Grammatical subject of research articles: What novice writers can learn from professional writers in writing introduction?

Refnaldi[1*], Andi Muhammad Irawan[1], Slamet Setiawan[2], and Sisri Ovalina[1]

[1]*English Language Education Study Program, Faculty of Language and Art, Universitas Negeri Padang, Jl. Prof. Dr. Hamka, Air Tawar Barat, Kec. Padang Utara, Padang, West Sumatera25171, Indonesia*
[2]*English Literature Study Program, Faculty of Language and Art, Universitas Negeri Surabaya,Jl. Ketintang, Ketintang, Kec. Gayungan, Surabaya, Jawa Timur 60231, Indonesia*

ABSTRACT

Grammatical subjects are critical as they warrant text coherence and assist in reader's understanding of the text. They are of particular importance in a research article introduction to set up the study's context and relevance within the wider academic discourse. This research aims to disclose what novice writers can learn from professional writers in using Grammatical Subjects (GS). It compared GS realizations and functions in Research Article Introductions (RAIs) employed by both professional and novice writers published in applied linguistic-related Journals. One hundred introduction sections from ten journals were purposely selected. 50 of them were authored by professional writers while the other 50 by novice writers. All of these sections were then examined with the use of Gosden's (1993) analytical framework. Findings revealed that both professional and novice writers employed all types of GS. These types encompassed Discourse Participant, Interactive Participant, Discourse Event or Process, Macro and Micro Discourse Entities, Hypothesized Viewpoint, and Empty Real-World Theme. Nevertheless, the results also highlighted noticeable differences concerning the realizations of the Participant Viewpoint, Interactive Discourse Entity, Empty Discourse Theme, Objective Viewpoint, Hypothesized Entity, Empty H and O Themes, Mental Process, Real-World Entity, and Real-World Event or Process. These findings could help novice writers, especially non-native novice writers, navigate the use of grammatical subjects and become aware of their functions and linguistic features to establish coherent ideas in writing a research article introduction section.

Keywords: Academic discourse; grammatical subjects; research article introductions

INTRODUCTION

Research articles are fundamental to the academic world. They often serve as one of the most important means through which scholars disseminate their findings and become engaged with the broader academic community, who increasingly values research output as a measure of success (Chen, 2015; Green, 2008). In higher education institutions, the emphasis on conducting research over teaching has caused lecturers to perpetually improve their competence in writing their scientific works for publication. This is due to the increasing importance of the number of publications as evidence of their academic excellence and knowledge in their field (Schilhan et al., 2021), creating a landscape where the quality can vary widely (Chou et al., 2011). The disparity highlights the necessity for an initiative that is aimed at improving writing skills to produce high-quality research articles, particularly for those with little

experience and lack of adequate training in academic writing (Lin & Yang, 2011). Stensaker et al. (2017) further gives prominence to strategies that foster the development of the essential skills in academic writing, which eventually strengthens not only teaching and learning but also research as well. This calls for a reassessment of the pedagogical approaches used to support academic writing development.

Research articles are composed of several main sections called IMRD (Introduction, Method, Result, and Discussion) (Ahmed & Afzal, 2020). Of these sections, the introduction section is pivotal with multiple functions. The section serves as a gateway for the readers and a roadmap for the study. It lays the foundation for the entire work, establishing a space for authors to delineate the context and novelty, argue for its significance, and outline the main objectives and contribution of the study (Zein et al., 2023). In addition, it enables the authors to situate their current study in the wider discourse community and among the existing previous studies to capture the interest of the readers (Bajwa et al. 2020; Swales, 1990). Characteristics of the introduction section, however, may vary across different academic fields. Its effective writing follows particular genre conventions that conform to the discipline of which it is affiliated to (Bajwa et al., 2020). Suryani et al. (2018), for instance, provided evidence-based insights into the unique characteristics of twenty introduction sections in computer science research, demonstrating how disciplinary norms influence the authors' rhetorical choices. These findings suggest that writers, both professional and novice, must be informed with such disciplinary standards to produce a coherent research article introduction (RAI). Novice writers, in particular, must be able to navigate these conventions carefully, adapting their writing to align with the expectations of their specific fields. They can benefit from analyzing successful examples in their fields to understand how professional writers articulate their research context and objectives.

