

# A Model for Topic Closers: Improving Paragraph-Ending Sentences in English L2 Academic Writing

Philip M. McCarthy<sup>1</sup>, Anuja M. Thomas<sup>2</sup>, Nicholas D. Duran<sup>3</sup>, Adrian M. Zytoskee<sup>1</sup>

<sup>1</sup> Department of English, American University of Sharjah, Sharjah, United Arab Emirates

<sup>2</sup> Department of Psychology, Florida State University, Tallahassee, Florida, United States

<sup>3</sup> Department of Psychology, University of Texas at Austin, Austin, Texas, United States

Correspondence: Adrian M. Zytoskee, Department of English, American University of Sharjah, Sharjah, P.O. Box 26666, United Arab Emirates.

Received: November 20, 2025

Accepted: January 7, 2026

Online Published: April 17, 2026

doi:10.5430/wjel.v16n4p456

URL: <https://doi.org/10.5430/wjel.v16n4p456>

## Abstract

Topic closers refer to strategies for constructing paragraph-final sentences in academic English writing. Their functional relationship to topic sentences is comparable to that between discussion and introduction sections in research articles. This study evaluates a model of ten topic-closer types (Labels) by examining whether English L2 student writers can learn to judge their functional appropriateness more consistently. Using a pretest–intervention–posttest design, participants rated 100 topic-closer sentences for suitability as paragraph endings. Learning was assessed using two complementary measures: appropriateness ratings and rating inconsistency. Mixed-effects analyses revealed a reliable overall increase in appropriateness ratings following instruction, with no corresponding time × label interaction, indicating that improvement was general rather than category-specific. In contrast, rating inconsistency showed a significant and uniform decrease across labels, suggesting that participants converged on more stable evaluative criteria for paragraph-final function. These findings indicate that brief instruction can strengthen discourse-level judgment even before fine-grained categorical distinctions emerge. Therefore, the model shows promise as both a pedagogical framework for teaching paragraph endings and a foundation for computational applications, including automated feedback systems for L2 academic writing.

**Keywords:** English L2 writing, topic closers, paragraph endings, concluding sentences, topic sentences, paragraph structure, student writing, automated feedback tools

## 1. Introduction

### 1.1 Topic Sentences and Discourse Analysis

Discourse pedagogy has a long history of researching paragraph structure, with a predominant focus on the pivotal role of topic sentences. As Braddock (1974) and Aulls (1975) highlighted, a well-crafted topic sentence provides a clear focus and direction for the ensuing discussion. Consequently, it is paramount for students to grasp the varying functions and effectiveness of topic sentences, as this understanding can enhance the accessibility and engagement of their ideas. However, although topic sentences undoubtedly constitute a critical component in composition, they are not the only component. Their counterpart, often described as paragraph ending sentences, also plays an important role in a text's coherence and organization (Rass, 2015; Rustipa, 2016). Despite their importance, both in pedagogical and empirical work, the contribution of paragraph ending sentences is seldom recognized and often poorly specified.

### 1.2 The Overlooked Role of Paragraph-Ending Sentences

With few exceptions, paragraph ending sentences are typically assigned a one-dimensional role of providing conclusions to the information detailed in their corresponding paragraphs. In this study, we argue that paragraph ending sentences, conceptualized here as topic closers, take on multiple roles, e.g., summarizing, making a prediction, drawing a conclusion, suggesting a recommendation, among other possible functions (also see McCarthy & Ahmed, 2021; Thomas et al., 2022). Furthermore, when termed as “more effective” and/or “less effective,” these roles can be recognized and applied by student writers.

### 1.3 Topic Closers as a Multi-Functional Construct

Our aim in this study is to develop a fine-grained taxonomy for topic closers that draws important distinctions between 10 proposed types of paragraph-ending sentences. Importantly, we assess the validity of this taxonomy when topic closers are presented in isolation, divorced from the context of the paragraph with which it is associated. This assessment method follows analogous work conducted by McCarthy et al. (2008) on topic sentences. Although paragraph-ending strategies exist across languages, this study focuses specifically on academic English writing. The linguistic signals examined in the model (e.g., *therefore*, *in sum*, *in short*) reflect conventions of English academic discourse and may not map directly onto other languages. As such, the present study investigates topic closers as they function within English academic prose, particularly for English L2 writers.

#### 1.4 Research Questions

In sum, in this paper, we aim to answer the following research questions:

- (i) Can college-level student writers learn to distinguish between more and less effective topic closers?
- (ii) And if so, what role do various linguistic signals play in this process?

The remainder of this paper is structured as follows. Section 2 reviews prior work on cognitive processing, pedagogical treatments, and previous taxonomies of topic closers. Section 3 outlines our model and database development. Section 4 describes the methodology for testing the model. Section 5 reports results, and Section 6 discusses their significance. Section 7 considers pedagogical and computational implications, followed by limitations in Section 8 and a conclusion in Section 9.

## 2. Literature Review

### 2.1 Cognitive Processing Consequences of Topic Closers

Although the cognitive effects of topic closers remain under-researched, their potential can be inferred from studies on related discourse elements: topic sentences and concluding sections. Topic sentences aid comprehension and recall (Aulls, 1975; Braddock, 1974; Goldman et al., 1999; Kieras, 1978; McCarthy et al., 2008; McNamara et al., 1996; Smith, 2008), and strong conclusions similarly enhance textual memory (Brooke, 2015; Glynn et al., 2019; Harvey, 2013). These effects reflect primacy and recency biases, where information at a text's start or end is more easily remembered (Baddeley & Hitch, 1993; Bruce & Papay, 1970; Crano, 1977; Watkins & Peynircioglu, 1983). Readers recall end-positioned information particularly well, even when sentence order is scrambled (Freebody & Anderson, 1986; Russell & Sewall, 1972). Such findings highlight topic closers as a neglected but pedagogically valuable discourse element (Rustipa, 2016).

### 2.2 Pedagogical Treatments of Topic Closers

Topic closers are acknowledged in previous research but rarely modelled explicitly (Budiharso, 2017; Rass, 2015; Rustipa, 2016; Shahhoseiny, 2015; Wali & Madani, 2020; Yamin, 2019). Studies highlight persistent student difficulties, including stylistic deficits and transfer effects (Crompton, 2011; Rass, 2015; Saffari et al., 2017; Shokrpour & Fallahzadeh, 2007), such as the use of informal phrases by Arabic speakers (Rass, 2015) and frequent errors among Farsi speakers (Shahhoseiny, 2015; Shpit & McCarthy, 2022).

Online resources offer inconsistent advice: some stress summarizing endings, others emphasize signals like *therefore* or *in short* (literarydevices.net; study.com). Textbooks similarly vary: many highlight paragraph-final functions and signals (Boardman & Frydenberg, 2008; Glynn et al., 2019; McCarthy & Ahmed, 2021; Weaver et al., 2016), while others omit or oversimplify the feature (Bazerman, 1997; Lumen Learning, 2014; Inoshita et al., 2019; King, 1991). This inconsistency may limit students' mastery of closers as beyond drawing conclusions, topic closers can summarize, predict, recommend, or otherwise complete paragraphs (McCarthy & Ahmed, 2021; Thomas et al., 2022). The current study adopts this broader functional view.

