi.org/10.21832/9781853596773-004

Duran, N. D., McCarthy, P. M., Graesser, A. C., & McNamara, D. S. (2007). Using temporal cohesion to predict temporal coherence in narrative and expository texts. *Behavior Research Methods, 39*(2), 212-223. https://doi.org/10.3758/BF03193150

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text, 20*(1), 29-62. https://doi.org/10.1515/text.1.2000.20.1.29

Fontaine, S. I. (1988). Using what they know: 9-, 13-, and 18-year-olds writing for different audience. In B. Rafoth & D. Rubin (Eds.), *The social construction of written communication* (pp. 99-116). Ablex Publishing Corp.

Gibbons, S. (2019). Death by PEEL? The teaching of writing in the secondary English classroom in England. *English in Education, 53*(1), 36-45. https://doi.org/10.1080/04250494.2019.1568832

Graesser, A. C., Olney, A., Haynes, B. C., & Chipman, P. (2005). AutoTutor: A cognitive system that simulates a tutor that facilitates learning through mixed-initiative dialogue. In C. Forsythe, M. L. Bernard & T. E. Goldsmith (Eds.), *Cognitive systems: Human cognitive models in systems design*. Lawrence Erlbaum Associates, Inc.

Graff, G., & Birkenstein, C. (2014). *They say / I say: The moves that matter in academic writing* (3rd ed.). W.W. Norton & Company.

Graham, A. (2018). *English for academic purposes: A handbook for students* (Ser. Critical study skills). Critical Publishing.

Heng, C. S., & Tan, H. (2010). Extracting and comparing the intricacies of metadiscourse of two written corpora. *International Journal of Education and Development using Information and Communication Technology, 6*(3), 124-146.

Hirvela, A. (2017). Argumentation and second language writing: Are we missing the boat? *Journal of Second Language Writing, 36*, 69-74. https://doi.org/10.1016/j.jslw.2017.05.002

Ho, V., & Li, C. (2018). The use of metadiscourse and persuasion: An analysis of first year university students' timed argumentative essays. *Journal of English for Academic Purposes, 33*, 53-68. https://doi.org/10.1016/j.jeap.2018.02.001

Hyland, K. (2018). *Metadiscourse: Exploring interaction in writing* (2nd ed.). Bloomsbury.

Hyland, K., & Tse, P. (2004). Metadiscourse in academic writing: A reappraisal. *Applied Linguistics, 25*(2), 156-177. https://doi.org/10.1093/applin/25.2.156

Ivanic, R. (1998). *Writing and identity: The discoursal construction of identity in academic writing*. John Benjamins Publishing Company. https://doi.org/10.1075/swll.5

Johns, A. (1993). Written argumentation for real audiences: Suggestions for teacher research and classroom practices. *TESOL Quarterly, 27*(1), 75-90. https://doi.org/10.2307/3586952

Kamimura, T. (1996). Composing in Japanese as a first language and English as a foreign language: A study of narrative writing. *RELC Journal, 27*(1), 47-69. https://doi.org/10.1177/003368829602700103

Knudson, R. E. (1992). The development of written argumentation: An analysis and comparison of argumentative writing. *Child Study Journal, 22*(3), 167-184.

Kuhn, D. (1991). *The skills of argument*. Cambridge University Press. https://doi.org/10.1017/CBO9780511571350

Kuteeva, M. (2011). Wikis and academic writing: Changing the writer-reader relationship. *English for Specific Purposes, 30*(1), 44-57. https://doi.org/10.1016/j.esp.2010.04.007

Lancaster, Z. (2016). Do academics really write this way? A corpus investigation of moves and templates in "*They Say / I Say*". *College Composition and Communication, 67*(3), 437-464.