The literature reveals a growing body of research investigating RAIs in the last two decades with a particular focus on its structural and rhetorical elements of various research articles (i.e., Abdi & Sadeghi, 2018; Akoto & Afful, 2020; Gray, 2018; Kafes, 2018; Ozturk, 2018; Piromkij & Phoocharoensil, 2022). Gray (2018) examined the evolving nature of introduction sections of medical research articles and their critical role in engaging readers from the outset, providing an educational review of how to create effective RAIs for this field. Conversely, Ozturk (2018) highlighted the structural comparison of RAIs in Applied Linguistics research, discovering that RAIs with subsections were more rhetorically organized compared to those without. Kafes (2018) further studied the rhetorical organization of RAIs in the field of social sciences,

consisting of 75 RAIs written in English by American academic writers and in English and Turkish by Turkish academic writers. Its findings indicated that the three corpora followed the adapted version of Swales' (1990) framework by employing three moves, namely establishing a territory, providing a niche, and occupying the niche, to a great extent. A similar result was also displayed in Chemical Engineering RAIs. Piromkij and Phoocharoensil (2022) identified move analysis and lexical bundles with the same three conventional moves. Variability in establishing writer authority and credibility, on the other hand, was evident in the Applied Linguistics RAIs (Abdi & Sadeghi, 2018). The study showed that professional writers tended to use a variety of grammatical constructions that effectively position their research within the existing body of knowledge allowing them to assert centrality and relevance. Meanwhile, most novice writers struggled with this aspect, leading to less impactful introductions. Metadiscourse, or the use of language to organize and comment on the discourse itself, plays a crucial role in establishing an authorial voice. Akoto and Afful (2020) explored variations in writers' usage of metadiscourse within RAIs, emphasizing its importance for clarity and engagement. This finding is in line with Ebrahimi (2017) who examined the functional aspects of grammatical subjects in RAI, noting that effective use of metadiscourse can enhance the perceived centrality of the author's contribution. Understanding how to employ metadiscourse is of significance for novice writers, who may struggle to assert their voice in their writing, empowering them to create more engaging and persuasive RAIs. In light of these findings, a focus on RAI should offer a fruitful avenue for research intervention.

The importance of effective writing in academic contexts cannot be overstated. Writing RAI is more than a mere formality; it is a critical component that influences the research reception. The review of the literature encompasses a diverse range of studies focusing on the writing of RAI. There is, however, a dearth of research on the use of grammatical subjects (GS) in RAI. Ebrahimi (2017) conducted a study examining GS usage in RAI across four disciplines, namely Applied Linguistics and Psychology (representing soft sciences), and Chemistry and Environmental Engineering (representing hard sciences). The study involved a corpus of 40 RAIs, 10 from each discipline, which were analyzed with a modified model concerning grammatical subject types and discourse functions (Ebrahimi, 2014). The findings showed that the nature of the RAIs served as a guide in the selection of the appropriate grammatical subject. The study, however, focused on the frequency of the categories of GS used in the two main science divisions. In addition, the primary determinants of the discourse functions of the GS types consisted of the divisions

of the hard and soft sciences and the internal structure of the RAIs. Therefore, there is a need for a focus on a more comprehensive functional analysis of the use of GS, especially in the field of Applied Linguistics, as one of the critical elements in writing RAI.

Such a focus will considerably benefit novice writers who often grapple with effectively crafting RAIs. It is also particularly relevant to Indonesian context where English, the primary language of academic publications, is a foreign language still to be mastered by the writers. Articulating ideas coherently in the language along with mastering the art of effectively constructing RAI is essential and concurrently challenging for many novice writers in Indonesia. Critically examining the deployment of GS will provide the best evidence of professional writers' effective writing practices and expertise, which novice writers can glean from to develop their own RAIs. In addition, by examining the patterns and strategies employed by the professional writers, the novice writers who are at the early stages of their research careers with little experience in research writing can gain valuable insights into the expectations and conventions of academic discourse. The present study then aims to contribute to the ongoing dialogue with a thorough analysis of the use of GS in RAIs of applied linguistics journals. Findings from this study are expected to provide assistance to novice writers in crafting the path to success in their research publications in reputable international journals.