### 2.3 Previous Taxonomy Attempts

Thomas et al. (2022) proposed a three-dimensional taxonomy of topic closers, distinguishing function, sentence type, and signaling, with 12 classifications. While promising, the model is complex for students. Our approach here builds on this foundation by offering a more parsimonious, student-friendly framework, supported by a larger database to capture a wide range of closing strategies for L2 writers. Developing such a model responds to cognitive and pedagogical evidence highlighting the value of effective paragraph endings (Olson & Land, 2007).

Recent work has further reinforced the importance of evaluative judgment and structured assessment in L2 writing development. Sharma (2016) shows that engagement in error analysis can strengthen learners' metalinguistic awareness and sensitivity to discourse-level writing quality, even when productive gains are gradual. In parallel, rubric-based approaches continue to be associated with improved analytic attention to textual features and clearer performance criteria (Sharma, 2019). More recent studies have reported positive effects of analytic rating frameworks on writing development in EFL and EMI contexts (Alghizzi & Alshahrani, 2024; Dahmash, 2025; Keller et al., 2024). Work on metadiscourse and interactional features likewise highlights the role of explicit signals in shaping readers' interpretation of rhetorical function and textual coherence (Esfandiari & Allaf-Akbary, 2024; Hyland & Jiang, 2023; Lee & Park, 2023). Together, these findings support treating learners' sensitivity to functional sentence appropriateness as a meaningful instructional and assessment target.

## 3. The Topic Closer Model

### 3.1 Data Source (TCD 1.0)

We created the Topic Closer Database 1.0 (TCD 1.0) to identify more and less effective topic closers. TCD 1.0 was seeded with authentic student writing but later modified to produce suitable teaching and experimental materials. As such, TCD 1.0 is not a conventional corpus and is not intended for large-scale quantitative analysis, though it is available to researchers upon request. The database draws on 111 papers from six Advanced Academic Writing classes taught by four instructors. These papers, produced by multilingual students, underwent instructor feedback and revisions and are generally of college-level quality despite minor errors. From these papers, we extracted all paragraph-final sentences, yielding 1,943 items. These sentences informed the model described below and were further modified to create sufficient exemplars for teaching and testing.

### 3.2 Data Preparation

Revisions targeted three issues: adding transitional markers (e.g., *therefore*, *thus*) to create varied conclusion signals; adding list items to diversify summary signals; and inserting in-text citations for support signals. Additional edits improved grammar and vocabulary to meet academic standards. These modifications produced the 500 sentences that comprise TCD 1.0.

### 3.3 Label Taxonomy

The topic closer model used in this study has some similarities to the model described in Thomas et al. (2022). However, unlike the complex, multiple dimension approach taken in that study, our model is more parsimonious and focuses directly on the various possible functions of topic closer sentences. Accordingly, our examination of the original data that formed TCD 1.0 led us to identify ten possible functions of paragraph-ending sentences. For each of these functions, we ascribe a *Label*. Accordingly, the 10 labels that comprise TCD 1.0 are as follows: *Simple Conclusions*, *Complex Conclusions*, *Summaries*, *Recommendations*, *Predictions*, *Speculations*, *Supplementals*, *Supports*, *Elaborations*, and *Topic Sentences*. The functions of these sentences are typically indicated by the use of linguistic signals (e.g., *therefore*, *in sum*, *in addition*). It should also be noted that some signals, particularly less ambiguous signals like *in sum* and sentence initial signals like *therefore*, are higher in saliency than non-sentence initial or highly versatile signals. The labels, along with the signals that indicate their purpose are described below (see Table 1):

#### 3.3.1 Simple Conclusions

Simple conclusions draw inferences from paragraph content. Signals include transitionals (*therefore*, *thus*, *as such*, *in conclusion*) or the structure *this + noun + conclusion verb* (e.g., *shows*, *demonstrates*).

*Examples:*

“Thus, landfills are a major contributor to the climate change problem.”

“This statistic illustrates that the efficiency of the existing public healthcare and the private insurance system is low.”

#### 3.3.2 Complex Conclusions

Complex conclusions combine a simple conclusion with an explanatory clause. They typically contain both a conclusion signal (*in conclusion*, *as such*, *accordingly*, *therefore*) and a causal connector (*because*, *since*, *as*).

*Example:* “Accordingly, the increase in high minimum wage is not beneficial to the economy because the demand for labor decreases simultaneously.”

#### 3.3.3 Summaries

Summary sentences restate paragraph content. They are often signaled by expressions such as *in other words*, *in short*, *in sum*, *overall*, or *in brief*, or by list structures summarizing key points.

*Example:* “In summary, artists who work hard, are social, and believe in their work are more likely to succeed.”

#### 3.3.4 Predictions

Predictions make claims about future outcomes based on paragraph information. Signals include the modal *will* and expressions like *ultimately* or *in the future*.

*Example:* “Serious actions taken against climate change will lead to a decrease in the predicted number of deaths by another 2.6 million.”

#### 3.3.5 Recommendations

Recommendations urge action based on paragraph content. Signals include modals such as *should*, *must*, *need to*, *have to*, and *ought to*.

*Example:* “Therefore, society must aim towards replacing non-renewable energy resources with renewable resources.”

#### 3.3.6 Speculations

Speculations project tentative future outcomes and are signaled by modals such as *would*, *may*, *might*, or *could*.

*Example:* “As such, logging companies could be one of the solutions to large scale unemployment.”

#### 3.3.7 Supplementals

Supplementals add new claims, often using additive (*in addition*), contrastive (*by contrast*), or hierarchical (*first*, *second*, *finally*) transitionals.

*Example:* “Another concern is that adolescents are often inexperienced in handling stressors and may use guns to resolve conflicts.”

#### 3.3.8 Supports

Supports provide evidence, typically through citation phrases, parenthetical references, or signals like *for example* and *for instance*.

*Example:* “Hoberman (2008) suggests that media attention attracts foreign investment and creates jobs.”

#### 3.3.9 Elaborations

Elaborations offer additional commentary or explanation and rarely include specific signals.

*Example:* “Soccer in the U.S. has recently improved in terms of gender wage gaps, but full equity remains distant.”

#### 3.3.10 Topic Sentences

Topic sentences introduce new claims without transitions. Although structurally suited to openings, they sometimes appear at paragraph ends. They are usually short, noun-phrase initial, and loosely connected to preceding material.

*Example:* “Anxiety can inflict harmful effects on the overall well-being of a person.”

Table 1. Topic-Closer Labels and Defining Features

Label	Core rhetorical function	Defining linguistic / structural features	Typical signals or cues
<b>Simple Conclusions</b>	Draw inferences from paragraph content	Express a conclusion derived from prior information; often use either a conclusion transition or a “this + noun + conclusion verb” structure	therefore, thus, as such, in conclusion; this + noun + shows/demonstrates
<b>Complex Conclusions</b>	Combine a conclusion with an explanation	Contain both a conclusion signal and a causal connector that links the conclusion to a reason	therefore, accordingly, as such + because, since, as
<b>Summaries</b>	Restate key paragraph content	Rephrase or condense main ideas; may use list structures to summarize multiple points	in other words, in short, in sum, overall, in brief; list structures
<b>Predictions</b>	Make claims about future outcomes	Assert what will happen based on paragraph information	will, in the future, ultimately
<b>Recommendations</b>	Urge actions based on paragraph content	Direct readers or institutions toward a course of action	should, must, need to, have to, ought to
<b>Speculations</b>	Project tentative future outcomes	Express uncertainty or possibility rather than certainty	would, may, might, could
<b>Supplementals</b>	Add new or extended claims	Introduce additional, contrastive, or hierarchically ordered ideas beyond the main paragraph claim	in addition, by contrast, first, second, finally, another
<b>Supports</b>	Provide evidence	Supply examples or references that support previous claims	for example, for instance; citations; parenthetical references
<b>Elaborations</b>	Offer further explanation or commentary	Expand on ideas without relying on formal signals	often no explicit signal
<b>Topic Sentences</b>	Introduce a new claim	Stand-alone claim, usually short, noun-phrase initial, loosely connected to prior material, and without transitional signals	typically none

**4. Method**

In addition to developing the model described above, we validate in this study the teachability of the model. Specifically, we examine the following:

- (i) The degree to which students give each type of topic closer a quality rating that reflects its appropriateness at the paragraph final position before and after a teaching intervention.
- (ii) How similarly students were rating sentences that fell into each distinct category before and after a teaching intervention.