Lee, J. J., & Deakin, L. (2016). Interactions in L1 and L2 undergraduate student writing: Interactional metadiscourse in successful and less-successful argumentative essays. *Journal of Second Language Writing, 33*, 21-34. https://doi.org/10.1016/j.jslw.2016.06.004

Leitao, S. (2003). Evaluating and selecting counterarguments. *Written Communication, 20*, 269-306. https://doi.org/10.1177/0741088303257507

Lewis, M. (1997). Pedagogical implications of the lexical approach. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 255-270). Cambridge University Press. https://doi.org/10.1017/CBO9781139524643.018

Li, J., & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing, 18*(2), 85-102. https://doi.org/10.1016/j.jslw.2009.02.001

Li, T., & Wharton, S. (2012). Metadiscourse repertoire of L1 Mandarin undergraduates writing in English: A cross-contextual, cross-disciplinary study. *Journal of English for Academic Purposes, 11*(4), 345-365. https://doi.org/10.1016/j.jeap.2012.07.004

Lippi, M., & Torroni, P. (2016). Argumentation mining: State of the art and emerging trends. *ACM Transactions on Internet Technology, 16*(2), 1-25. https://doi.org/10.1145/2850417

Liu, M. S. M., & Lim, J. M. H. (2016). Handling counter-arguments in written argumentative discourse: Implications for the teaching of English for academic purposes. *Borneo Akademika, 1*(2), 34-46. Retrieved from https://ir.uitm.edu.my/id/eprint/16467

Martin, J., & White, P. (2005). *The Language of evaluation: Appraisal in English*. Palgrave Macmillan.

McCarthy, P. M., & Ahmed, K. (in press). *Writing the research paper: Multicultural perspectives for writing in English as a second language*. Bloomsbury.

McCarthy, P. M., Al-Harthy, A., Buck, R. H., Ahmed, K., Duran, N. D., Thomas, A. M., Kaddoura, N. W., & Graesser, A. C. (2021). Introducing Auto-Peer: A computational tool designed to provide automated feedback. *Asian ESP Journal, 17*, 9-43.

McCarthy, P. M., Lehenbauer, B. M., Hall, C., Duran, N. D., Fujiwara, Y., & McNamara, D. S. (2007). A Coh-Metrix analysis of discourse variation in the texts of Japanese, American, and British scientists. *Foreign Languages for Specific Purposes, 6*, 46-77.

McCarthy, P. M., Renner, A. M. Duncan, M. G., Duran, N. D., Lightman, E. J., & McNamara. D. S. (2008). Identifying topic sentencehood. *Behaior, Research and Methods, 40*, 647-664. https://doi.org/10.3758/BRM.40.3.647

McCarthy, P. M., Watanabe, S., & Lamkin, T. A. (2012). The Gramulator: A Tool to identify differential linguistic features of correlative text types. In P. M. McCarthy & C. Boonthum-Denecke (Eds.), *Applied natural language processing: Identification, investigation, and resolution* (pp. 312-333). Hershey, PA: IGI Global. https://doi.org/10.4018/978-1-60960-741-8.ch018

McGee, I. (2020). Macro- and micro-linguistic management of the argumentative essay: Implications for teaching. *Educational Studie*s, *46*(6), 640-657. https://doi.org/10.1080/03055698.2019.1627661

McKnight, L. (2020). Teaching writing by formula: Empowerment or exclusion? *International Journal of Inclusive Education*. https://doi.org/10.1080/13603116.2020.1864790

McKnight, L. (2021). Since feeling is first: The art of teaching to write paragraphs. *English in Education, 55*(1), 37-52. https://doi.org/10.1080/04250494.2020.1768069

McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes, 22*(3), 247-288. https://doi.org/10.1080/01638539609544975

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press. https://doi.org/10.1017/CBO9780511894664

Miller, R. T., & Pessoa, S. (2016). Where's your thesis statement and what happened to your topic sentences? Identifying organizational challenges in undergraduate student argumentative writing. *TESOL Journal, 7*(4), 847-843. https://doi.org/10.1002/tesj.248

Miller, R. T., Mitchell, T. D., & Pessoa, S. (2014). Valued voices: Students' use of engagement in argumentative history writing. *Linguistics and Education, 28*, 107-210. https://doi.org/10.1016/j.linged.2014.10.002

Mohamed, A. F., & Rashid, R. A. (2019). Interactive resources of metadiscourse markers in good and weak undergraduate essays. *Trends in Social Sciences, 1*(1), 1-6.