The concerns previously addressed have led to the formation of the following research questions:

1. What are the similarities and differences between GS used in RAIs written by professional writers and novice writers?
2. What can the novice writers learn from the professional writers in the use of GS to develop their RAIs?

METHOD
Research Design
To address the research questions, a mix of quantitative-qualitative methodology (QQM) was employed. This methodology allows for collection of various types of GS deployed by professional writers in applied linguistic journal articles and more empirically visible research results (Grim et al., 2006). The types of GS were determined using a percentage formula, while discourse analysis was selected as the primary approach to elaborate quantitative data outcomes. The JASP 0.8.5.1 program, Classic Test Theory Independent Sample T-Test and Bayesian Independent Sample T-Test were utilized for the data analysis.

Data Collection
This study analyzed 100 articles published between 2016 and 2021 in applied linguistics journals. It focused on the Introduction section only, as this section is a specific and essential 'core section' (Dudley-Evans, 1997), and a critical section where the author reviews and evaluates what other researchers have explored or discovered in previous relevant studies (Arsyad et al., 2018; Swales, 1990). In addition, the selection of this particular number of research articles from the applied linguistics journals is essential to provide a representative sample of professional writing practices. This focus ensures that novice writers can observe diverse grammatical subject usage, thereby enhancing their understanding of effective RAI writing in their specific academic field. Of these 100 articles, 50 were authored by professional writers published in 5 different reputable international journals, consisting of TESOL Journal, TESOL Quarterly, Journal of L2 Writing, RELC Journal and ELT Journal with 10 articles from each journal. The rests were authored by novice writers, which were also published in 5 different journal, namely the ELT Forum Journal of English Language Teaching, Journal of English Language and Education, Journal of English Language Teaching, Journal of English Teaching and Research and Wiralodra English Language Journal with also 10 articles from each. The analysis of these articles would identify four domains of GS: Participant, Discourse, Hypothesized and Objectivized, and Real-world domains (Gosden, 1993).

Analysis Framework
Gosden's (1993) categorization of grammatical subjects (GS) was chosen as the basis for analyzing the data in this study. He classifies GS into four domains that indicate its discourse functions. The classifications include Participant, Discourse, Hypothesized and Objectivized, and Real-world domains (See Figure 1). The Participant domain constitutes: (a) Discourse participant that uses the first-person pronouns 'we' and 'I' as GS; (b) Participant viewpoint that refers to the pronoun 'our' (For example: our analysis); and (c) Interactive participant that deals with citation (For instance: David, 1987). The second domain is the Discourse domain that consists of: (a) Discourse event or process that refers to the discourse act or process of the reporting (For example: the conclusion); (b) Macro discourse entity that is an integral unit of discourse (For instance: this paper); (c) Micro discourse entity that is concerned with internal entities of a discourse (For example: the result); (d) Interactive discourse entity that is indicative of the previous study (For instance: the previous studies); and (e) Empty discourse theme: it emphasizes domain criteria through 'it/there' (example: it is concluded).

The last two domains are composed of Hypothesized and Objectivized, and Real-world domains. The Hypothesized and Objectivized domain encompasses: (a) Hypothesized viewpoint that refers to comment or judgment of the research (For example: it is a possibility); (b) Objectivized viewpoint which reflects evaluative modification an acknowledgment or status (For example: one reason); (c) Hypothesized entity that indicates the testing and meaning of the expression (For instance: the design); (d) Empty hypothesized and objectivized theme that is characterized by the pattern (For instance: there is evidence). The last domain is the real-world domain. This domain includes several subcategories: (a) Mental process which focuses on intellectual processes and entities that are part of the investigation (For example: deduction); (b) Real-world entity that deals with physical or material entities in research (For instance: participants, subjects); (c) Real-world event or process that describes the action or procedure in resulting research activities (For example: A focus on the project as the main role of activity can be evaluated); (d) Empty real-world theme that is process or reference to mental process of research (For example: it was found).