*4.1 Participants*

A total of 22 participants (Male = 11, Female = 11) were included in the study. All participants were English L2 writers who had passed the university’s advanced writing courses and were within one year of graduation. Such a high level of achievement was chosen as a reasonable point of departure for assessing the learning ability of the target structure. Participants were rewarded with a gift card for fully completing the assessment.

*4.2 Materials*

Three sets of materials were used in this study: 1) the TCD 1.0 database, 2) the training material, and 3) the testing material. From TCD 1.0, we randomly selected a total of 155 sentences that would be included in training and testing material. The training material is a document that highlights the importance of topic closers and all the forms of topic closers used in the model. A total of 55 of the 155 randomly selected sentences were used in the training material. The testing material, featuring the remaining 100 sentences, was further randomly divided into two sets of 50 to form pretest and posttest modules. All materials used in this study are available on the OSF repository ([https://osf.io/mts5f/?view\\_only=64ef94c06dc7468089511d203f146598](https://osf.io/mts5f/?view_only=64ef94c06dc7468089511d203f146598)).

*4.2.1 Training Material*

The training material is a comprehensive guide to topic closers. This material contains an explanation of the importance of topic closers, definitions of the various labels, and prototypical examples of sentences belonging to each Label. Furthermore, a verification test containing simple questions pertaining to the material described above is included in the training document in order to ensure that participants read the document carefully.

With regard to the complexity of the training material, we followed Simensen (1987) in ensuring that the material contains easy-to-process examples that are free from distractors. Such examples serve the purpose of drawing attention to the linguistic features under analysis. Additionally, signals that help identify the function of the sentence are highlighted in the material. By doing so, we aimed to maximize the learnability of the features and provide participants with clear examples of the sentences from the various labels.

*4.2.2 Testing Material*

The testing material contains a set of topic closer sentences that were randomly selected from TCD 1.0. The material contains ten

sentences from each of the ten labels. Thus, a total of 100 sentences is used in the testing phase. The 100 sentences in the testing material are equally divided into two complementary sub-tests, titled *Test S* and *Test T*. This division was deemed necessary to allow for the administration of both a pretest and a posttest. In order to mitigate the differences between the two versions, half the participants in the pretest were presented with Test S and the other half were presented with Test T. Similarly, in the posttest, students who were initially presented with Test S were presented with Test T and vice versa. Both the pretest and posttest were designed to take approximately 25 minutes to complete although there was no time limit set for either test. The testing material instructs participants to rate sentences along a Likert scale, ranging from 1 to 6, where a score of 1 indicates that the given sentence is an exceptionally poor example of a topic closer and a score of 6 indicates that the sentence is an exceptionally good example.

#### 4.3 Procedure

Participants were first instructed to take a pretest to determine their knowledge of the material prior to the intervention. The test was conducted through the survey collection platform SurveyMonkey. Following the pretest, participants were then instructed to complete an intervention. The intervention involves reading a document on the topic closer model. Following the intervention, participants were instructed to take a posttest. The posttest determines how much learning has taken place as a result of the intervention material.

#### 4.4 Scoring and Hypotheses

Assessment of learning is conducted through the use of two evaluations: appropriateness ratings and rating inconsistency. Note that we operationalize the term *accuracy* here as relative appropriateness ratings rather than objective correctness. The measures (appropriateness ratings and rating inconsistency) reflect participants' judgments of sentence appropriateness rather than explicit identification of topic-closer labels. Accordingly, the terms "appropriateness ratings" and "rating inconsistency" are used in what follows to avoid implying direct category recognition. Both evaluations provide evidence for changes in participants' sensitivity to the functional appropriateness of paragraph-final sentences following the intervention. The use of structured rating criteria is consistent with prior findings that rubric-guided evaluation can enhance learners' analytic awareness and instructional transparency in EFL writing contexts (Sharma, 2019).

#### 4.5 Scoring Procedures

##### 4.5.1 Appropriateness Ratings

The primary goal of the appropriateness ratings is to assess the degree to which evaluations rise or fall from pretest to posttest. Accuracy for appropriateness is derived from the mean of the five evaluated sentences per participant, per Label, per test. Accordingly, each participant contributes ten accuracy scores for the pretest and another ten for the posttest.

To facilitate the interpretation of the analysis, four critical levels of *desirability* were determined: (i) *highly desirable*, with an accuracy score of 4.35 or above, (ii) *more desirable*, with a score between 4.34 and 3.5, (iii) *less desirable*, with a score between 3.49 and 2.66, and (iv) *undesirable*, with a score of 2.67 and below.

##### 4.5.2 Hypotheses for Appropriateness Ratings

In many experimental studies, higher posttest accuracy scores are indicative of participant learning. However, the current study requires a more specific interpretation of evaluation. Accordingly, we present below the hypotheses and reasoning for each of the four divisions of topic closers as they pertain to accuracy.

We predicted accuracy patterns for Label groups as follows:

**Simple & Complex Conclusions, Summaries:** High pretest scores, with modest posttest gains from recognizing less prototypical structures (e.g., *this + noun*) and list-based summaries; expected to remain above the 4.35 "highly desirable" threshold.

**Recommendations, Predictions, Speculations:** Relatively high evaluations throughout, likely near but not exceeding 4.35.

**Supplementals, Supports:** Posttest scores expected to rise as students learn these less typical closers are legitimate endings.

**Elaborations, Topic Sentences:** Consistently low evaluations (< 3.5), possibly lower posttest as instruction highlights their inappropriateness as closers.

##### 4.5.3 Recognition Inconsistency

The primary goal of the recognition inconsistency assessment is to gauge the degree to which fluctuation occurs in participant evaluations for each individual Label. That is, if the values provided by a participant for a Label fluctuate highly, then that participant is unlikely to be recognizing the Label of the sentences and their associated functions. By contrast, if the sentence values for a Label remain relatively stable then the participant is presumably recognizing the function of a sentence and evaluating those sentences consistently. Like accuracy for appropriateness, the inconsistency value is derived from the participant assessment of the five sentences (individually assessed for pretest and posttest). Specifically, each value for each sentence is assessed for the degree to which it varies from the accuracy value of all five sentences. For example, if a sentence is evaluated as a 4, and the accuracy evaluation for all five sentences is 3.5, then that particular sentence is scored as 0.5 for inconsistency. All individual inconsistency values are positive, so if a sentence is valued as a 3, and the accuracy value for all five sentences is 3.5, then that sentence is also evaluated as 0.5 for inconsistency. The final inconsistency evaluation for each participant is assessed as the average of the five individual scores of sentence inconsistency. The inconsistency value is particularly useful for assessing learning. That is, if learning has taken place, then inconsistency is predicted to be lower in posttest

analysis. Thus, as the intervention provides material for all 10 Labels of topic closers, we hypothesize that inconsistency values will correspondingly decrease.