Noorian, M., & Biria, R. (2010). Interpersonal metadiscourse in persuasive journalism: A study of texts by American and Iranian EFL columnists. *Journal of Modern Languages, 20*, 64-79. Retrieved from https://jml.um.edu.my/article/view/3374

Nussbaum, E., & Kardash, C. (2005). The effects of goal instructions and text on the generation of counterarguments during writing. *Journal of Educational Psychology, 97*(2), 157-169. https://doi.org/10.1037/0022-0663.97.2.157

Nussbaum, E., & Schraw, G. (2007). Promoting argument-counterargument integration in students' writing. *The Journal of Experimental Education, 76*(1), 59-92. https://doi.org/10.3200/JEXE.76.1.59-92

Pessoa, S., Mitchell, T. D., & Miller, R. T. (2017). Emergent arguments: A functional approach to analyzing student challenges with the argument genre. *Journal of Second Language Writing, 38*, 42-55. https://doi.org/10.1016/j.jslw.2017.10.013

Qin, J., & Karabacak, E. (2010). The analysis of Toulmin elements in Chinese EFL university argumentative writing. *System, 38*(3), 444-456. https://doi.org/10.1016/j.system.2010.06.012

Rafoth, B. A. (1985). Audience adaptation in the essays of proficient and non-proficient freshman writers. *Research in the Teaching of English, 19*, 237-253. Retrieved from https://www.jstor.org/stable/40171048

Roen, D. H., & Willey, R. J. (1988). The effects of audience awareness on drafting and revising. *Research in the Teaching of English, 22*, 75-88. Retrieved from https://www.jstor.com/stable/40171133

Rothery, J. (1996). Making changes: Developing an educational linguistics. In R. Hasan & G. Williams (Eds.), *Literacy in society* (pp. 86-123). Addison Wesley.

Rusfandi. (2015). Argument-counterargument structure in Indonesian EFL learners' English argumentative essays: A dialogic concept of writing. *RELC Journal, 46*(2), 181-197. https://doi.org/10.1177/0033688215587607

Rustipa, K. (2014). Metadiscourse in Indonesian EFL learners' persuasive texts: A case study at English department, UNISBANK. *International Journal of English Linguistics, 4*(1), 44-52. https://doi.org/10.5539/ijel.v4n1p44

Ryshina-Pankova, M. (2014). Exploring academic argumentation in course-related blogs through engagement. In G. Thompson & L. Alba-Juez (Eds.), *Evaluation in context* (pp. 281-302). John Benjamins. https://doi.org/10.1075/pbns.242.14rys

Schleppegrell, M. J. (2004). *The language of schooling: A functional linguistics perspective*. Lawrence Erlbaum. https://doi.org/10.4324/9781410610317

Smith, C. G (2008). Braddock revisited: The frequency and placement of topic sentences in academic writing. *The Reading Matrix, 8*(1), 78-95.

Stab, C., & Gurevych, I. (2014). Annotating argument components and relations in persuasive essays. In *Proceedings of the 25th International Conference on Computational Linguistics (COLING 2014),* Dublin, Ireland.

Stapleton, P. (2001). Assessing critical thinking in the writing of Japanese university students. *Written Communication, 18*(4), 506-548. https://doi.org/10.1177/0741088301018004004

Strongman, L. (2013). *Academic writing*. Cambridge Scholars Publishing.

Tan, H., & Eng, W. B. (2014). Metadiscourse use in the persuasive writing of Malaysian undergraduate students. *English Language Teaching, 7*(7), 26-39. https://doi.org/10.5539/elt.v7n7p26

Tannen, D. (1998). *The argument culture: Moving from debate to dialogue*. Random House.

Tannen, D. (1999). *The argument culture: Stopping America's war of words*. Ballantine Books.

Van Eemeren, F. H., Grootendorst, R., & Henkemans, F. S. (1996). *Fundamentals of argumentation theory: A handbook of historical backgrounds and contemporary developments*. Lawrence Erlbaum. https://doi.org/10.2307/358423

Vande Koppel, W. J. (1985). Some exploratory discourse on metadiscourse. *College Composition and Communication, 36*(1), 82-93. https://doi.org/10.2307/357609

Walton, D. N. (2007). *Dialog theory for critical argumentation*. John Benjamins. https://doi.org/10.1075/cvs.5

Weaver, R., Bost, L., Kassorla, M., McKinney-Holley, K., Crowther, K., Curtright, L., Gilbert, N., Hall, B. Ravita, T., & Swenson, K. (2016). *Successful college composition* (3rd ed.). Retrieved from https://oer.galileo.usg.edu/cgi/viewcontent.cgi?article=1007&context=english-textbooks

Wolfersberger, M. (2003). L1 to L2 writing process and strategy transfer: A look at lower proficiency writers. *TESL-EJ, 7*(2), 1-13.

Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics, 21*(4), 463-489. https://doi.org/10.1093/applin/21.4.463

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press. https://doi.org/10.1017/CBO9780511519772

Yeh, S. (2002). Tests worth teaching to: Constructing state-mandated tests that emphasize critical thinking. *Educational Researcher, 30*(9), 12-17. https://doi.org/10.3102/0013189x030009012

Appendix

Signals and Marked Language in Counterarguments and Supporting Arguments

Table A1. Examples of Signals and Marked Language in Counterarguments

Sentence	Example
1	**Looking from another perspective**, there are *certain drawbacks* if artists are not restricted.
2	**However**, **the opponents** *believe* that the dominated status of English has destroyed culture diversity and many less-used languages become extinct.
3	**Nevertheless**, **opponents** of online-degrees *would argue* that attending lectures in person provides students an opportunity to communicate with teachers and other classmates.
4	**To begin with**, **those who advocate** the impracticability of university have several reasons.
5	**On the other hand**, there are *some drawbacks* from saving minority languages.
6	**Undeniably**, making the fuel cost more money *could* limit the number of vehicles to *some extent*.
7	First, **critics of** zoology argue that keeping animals in captivity is wrong because animals should be considered non-human persons and they should have the right of having freedom.
8	*It is true* that modern games *may be*, to *some extent*, beneficial for children to foster *some* skills, such as computer skills or the capacity to keep up with the latest trend.
9	**Nevertheless**, **opponents of** modern communications *claim* that *only the more* privileged are able to enjoy the benefits.
10	**On the other hand**, *complaints* about the *drawbacks* of zoos are often heard.
11	**However**, *it can be seen* that **some people** have *argued* against advertising due to its negative impacts.
12	*What is worse*, the labor market for highly educated students is now *too competitive* to get a job after finishing their studying.
13	*Not only that*, the mobile phones *can also* be used to make calls with the companies in which they would like to work with.
14	**Admittedly**, **opponents** may blame that English is making lesser-known languages disappear ever year.

Note. Words in bold indicate signal words; italicized words indicate marked language

Table A2. Examples of Signals and Marked Language in Supporting Arguments

Sentence	Example
1	**Furthermore**, *most* economies of developing countries depends on exporting raw material and agricultural products as rice, corn.
2	**Moreover**, children who play games *too much* on computer can *seriously* lack communicating skills, they will *know little* about the outside life.
3	**Additionally**, technology enriches the way of displaying traditional cultural heritage, making it more vivid and appealing.
4	**Secondly**, *it is crucial* to keep one's identity for they need a connection back to their country as well as teach their children their value of origin.
5	**To begin with**, mobile phones and other tools of modern communication facilitate *not only* contact with friends and relatives in faraway places *but also* global business.
6	**First of all**, through cooperation, children can learn about interpersonal skills which are *significant* in the future life of *all* students.
7	*The benefits* of teaching students different units are *easily* recognized.
8	**Also**, employers are *mostly* looking for people who have international and language skills.
9	**Last, but not least**, when taking environment into consideration, people *must* conceive that *the more* newspapers are published, *the more* trees are cut down.
10	Not cooking fresh food *will also* lead to the lack of nutrition.
11	*Besides*, there are other *pressing demands* on the government expenditure, such as health care service and education, so addressing unemployment *should* have low priority.
12	**In addition** to that, the students will become *more* independent and it will build their characters in the marketplace *as well*.
13	**Also**, *the more* an advertisement of a product takes place in mass media, *the more* popular the product becomes.
14	*What is more*, the factory is not isolated, other industries and business will be *quickly* established, resulting in economic growth of local area.

Note. Words in bold indicate signal words; italicized words indicate marked language