Figure 1
The Classification of Grammatical Subject (Gosden, 1993)
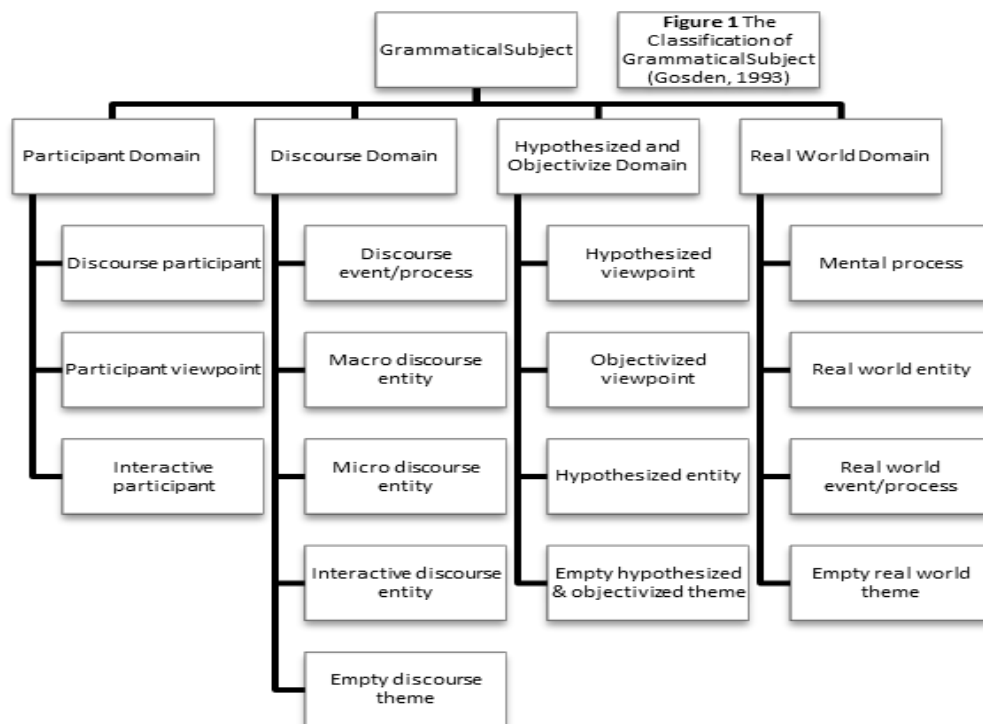
Data Analysis

Classic Test Theory Independent Sample T-Test and Bayesian Independent Sample T-Test were used to analyze the corpora and calculate the results accurately. First, the classification of GS is adopted from Gosden's classification (1993), which consists of mainly four domains as previously discussed. Each domain used by both professional and novice writers in the articles was highlighted. The identified domains were calculated based on the location from which journal they were employed. Second, the calculation result was rechecked manually to minimize incorrect identification. Every GS domain identified and categorized was retrieved only from independent clauses in the selected RAIs. The collected data were qualitatively examined to address the research questions.

The quantitative analysis was conducted to obtain statistical data, which were calculated with the Classic Test Theory Independent Sample T-Test and Bayesian Independent Sample T-Test. The quantitative analysis results supported the descriptive results to answer the types of GS commonly used by novice writers. The use of QQM accurately revealed the different categories that the novice and professional writers used in RAIs.

FINDINGS AND DISCUSSION

Descriptive Analysis

Participant Domain

Participant domain is intended to describe the way writers deliver their thoughts or ideas with the use of pronouns in writing RAIs. It consists of three subcategories, which are (1) Discourse Participant, (2) Participant Viewpoint, and (3) Interactive Participant. The frequency of each subcategory of the participant domain is presented in Figure 2.