## 5. Results

Results of this study are analyzed from two perspectives. The first perspective focuses on the changes in appropriateness accuracy and inconsistency values, referred to as *scoring analysis*. The second perspective focuses on the role of signals in the accuracy values, referred to as *signal analysis*.

### 5.1. Scoring Analysis

The objective of the scoring analysis is to assess whether and the degree to which learning has taken place as a result of the intervention. Evidence for the success of the intervention is provided by demonstrating that 1) accuracy values from pretest to posttest moves in the predicted directions and 2) inconsistency values in the posttest decreases for all Labels. The degree to which these two factors conform to the predicted outcomes is the degree to which initial validity can be established for the Topic Closer Model.

All analyses were conducted using linear mixed-effects models in R (version 4.5.1) with the lme4 and lmerTest packages (Bates et al., 2015; Kuznetsova et al., 2017). Mixed-effects models were selected for their capacity to account for the hierarchical structure of the data, in which individual ratings are nested within participants and items. Post-hoc comparisons were conducted using the emmeans package (Lenth, 2024). Three complementary analyses were conducted: (i) a theoretically-driven analysis grouping labels by pedagogical effectiveness (ii) an analysis of recognition accuracy examining changes in mean ratings across labels, and (iii) an analysis of recognition inconsistency examining changes in rating variability. A fully documented and reproducible version of the analysis workflow is available in an online methodological appendix (<https://nickduran.github.io/topics-closer/data-analysis-documented.html>).

### 5.2 Direct Test of Pedagogical Effectiveness Hypothesis

To provide a direct test of the core theoretical hypothesis, that learning differentially affects ratings based on the pedagogical appropriateness of different closer types, we grouped labels by their effectiveness as paragraph-ending structures. Based on our pedagogical framework, Labels were classified into three categories: Effective closers (Simple Conclusions, Complex Conclusions, Summaries, Predictions, Recommendations, Speculations;  $n = 6$  labels), Underrecognized closers (Supplementals, Supports;  $n = 2$  labels), and Ineffective closers (Elaborations, Topic Sentences;  $n = 2$  labels).

A linear mixed-effects model was fitted with the Time (Pretest, Posttest)  $\times$  Effectiveness interaction as fixed effects and random intercepts for participants and unique items. Type III tests revealed significant main effects of Time,  $F(1, 94) = 4.95, p = .028$ , and Effectiveness,  $F(2, 94) = 44.30, p < .001$ . However, the Time  $\times$  Effectiveness interaction was not significant,  $F(2, 94) = 0.70, p = .499$ . This pattern of results indicates that while participants reliably differentiated among effectiveness categories and showed overall improvement from pretest to posttest, the magnitude of improvement was comparable across all three effectiveness groups. In other words, the intervention produced a uniform increase in ratings rather than the hypothesized differential pattern in which effective closers would increase while ineffective closers would decrease or remain stable.

Although the non-significant interaction precludes strong claims about differential learning effects across effectiveness categories, we examined the simple effects of Time within each category to characterize the descriptive patterns. These comparisons should be interpreted cautiously given the absence of a significant omnibus interaction. For Effective closers, ratings increased modestly from pretest ( $M = 3.99$ ) to posttest ( $M = 4.21$ ), though this difference did not reach statistical significance,  $t(94) = 1.51, p = .134, d = 0.16$ . For Underrecognized closers, ratings showed the largest increase, from  $M = 2.75$  at pretest to  $M = 3.27$  at posttest,  $t(94) = 2.04, p = .044, d = 0.37$ . For Ineffective closers, ratings changed minimally from pretest ( $M = 2.95$ ) to posttest ( $M = 3.07$ ),  $t(94) = 0.48, p = .630, d = 0.09$ . The pattern of effect sizes suggests that the intervention may have had its strongest impact on Underrecognized closers. These are precisely those structures (Supplementals and Supports) that the training materials highlighted as legitimate but often overlooked paragraph-ending strategies. However, given the non-significant interaction, this interpretation remains tentative and would benefit from replication with a larger sample.

### 5.3 Recognition Accuracy

As an additional analysis, we examined changes in recognition accuracy from pretest to posttest across the ten individual Labels (see Table 2). A linear mixed-effects model was fitted with Rating as the dependent variable and the interaction between Time (Pre, Post) and Label (10 levels) as fixed effects, with random intercepts for participants and unique items. Type III tests revealed significant main effects of Time,  $F(1, 80) = 6.11, p = .016$ , and Label,  $F(9, 80) = 14.19, p < .001$ . However, the Time  $\times$  Label interaction was not significant,  $F(9, 80) = 0.53, p = .852$ , indicating that the instructional intervention produced a general improvement in recognition accuracy rather than selectively benefiting particular Label types.

The results were largely consistent with our hypotheses. Accuracy values for Simple Conclusions, Complex Conclusions, and Summaries were higher in the posttest than in the pretest. Similar trends were observed for Predictions, Speculations, Supplementals, and Supports. However, contrary to our hypotheses, accuracy values also rose modestly for Elaborations and Topic Sentences. The only label to move in a lower direction was Recommendations, though the difference was negligible (from 3.98 to 3.92).

There was also support for the validity of the model from an analysis using the Desirability scale. Both Summaries ( $M = 4.55$ ) and Simple

Conclusions ( $M = 4.35$ ) moved from More Desirable in pretest to Highly Desirable in the posttest. Speculations rose from Less Desirable ( $M = 3.40$ ) to More Desirable ( $M = 3.83$ ). Supplementals evaluations also rose to 3.48, making it just under the threshold to move from Less Desirable to More Desirable. Most notably, Supports rose from Undesirable ( $M = 2.25$ ) to Less Desirable ( $M = 3.06$ ), the largest gain observed for any label.

Of the ten labels, only Supports showed a statistically significant increase,  $t(80) = 2.45, p = .016$ . However, given the non-significant omnibus interaction, this result should be interpreted cautiously as a potential signal rather than a confirmed effect. The pattern suggests that the intervention may have been particularly effective at drawing attention to citation-based evidence as a legitimate closing strategy. This is a function that students may have previously undervalued.

Examination of the random effects indicated that items accounted for 7.6% of the total variance, participants accounted for 11.3%, and the residual variance was 81.1%. The relatively modest item-level variance suggests that sentences within each label elicited reasonably consistent ratings, while the participant variance reflects individual differences in overall rating tendencies.

5.4 Recognition Inconsistency

Beyond accuracy, we assessed recognition inconsistency to determine whether the intervention led to more stable, coherent evaluations within each Label. Inconsistency was operationalized as the mean absolute deviation of a participant's five ratings within each label-time cell, reflecting the degree to which ratings fluctuated around the participant's mean for that Label. Lower inconsistency values indicate more uniform evaluations within a category, suggesting recognition of shared functional properties.

To determine whether changes in inconsistency varied by Label, we compared an additive model (Time + Label) to an interaction model (Time  $\times$  Label), both with random intercepts for participants. The likelihood ratio test indicated that the interaction term did not significantly improve model fit,  $\chi^2(9) = 6.86, p = .652$ . The additive model was therefore retained, indicating that the effect of time on inconsistency was uniform across labels.