Figure 2
Types and Frequency of Participant Domain
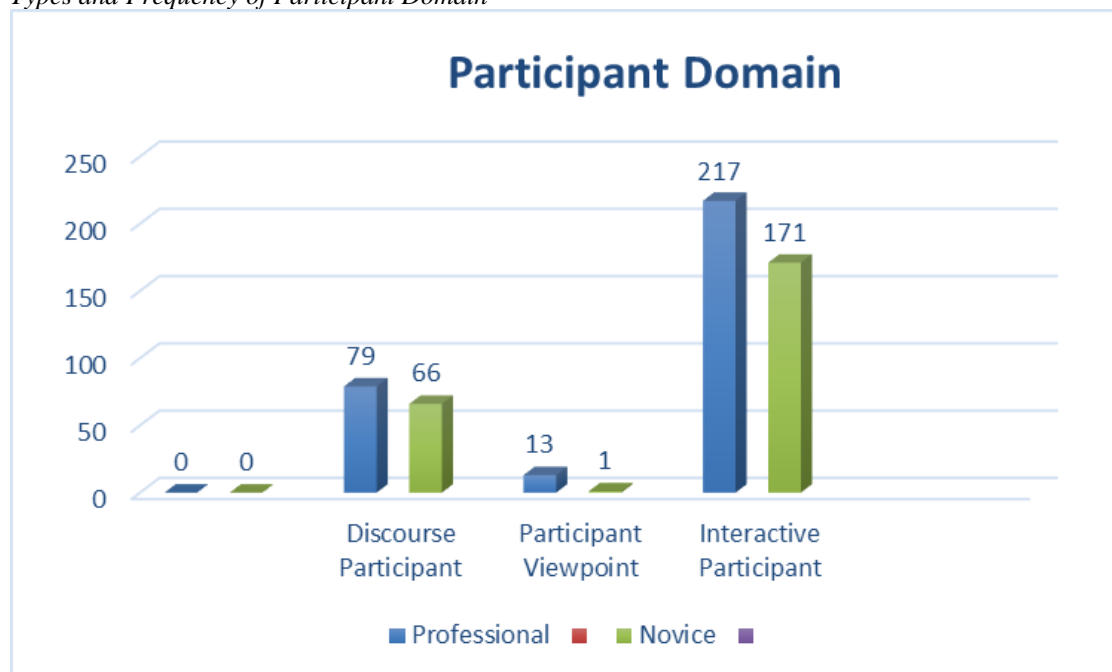
As shown in Figure 2, the Interactive Participant was predominant among the three subcategories of the Participant Domain for both professional and novice writers. In addition, there were similarities in the use of Interactive Participants and Discourse Participants. The use of Interactive Participants among novice writers was slightly higher than that of professional writers with 71.85% and 70.23% respectively. A similar phenomenon was also observed with the use of Discourse Participants with 27.73 % among novice writers and 25.57 % among professional writers.

In contrast, the Participant Viewpoint's usage as GS was different from the other domains. Professional writers used as much as 4.21%, while novice writers only used 0.42% of this domain. The discourse function of the Participant Viewpoint is presenting the author's perspective and finding (Ebrahimi, 2017). It is crucial in writing article, serving as one of the essential parts in delivering research findings based on the data previously analyzed. Based on this result, it can be concluded that professional writers were more successful in providing and writing findings compared to novice writers. Many novice writers seemed to not pay attention to the importance of finding, as they wrote it to a lesser degree in constructing academic articles.

The discrepancy between professional and novice writers could have been more balanced. Professional writers used multiple participant domains in their writing, while novice writers rarely used various Participant Domains. The main function of Participant Domains was to state the aim, method, and finding of the study. This is supported by Ebrahimi and Farnia (2022) who argued that professional writers described the aim, method, and finding in various ways to shape the reader's understanding and show clear-cut information about a study.

The presence of the Participant Viewpoints can be seen in the following examples:

*(1) **Her research** showed that Finding the Main Idea, Implied Detail Question, Unstated Detail Question, and Unstated Detail Question belong to four difficult skills. (Novice writers)*
*(2) **Our comprehensive analysis** of the published empirical studies in this journal over its first 25 years aims to indicate how the field of L2 writing has developed over time in terms of salient contextual, theoretical, and methodological aspects. (Professional Writers)*

Discourse Domain
Essentially, this domain delves into how writers manage the flow of thought and presentation of information in their writings. This section examined how writers employ grammatical subjects within their research articles, encompassing various types such as "Discourse Event/Process," "Macro Discourse Entity," "Micro Discourse Entity," "Interactive Discourse Entity," and "Empty Discourse Theme. The frequency of each subcategory of the participant domain is presented in Figure 3.