Type III tests from the additive model revealed significant main effects of both Time,  $F(1, 408) = 14.99, p < .001$ , and Label,  $F(9, 408) = 7.46, p < .001$ . Estimated marginal means indicated that inconsistency decreased from pretest ( $M = 1.05$ ) to posttest ( $M = 0.90$ ), with a pairwise contrast of 0.15,  $SE = 0.04, t(408) = 3.87, p < .001$ . The standardized effect size ( $d = 0.37, 95\% CI [0.17, 0.57]$ ) indicates a small-to-medium reduction in rating variability following the intervention.

This uniform decrease in inconsistency supports the hypothesis that the instructional intervention promoted more coherent category recognition. When participants can reliably identify the function of a sentence, their evaluations of sentences sharing that function become more stable. The absence of a significant interaction suggests that this benefit extended across all topic closer types, including those that were initially less familiar to participants.

Table 2. Changes in Accuracy Scoring for Appropriateness Ratings After Intervention

Label	Pretest Accuracy	Posttest Accuracy	Change in Accuracy
Summary	4.25	4.55	Rose
Simple Conclusion	4	4.35	Rose
Complex Conclusion	4.23	4.31	Rose
Prediction	4.08	4.31	Rose
Recommendation	3.98	3.92	Fell
Speculation	3.4	3.83	Rose
Supplemental	3.26	3.48	Rose
Topic Sentence	3.13	3.23	Rose
Supports	2.25	3.06	Rose*
Elaboration	2.76	2.91	Rose

Note. \* indicates statistical significance at  $\alpha = .05$

5.5 Signal Analysis

The role and importance of signals in topic closers is another avenue of interest. In general, the significance of signals is highlighted by the importance of metadiscourse (Hyland & Tse, 2005). Consequently, this study features a qualitative and exploratory signal analysis. That is, the database for topic closers was primarily geared towards Division, Label, and their respective functions. As such, while care was taken to ensure signals were varied and distributed, they were not categorized and enumerated. For this reason, extensive statistical analysis of the signals is not provided. That having been said, the signal analysis is useful because it provides qualified evidence in support of the Topic Closer Model. Moreover, the analysis provides direction as to future research that may offer insight for improving the intervention and further development of the model. The analysis below is divided into three sections and based on the Desirability scale. For ease of analysis, Highly Desirable and More Desirable are discussed separately whereas Less Desirable and Undesirable are compared and contrasted.

5.5.1 Highly Desirable

In total, 26 sentences had average evaluations above 4.35 and were classified as Highly Desirable. The leading indicator for these sentences was high saliency signaling. For example, *In short* and *To summarize* were sentence initial transitionals for the four highest

ranked sentences. A further 12 examples featured sentence initial transitionals such as *Therefore*, *Thus*, *As a consequence*, *In conclusion*, and *Accordingly*. There were also three sentences that used high saliency signals such as *Based on this information* or *Based on these studies*. Only two Predictions (both using *will*) did not use similarly high saliency signals. Instead, one of these sentences featured a listed ending (common with Summaries) and the other used the transitional phrase *It would seem*. The two Supplemental sentences featured the transitionals of *Indeed* and *Above all*. Overall, the analysis suggests that prototypical sentence initial signals are highly indicative of highly desirable topic closers.

#### 5.5.2 More Desirable

In total, 32 sentences had average evaluations between 3.5 and 4.35 and were accordingly classified as More Desirable. For Simple Conclusions and Complex Conclusions the leading indicator for these sentences was non-prototypical sentence initial signaling: In other words, *Accordingly*, *Taken as a whole*, *As such*, and *Evidently*. Other sentences featured *this + noun* together with a conclusionary verb. For Predictions, and Speculations, all 11 sentences featured positive forms of modal verbs, with *will* being the most common. Only one of the sentences contained a transitional, namely, *As such*. For the two Supplementals, both appear to have been boosted by the inclusion of the non-prototypical transitionals *Indeed* and *Likewise*. For the three Topic Closers, the highest scoring sentence (4.27) began with a bridging clause, which may have increased its evaluation. The remaining two sentences received the lowest of the evaluations (from 3.5 to 3.54), barely reaching the category of More Desirable. Similarly, the two Elaboration sentences received low enough evaluations to be considered at the cusp of Less Desirable. In sum, the analysis suggests that non-prototypical sentence initial signals are better recognized than less salient internal sentence features.

#### 5.5.3 Less Desirable and Undesirable

In total, 22 sentences had evaluations between 2.67 and 3.5 and were classified as Less Desirable, while 20 had average evaluations below 2.67 and were classified as Undesirable. Of these sentences, all three Conclusions were signaled through *this + noun*, with no explicit transitionals. The one Prediction sentence (usually a high scoring topic closer) was an example of *will + not*. Of the remaining six Predictions and Speculations and one Recommendation, two were negatives and both were evaluated as Undesirable. Such analysis raises the possibility that negated modals are lower in saliency for learners. However, given that negation was not explicitly addressed in the instructional materials, the observed pattern may instead reflect reliance on surface-level form matching, with learners responding to familiar structures rather than generalizing the underlying modal meaning to new forms such as negation. Further targeted instruction would be required to distinguish between these explanations. Turning to the Supports sentences, the highest three examples each feature sentence ending parenthetical references, a feature of just two of the six that appear in the Undesirables category. The distinction may again be one of saliency, inasmuch as sentence initial referencing, such as "*Elbahnasawy and Revier (2012) posit that ...*" is higher in saliency with regard to mid-paragraph sentences. Meanwhile, the distinction between the Topic Closers and Elaborations appearing as Less Desirable and those appearing as Undesirable appears to be attributable to signaling through *the + noun*. That is, eight of the nine Less Desirable sentences did not begin with *the + noun* whereas five of the six Undesirable sentences did (the five lowest ranking sentences of all).

Taken as a whole, the analysis supports our hypothesis that signals determine, reinforce, or clarify the functions of topic closers, with greater levels of saliency facilitating the effect further. Lower saliency signals (e.g., *this + noun*) appear to be relatively effective for most student-writers as they are now able to better recognize the functions of such sentences. However, greater training is likely to improve their levels of recognition (and presumably subsequent inclusion).

## 6. Discussion

In this paper, we expanded the nomenclature proposed by Thomas et al. (2022) to develop a topic closer model with ten Labels and evaluated its instructional potential by examining whether students could learn to more consistently judge paragraph-final sentence suitability. The most robust outcome concerns rating inconsistency, which decreased reliably following the intervention and closely matched theoretical expectations. This reduction in variability indicates that participants converged on more stable judgments of functional appropriateness, suggesting genuine learning at the level of discourse evaluation rather than random response fluctuation. Therefore, improvement is most clearly observed in how consistently students applied the construct, not merely in whether individual ratings shifted upward or downward. This pattern supports the interpretation that the intervention strengthened shared evaluative criteria for paragraph-final function.

Changes in recognition accuracy, by contrast, followed a different pattern from that originally hypothesized. The study predicted differential movement across categories, with effective topic closers increasing in perceived appropriateness and ineffective ones remaining stable or declining. However, the observed data show a more uniform increase across Labels. That is, the intervention did not sharpen discrimination between strong and weak closers but instead elevated perceived suitability across categories. This outcome suggests that instruction primarily increased general sensitivity to paragraph-final function rather than refining categorical boundaries among topic-closer types. Such a pattern is consistent with early-stage conceptual learning, in which learners acquire a broader sense of what constitutes acceptable discourse behavior before developing more fine-grained distinctions.