Figure 3
Types and Frequency of Discourse Domain
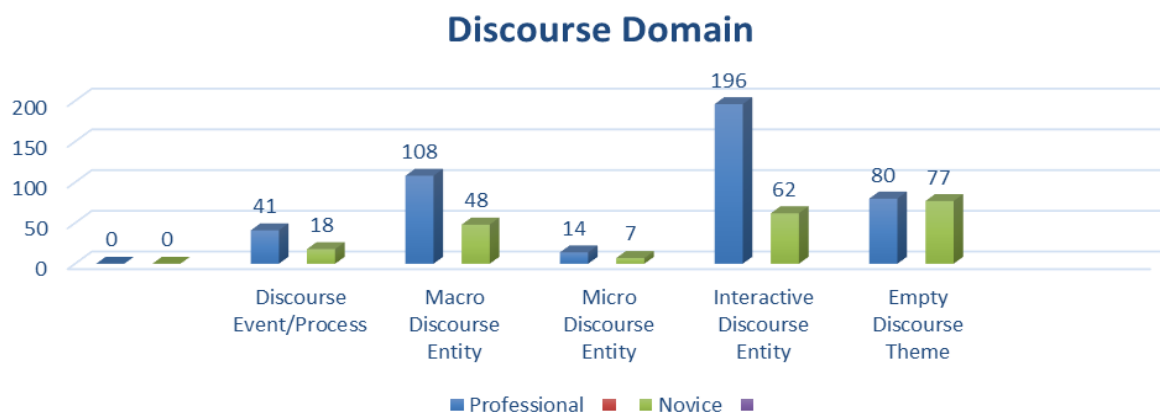
The result indicates that the dominant subcategories of the Discourse Domain.used by professional writers and novice writers were different – the Interactive Discourse Entity for professional writers and the Empty Discourse Theme for novice writers. The interactive Discourse Entity had the highest usage percentage in RAI. A similar pattern was also seen in the use of the Discourse Event/ Process, the Macro Discourse Entity, and the Micro Discourse Entity in RAI. 9.34% of Discourse events/processes were used by professional writers, while it was only 8.49% by novice writers. The use of Macro Discourse Entity by professional writers, on the other hand, amounted to 24.60%, and 22.64% by novice writers. Similarly, the Micro Discourse Entity was used as much as 3.19% by professional writers and 3.30% by novice writers in writing RAI.

Nevertheless, professional writers and novice writers significantly differed in using the Interactive Discourse Entity and the Empty Discourse Theme. It reached 44.65% for professional and 29.25% for novice writers regarding the use of the Interactive Discourse Entity. This implies that professional writer argued their statement by looking at the previous study to pinpoint the gap within the current research, while novice writers were limited to describing the previous studies in their writing, and it was not as proper as the professional ones. According to Wilang et al. (2018), novice writers need to attend workshops, seminars, or research article coaching to elevate their understanding and level up their confidence in writing research articles.

The discourse function of the Interactive Discourse Entity as GS is to indicate the community's validation through the presentation of some data from previous studies (as seen in the following examples 3 and 4). This may convince the scientific literature readers to believe in the study being investigated or the result of the study.

(1) *Several studies (e.g. Ellis, Sheen, Murakami, & Takashima, 2008; Bitchener & Knoch, 2010; Sheen, 2010) also supported this argument,* *showing that WCF can be helpful for students' writing accuracy.* (Novice Writers)

(2) *Several studies investigated motivation in ESP courses.* (Professional Writers)

The use of the Empty Discourse Theme reached 18.22% by professional writers and 36.32% by Novice writers. In this domain, novice writers elaborated their statements by using evaluative comments and providing arguments that they read from previous studies. They described their information to emphasize the main idea of the earlier theories. As Ebrahimi and Chan (2015) pointed out, this implementation is based on the premise that producing these evaluative comments will help convince readers and the community in the field that the writers are familiar with the subject under study.

This domain has several different discourse functions. Thus, the findings showed that the use of these subcategories as GS was to: 1) provide evaluative comments on previous relevant studies, and 2) indicate a gap in the existing literature. They added that the writers can present their position in case of the literature existence and explain their study needs (See examples 5 and 6).