An additional interpretive consideration concerns participants' sensitivity to the distinctions encoded in the proposed model prior to instruction. The mixed-effects analysis revealed that several Labels designated as more effective paragraph-final strategies, particularly Complex Conclusions, were already rated significantly higher than less effective types at pretest. By contrast, Elaborations and Topic

Sentences, identified in the model as weak paragraph-ending options, were among the lowest-rated categories before the intervention. This pattern suggests that participants' intuitions broadly aligned with the model from the outset, potentially constraining the magnitude of observable accuracy gains following instruction. From this perspective, the primary instructional effect may lie less in reshaping judgments and more in stabilizing and consolidating evaluative criteria, an interpretation consistent with the observed reductions in inconsistency. One notable exception concerns the Supports sentences, which participants initially rated especially low despite their classification as acceptable paragraph-ending strategies in the model. The persistence of this judgment may reflect the distinctive surface features of citation-based closers or prior instructional norms, a possibility that warrants further investigation.

Within this general trend, several underrecognized categories displayed especially notable descriptive gains. In particular, the Supports Label showed one of the largest descriptive improvements and the largest effect size among the categories, despite not reaching statistical significance after correction. This pattern suggests that learners may initially undervalue certain functional strategies that do not align with prototypical summarizing or concluding forms. Accordingly, instructional attention to such categories may be especially productive in future interventions and merits targeted investigation.

Signal-based analyses further clarify the nature of learners' judgments. Sentences containing prototypical summarizing or concluding signals were evaluated more accurately than those with sentence-internal or non-prototypical cues, indicating that surface markers continue to play a substantial role in functional interpretation. However, exclusive reliance on explicit signals risks producing mechanical and stylistically constrained writing. Therefore, effective instruction should emphasize that paragraph-final function can be realized both with and without overt markers, provided that rhetorical intent remains unambiguous. Developing such flexibility likely requires sustained exposure and guided practice beyond brief instructional treatments.

Finally, it is important to distinguish statistical significance from educational significance when interpreting these results. Although several categories exhibited reliable pretest to posttest differences, absolute gains in accuracy were modest. Consequently, the present findings should be understood as evidence of emerging discourse awareness and improved evaluative stability rather than immediate improvement in writing proficiency. The mixed-effects analysis reinforces this interpretation by demonstrating sentence-level gains distributed across categories rather than effects concentrated in a limited subset of labels. Accordingly, the intervention appears to support foundational discourse sensitivity that may facilitate later, more substantial development under extended instruction.

## 7. Implications

### 7.1 Pedagogical Implications

Our study provides a wide array of useful findings that may have considerable implications for materials development and classroom teachers. As noted by Rass (2015), Rustipa (2016), and Shahhoseiny (2015), L2 students do not seem to have a strong knowledge base of how to end paragraphs. Therefore, the model developed in this paper could serve as an effective guide for students of academic writing. Teachers can also help students practice identifying and integrating effective topic closers using a variety of classroom exercises. For example, teachers can present paragraphs that may benefit from a final sentence and ask students to complete the paragraph. Alternatively, students could be asked to review their peers' papers and critique their choices of topic closers.

Because the model was developed within the conventions of academic English writing, its pedagogical applications are most directly relevant to instructors working with English L2 writers. Topic closers and their associated signals can vary considerably across languages, so adaptations may be required in instructional contexts involving non-English writing systems.

### 7.2 Pedagogical Applications

The proposed topic-closer taxonomy can be integrated into classroom practice in several low-overhead ways. In diagnostic contexts, instructors can present short paragraphs with alternative closing sentences and ask students to rate their functional appropriateness, allowing teachers to identify gaps in discourse-level awareness without requiring extended writing tasks. The framework can also be used in guided paragraph-completion activities, where learners select or compose endings that match specific rhetorical functions. In peer-review settings, the categories provide a shared metalanguage for discussing paragraph endings, helping students move beyond surface-level comments to function-oriented feedback. Such uses align with prior work suggesting that structured evaluation criteria can support analytic attention and instructional transparency in EFL writing classrooms (Sharma, 2019). Importantly, these activities emphasize judgment and interpretation rather than memorization of Labels, maintaining consistency with the construct measured in the present study.

### 7.3 Implications for Automated Feedback Tools

The identification of lexical and structural signals for topic closers has implications for Automated Feedback Tools (AFT). AFT, such as Auto-Peer (McCarthy et al., 2021; Thomas & McCarthy, 2023) are used to computationally assess written texts using a set number of parameters. For example, Auto-Peer examines written texts for writing issues such as cohesion (see McNamara et al. 2014), topic sentences (see McCarthy & Ahmed, 2021), sentence structure (see Shpit & McCarthy, 2022, 2023), and counterarguments (see McCarthy et al., 2021, 2022), among many other aspects. The developers of Auto-Peer have also included an algorithm that seeks to identify prototypical signals of appropriate topic closers. That is, these signals can be recognized by Auto-Peer, and by so doing, Auto-Peer may be able to identify the various kinds of corresponding topic closers, subsequently providing appropriate feedback as to their effectiveness.

## 8. Limitations and Future Research

Although this study has made a considerable step towards developing a greater understanding of topic closers, it is important to acknowledge that there are certain limitations. One of the limitations is that the intervention was presented in a written format. As a result, participants did not have the opportunity to engage in discussion about the material they were learning. Therefore, future research should conduct a similar experiment in which the intervention is carried out in the classroom where teaching and exchange of feedback are possible. A second limitation is that the study's participants were students who have already completed a course in Advanced Academic Writing. As such, these participants can be considered to have more experience in academic writing. Future research should focus on participants who are current students of Advanced Academic Writing. Such participants are less likely to already have been exposed to advanced discussions on paragraph structure. A third limitation is the distribution of signals in the intervention material. The results suggest that signals are primary in determining sentence function, so with careful consideration of these signals, the degree of learning may be better assessed. As such, future research should consider a study in which signals are changed from prototypical signals (such as *Therefore* or *Thus*) to less common signals (like *Accordingly*, *Evidently*, and *this + noun*). Such a study would corroborate the hypothesis that participants are guided primarily by the signals of sentences rather than the function.

A further methodological limitation concerns the nature of the materials themselves. The topic-closer sentences evaluated by participants were researcher-constructed and adapted from an initial pool of authentic learner writing, rather than drawn from intact, naturally occurring paragraphs. This design choice was made to allow systematic manipulation and control of sentence-level features, particularly linguistic signals and functional categories, and to ensure balanced representation of all ten Labels in both the training and testing materials. Such control would not have been feasible using unmodified learner texts, which typically contain uneven distributions of structures and substantial contextual variability. As a result, the findings should be interpreted as providing evidence for the conceptual validity and learnability of the proposed taxonomy rather than as direct evidence of how frequently or effectively learners deploy topic closers in authentic writing contexts. That is, the study demonstrates that advanced L2 writers can learn to recognize and differentiate functional types of paragraph-ending sentences under controlled conditions, but it does not claim to measure real-world writing performance or natural production behavior.

A related limitation concerns the relationship between recognition and production. The present study operationalized learning in terms of participants' ability to evaluate and differentiate topic closers, rather than their ability to produce them in their own writing. Although such evaluative sensitivity reflects an important component of discourse awareness, improved recognition does not necessarily entail immediate or automatic gains in productive writing ability. As a result, the observed improvements should not be interpreted as direct evidence that participants would consistently apply the model when composing original paragraphs. Therefore, future research should consider incorporating production-based measures, such as paragraph-writing or revision tasks, to examine whether instructional gains in recognition transfer to sustained changes in learners' written performance.

## 9. Conclusion

Despite the acknowledged limitations, the current study provides important insight. This study addressed two research questions: whether students can learn to distinguish between more and less effective topic closers, and what role linguistic signals play in this process. The results indicate that students were indeed able to recognize distinctions between topic closer types, particularly those signaled by salient linguistic markers. In addition, signal saliency emerged as a key factor influencing appropriateness ratings and consistency. Specifically, the developed model demonstrates that topic closers are a complex but important component of paragraphs, and that with a suitable intervention, student-writers are able to recognize this discourse feature. Although further research and intervention modifications will doubtlessly improve student comprehension and application of topic closers, this study provides instructors, materials developers, researchers, and students with a promising pathway to more effective writing.

## Acknowledgments

Not applicable.

## Authors' contributions

Philip M. McCarthy and Anuja M. Thomas contributed to study design and data collection. Nicholas D. Duran, Anuja M. Thomas, and Philip M. McCarthy conducted the statistical analysis. Anuja M. Thomas, Adrian M. Zytoskee, and Philip M. McCarthy contributed to manuscript organization and argument development, literature integration, articulation of pedagogical implications, and coordination of manuscript revision.

## Funding

This work was supported in part by FRG-22-AS1222 and FRG23-C-S66 from the American University of Sharjah.

## Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Informed consent

Obtained.

## Ethics approval

The Publication Ethics Committee of the Sciedu Press.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

#### **Provenance and peer review**

Not commissioned; externally double-blind peer reviewed.

#### **Data availability statement**

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

#### **Data sharing statement**

No additional data are available.

#### **Open access**

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).

#### **Copyrights**

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

#### **Declaration of Generative AI Use**

During the preparation of this manuscript, the authors used ChatGPT for limited editorial assistance, specifically for feedback on formatting and language clarity. No AI tools were used for drafting substantive content, conceptualizing the research, data analysis, interpretation of results, or generating or selecting references. All AI-assisted refinements underwent thorough review and revision by the authors, who take full responsibility for the accuracy, originality, and scholarly integrity of the manuscript. All cited works were personally curated and verified by the authors without the use of AI.

#### **References**

- Alghizzi, T. M., & Alshahrani, T. M. (2024). Effects of grading rubrics on EFL learners' writing in an EMI setting. *Heliyon*, 10(18). <https://doi.org/10.1016/j.heliyon.2024.e36394>
- Aulls, M. W. (1975). Expository paragraph properties that influence literal recall. *Journal of Reading Behavior*, 7(4), 391-399. <https://doi.org/10.1080/10862967509547156>
- Baddeley, A. & Hitch, G. (1993). The recency effect: Implicit learning with explicit retrieval? *Memory & Cognition*, 21(2), 146-155. <https://doi.org/10.3758/BF03202726>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48. <https://doi.org/10.18637/jss.v067.i01>
- Bazerman, C. (1997). *Involved: Writing for college, writing for your self*. Houghton Mifflin.
- Boardman, C. A., & Frydenberg, J. (2008). *Writing to communicate: Paragraphs and essays* (3rd ed.). Pearson Longman Publishing.
- Braddock, R. (1974). The frequency and placement of topic sentences in expository prose. *Research in the Teaching of English*, 8(3), 287-302. <https://doi.org/10.58680/rte197420099>
- Brooke, M. (2015). Deconstructing academic persuasive essays by constructing argument maps and analysing means of persuasive appeal. *Pertanika Journal of Social Sciences & Humanities*, 23(S), 113-126.
- Bruce, D. & Papay, J. (1970). Primacy effect in single-trial free recall. *Journal of Verbal Learning and Verbal Behavior*, 9(5), 473-486. [https://doi.org/10.1016/S0022-5371\(70\)80090-1](https://doi.org/10.1016/S0022-5371(70)80090-1)
- Budiharso, T. (2017). Developing prompts of argumentative essay writing for EFL college students. *Indonesian Journal of Linguistics-Teaching Language*, 2(3), 227-244. <https://doi.org/10.30957/ijoltl.v2n3.531>
- Crano, W. (1977). Primacy versus recency in retention of information and opinion change. *The Journal of Social Psychology*, 101(1), 87-96. <https://doi.org/10.1080/00224545.1977.9923987>
- Crompton, P. (2011). Article errors in the English writing of advanced L1 Arabic learners: The role of transfer. *Asian EFL Journal*, 50(1), 4-35.
- Dahmash, N. F. B. (2025). The analytic use of rubrics in writing classes by language students in an EFL context: Students' writing model and benefits. *Frontiers in Education*, 10, Article 1588046. <https://doi.org/10.3389/educ.2025.1588046>
- Esfandiari, R., & Allaf-Akbary, O. (2024). Assessing interactional metadiscourse in EFL writing through intelligent data-driven learning: The Microsoft Copilot in the spotlight. *Language Testing in Asia*. <https://doi.org/10.1186/s40468-024-00326-9>
- Freebody, P., & Anderson, R. C. (1986). Serial position and rated importance in the recall of text. *Discourse Processes*, 9(1), 31-36. <https://doi.org/10.1080/01638538609544630>