(3) *It is necessary to use Instagram for writing discussions in paragraph writing courses for English Department students. (Novice Writers)*

(4) *In other words, it is necessary to fine-tune the 'quality' descriptors and the strategy descriptors in both the CEFR and the ELP using linguistic information on what the learners are expected to do at each CEFR level, as analyzed through learner corpora.* (Professional Writers)

Pointing out the gaps in the existing literature is another function of the Empty Discourse Theme (See examples 7 and 8). The writers need to clearly indicate the significance of their study, like stating the reasons why they conducted this study and what makes their study differ from previous studies. As stated by Ebrahimi (2017), writers need to show the

importance of their research in relation to the existing professional literature.

(5) ***There were just a few studies*** *that tried to examine or analyze how Team Pair Solo or pictures are used by the teacher in the teaching and learning process.* (Novice Writers)

(6) ***It is widely acknowledged*** *that integrated writing is a fundamental skill that both first language (L1) and second language (L2) writers need to successfully complete academic writing tasks (Cumming, 2013; Currie, 1998; Gebril, 2009; van Weijen, Rijlaarsdam, & van den Bergh, 2019; Weigle, 2004).* (Professional Writers)

Hypothesized and Objectivised Domain

Within the framework of this descriptive analysis, the use of "Hypothesized and Objectivised Domain" was explored to shed light on how writers employ GS to convey their viewpoints and hypotheses in RAI. This domain is divided into four primary subcategories: "Hypothesized Viewpoint," "Objectivized Viewpoint," "Hypothesized Entity" and Objectivized Entity." The frequency of each subcategory of the participant domain is presented in Figure 4.

Figure 4

Types and Frequency of Hypothesized and Objectivised Domain
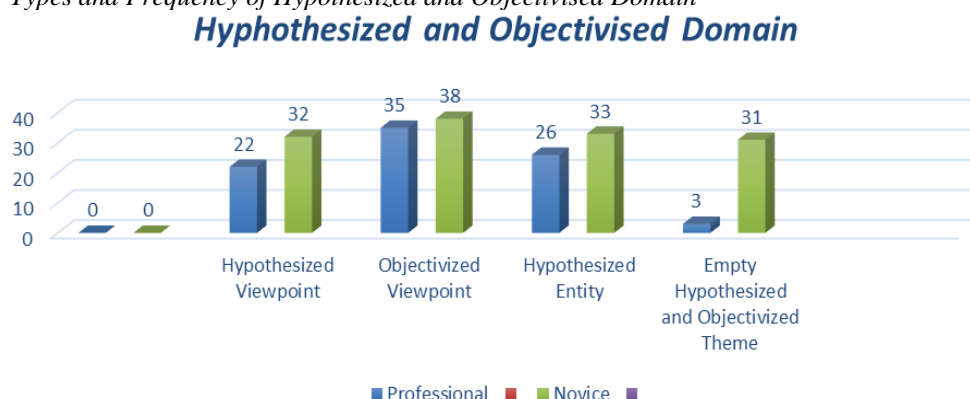
The analysis result reveals that the use of the Objectivized Viewpoint as GS dominated among all the other domains with 40.70% by the professional writers and 28.36% by the novice writers. A similar trend was also found with the Hypothesized Viewpoint. It reached 25.58% by professional writers and 23.88% by novice writers. On the other hand, they have total differences in the Objectivized Viewpoint, the Hypothesized Entity, and the Empty Hypothesized and Objectivized Theme. The use of the Objectivized Viewpoint by professional writers amounted to 40.70% and 28.36% by novice writers, while the same trend was also seen in the use of the Hypothesized Entity with 30.23% by the professional writers and and 24.63% by the novice writers. The use of the last sub-domain, the Empty H and O theme was approximately 3.49% by professional writers and 23.13% by novice writers. According to Gosden (1993), the Objective Viewpoint usually represents a highly judgmental correction toward perceived or given status, characterized by the timeless use of the present tense (See examples 9 and 10). The significant difference between professional and novice writers in the use of the Objectivized Viewpoint as GS in RAI occurred with approximately 35% and 38% respectively. It can be said that novice writers were rather judgmental in their RAI writing, as they provided minimal correction in their writing.

Another subcategory is in these articles is the Hypothesized Entity (See examples 11 and 12). Gosden (1993) stated that GS with this subcategory

refers to how research is tested and conducted and how it is expressed. The last subcategory is the Empty Hypothesized and Objectivized Theme (See examples 13 and 14).