- Glynn, A., Hallsten-Erickson, K. & Swing, A. (2019). *You, Writing! A Guide to College Composition*. Retrieved from <https://open.umn.edu/opentextbooks/textbooks/588>
- Goldman, S. R., Graesser, A. C., & van den Broek, P. (1999). Essays in honor of Tom Trabasso. In S. R. Goldman, A. C. Graesser, & P. van den Broek (Eds.), *Narrative comprehension, causality, and coherence* (pp. 1-10). Mahwah, NJ: Erlbaum. <https://doi.org/10.4324/9781410603135>
- Harvey, M. (2013). *The nuts and bolts of college composition*. Hackett Publishing Company, Inc.
- Hyland, K., & Jiang, F. (2023). Interaction in written texts: A bibliometric study of published research. *Studies in Second Language Learning and Teaching*, 13(4), 903-924. <https://doi.org/10.14746/ssl.40220>
- Hyland, K., & Tse, P. (2005). Metadiscourse in academic writing: A reappraisal. *Applied Linguistics*, 25(2), 156-177. <https://doi.org/10.1093/applin/25.2.156>
- Inoshita, A., Garland, K., Sims, K., Tsutsui, J. K., & Williams, T. (2019). *English composition: Connect, collaborate, communicate*. University of Hawaii Manoa. Retrieved from <https://open.umn.edu/opentextbooks/textbooks/774>
- Keller, S. D., Lohmann, J., Trüb, R., Fleckenstein, J., Meyer, J., Jansen, T., & Müller, J. (2024). Language quality, content, structure: What analytic ratings tell us about EFL writing skills at upper secondary school level in Germany and Switzerland. *Journal of Second Language Writing*, 65, 101129. <https://doi.org/10.1016/j.jslw.2024.101129>
- Kieras, D. E. (1978). Good and bad structure in simple paragraphs: Effects on apparent theme, reading time, and recall. *Journal of Verbal Learning and Verbal Behavior*, 17(1), 13-28. [https://doi.org/10.1016/S0022-5371\(78\)90496-6](https://doi.org/10.1016/S0022-5371(78)90496-6)
- King, A. (1991). Composition guide. In *Presbyterian College Handbook of Grammar and Usage*, (68-83). Retrieved from <https://www.presby.edu/doc/academicresources/Composition-Guide.pdf>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1-26. <https://doi.org/10.18637/jss.v082.i13>
- Lee, S., & Park, H. I. (2023). Changing patterns of interactive metadiscourse in *English Teaching* articles. *English Teaching*, 78(2), 83-102. <https://doi.org/10.15858/engtea.78.2.202306.83>
- Lenth, R. V. (2024). *emmeans: Estimated marginal means, aka least-squares means* [R package]. Retrieved from <https://CRAN.R-project.org/package=emmeans>
- Lumen Learning. (2014). *English composition I* [Open educational resource]. Retrieved from [https://s3.amazonaws.com/lumenlearning/success/Master+PDFs/English-Composition-I-1510697405\\_regexified.pdf](https://s3.amazonaws.com/lumenlearning/success/Master+PDFs/English-Composition-I-1510697405_regexified.pdf)
- McCarthy, P. M., & Ahmed, K. (2021). *Writing the research paper: Multicultural perspectives for writing in English as a Second Language*. Bloomsburg Publications. <https://doi.org/10.5040/9781350164192>
- McCarthy, P. M., Al-Harthy, A., Buck, R. H., Ahmed, K., Thomas, A. M., Kaddoura, N. W., ... Graesser, A. C. (2021). Introducing Auto-Peer: A computational tool designed to provide automated feedback for L2 writers. *Asian ESP Journal*, 17, 9-43
- McCarthy, P. M., Kaddoura, N. W., Buck, R. H., Thomas, A. M., Ahmed, K., Al-Harthy, A., & Duran, N. D. (2021). Metadiscourse and counterargument integration in student argumentative papers. *English Language Teaching*, 14(6), 96-113. <https://doi.org/10.5539/elt.v14n6p96>
- McCarthy, P. M., Kaddoura, N., Al-Harthy, A., Thomas, A. M., Duran, N. D., & Ahmed, K. (2022). Corpus analysis on students' counter and support arguments in argumentative writing. *Pegem Journal of Education and Instruction*, 12(1), 256-271. <https://doi.org/10.47750/pegegog.12.01.27>
- McCarthy, P. M., Renner, A. M., Duncan, M. G., Duran, N. D., Lightman, E. J., & McNamara, D. S. (2008). Identifying topic sentencehood. *Behavior Research Methods*, 40(3), 647-664. <https://doi.org/10.3758/BRM.40.3.647>
- 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>
- McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition & Instruction*, 14(1), 1-43. [https://doi.org/10.1207/s1532690xci1401\\_1](https://doi.org/10.1207/s1532690xci1401_1)
- Olson, C. B., & Land, R. (2007). A cognitive strategies approach to reading and writing instruction for English language learners in secondary school. *Research in the Teaching of English*, 41(3), 269-303. <https://doi.org/10.58680/rte20076014>
- Rass, R.A. (2015). Challenges face Arab students in writing well-developed paragraphs in English. *English Language Teaching*, 8(10), 49-59. <https://doi.org/10.5539/elt.v8n10p49>
- Russell, G. W., & Sewall, E. C. (1972). Serial-position effect and organization of recall of connected meaningful verbal material. *Psychological Reports*, 30(2), 443-446. <https://doi.org/10.2466/pr0.1972.30.2.443>

- Rustipa, K. (2016). The structure and adequate development of English as a Foreign Language learners' paragraphs: A case study. *International Journal of Humanities and Social Science*, 6(2), 133-139.
- Saffari, N., Noordin, S. M., Sivapalan, S., & Zahedpisheh, N. (2017). Transfer of mother tongue rhetoric among undergraduate students in second language writing. *International Education Studies*, 10(11), 23-32. <https://doi.org/10.5539/ies.v10n11p23>
- Shahhoseiny, H. (2015). A study of errors in the paragraph writing of EFL learners: A case study of first year translation students at University of Applied Science and Technology in Bushehr, Iran. *Theory and Practice in Language Studies*, 5(6), 1307-1312. <https://doi.org/10.17507/tpls.0506.26>
- Sharma, V. (2016). Perceptions and perspectives of error analysis in English writing: A case study of Saudi students. *Communication Studies and Language Pedagogy*, 2(1-2), 1-20.
- Sharma, V. (2019). Teacher perspicacity to using rubrics in students' EFL learning and assessment. *Journal of English Language Teaching and Applied Linguistics*, 1(1), 16-31.
- Shokrpour, N., & Fallahzadeh, M. H. (2007). A survey of the students and interns' EFL writing problems in Shiraz University of Medical Sciences. *Asian EFL Journal*, 9(1), 147-163.
- Shpit E. I., & McCarthy P. M. (2023). Novice Russian research writing: Prepositions and prepositional phrases. *Journal of Language and Education*, 9(2), 104-117. <https://doi.org/10.17323/jle.2023.15897>
- Shpit, E. & McCarthy, P.M. (2022). Addressing discourse differences in the writing of Russian engineering students and international researchers. *Language Teaching Research*, 0(0). <https://doi.org/10.1177/13621688221109809>
- Simensen, A. M. (1987). Adapted readers: How are they adapted? *Reading in a Foreign Language*, 4, 41-57. <https://doi.org/10.64152/10125/67003>
- Smith, C. G. (2008). Braddock revisited: The frequency and placement of topic sentences in academic writing. *The Reading Matrix*, 8(1), 78-95.
- Thomas, A. M., & McCarthy, P. M. (2023). Multi-agent teaching. In J. Muller, J. Adamson, S. Herder, & S. Brown (Eds.). *Re-envisioning language EFL Education in Asia*. (pp. 99-117). ITDI.
- Thomas, A. M., McCarthy, P. M., Ahmed, K., Kaddoura, N. W., & Duran, N. D. (2022). A model for teaching paragraph ending sentences to L2 students. *PEGEM Journal of Education and Instruction*, 12(3), 1-11. <https://doi.org/10.47750/pegegog.12.03.01>
- Wali, O., & Madani, A. Q. (2020). The importance of paragraph writing: An introduction. *International Journal of Latest Research in Humanities and Social Science*, 3(7), 44-50.
- Watkins, M. J., & Peynircioglu, Z. F. (1983). Three recency effects at the same time. *Journal of Verbal Learning and Verbal Behavior*, 22(4) 375-384. [https://doi.org/10.1016/S0022-5371\(83\)90237-2](https://doi.org/10.1016/S0022-5371(83)90237-2)
- Weaver, R., Bost, L., Kassorla, M., McKinney-Holley, K., Crowther, K., Curtright, L... Swenson, K. (2016). *Successful college composition* (3rd ed.). Retrieved from [https://oer.galileo.usg.edu/cgi/viewcontent.cgi?article=1007&context=englishtextbooks\\_](https://oer.galileo.usg.edu/cgi/viewcontent.cgi?article=1007&context=englishtextbooks_)
- Yamin, M. (2019). Learning from students' experiences in writing paragraph. *Metathesis: Journal of English Language Literature and Teaching*, 3(2), 188-202. <https://doi.org/10.31002/metathesis.v3i2.1736>