(9) ***Much of this work*** *has focused on how interaction can be influenced by factors of gender, ethnicity, and the role of social relationships of learners and their interlocutors, and by the nature of topics, tasks, and activities in which they engage.* (Novice Writers)

(10) ***These factors*** *can differentiate in part among classroom tasks. However, the current study draws mainly on the dimension of communicative stress that deals directly with the stakes and nature (speaking) involved in our research.* (Professional Writers)

(11) ***This technique*** *is very effective to be used by the teacher in the class because this technique can make the students taught in the class active in speaking activity.* (Novice Writers)

(12) ***This model*** *posits four types of articles based on two binary features of referentiality (see Table 1), depending on whether the NP in which the article appears is a specific referent and whether the NP can be assumed to be known to the hearer.* (Professional Writers)

(13) ***It can be seen*** *from the students' scores that the average writing was only 62, and if the teacher asked the students to write a sentence, almost all of the students had difficulty writing that sentence because there is a difference between English pronunciation and how it is written.* (Novice Writers)

(14) ***There are*** *two interpretations of zero article overuse.* (Professional Writers)

Real World Domain
In the context of descriptive analysis, the focus shifts to the "Real World Domain," where the examination centres on how writers utilize GS to portray elements from the real world within their research articles. This domain comprises a variety of subcategories, encompassing "Mental Process," "Real World Entity," "Real World Event/Process," and Empty Real World Theme. The frequency of each subcategory of the Real World Domain is presented in Figure 5.

Figure 5
Types and Frequency of Real World Domain
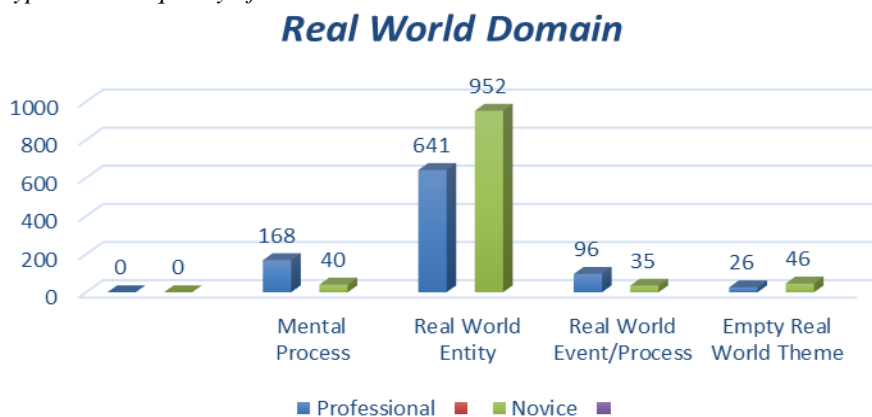
This research reveals the distinction of the use of the Real World Domain between professional and novice writers. The professional writers employed all subcategories, and it was significantly higher than novice writers did. The only similarity was the use of the Empty Real World Theme with 2.79% and 4.29% by professional and novice writers respectively. However, the use of the other three subcategories, Mental Process, Real World Entity, and Real World Event/ Process.was different, In terms of using the Mental Processes as GS in RAIs, professional writers used this category as much as 18.05%, while novice writers as low as 3.73%. On the other hand, the Real World entity contributed 68.85% to the total use of GS by professional writers and 88.72% by novice writers. This indicated that novice writers tended to use the Real World Entity in almost all sentences in their RAIs. The Real World Entities indicate physical things or an object that appears in writing RAI. The finding of Ebrahimi and Farnia (2022) showed that all professional writers tended to use the Real World Entity in their research article introduction.

Last, novice writers used the Real World Event/ Process as GS in their RAIs amounting to 3.26%, while professional writers used it as much as 10.31%. This domain has different discourse functions in RAI. The first category was the Mental Process, which indicates the process by using normalized form (See examples 15 and 16).

(15) **The ways** *they express their ideas are not similar and depend on their characteristics.* (Novice writer)

(16) **Course descriptions** *clarify that improving reading and translation skills and developing specialized vocabulary knowledge are the main course objectives. (Professional writer)*

The second category was the Real World Entity, which contained two discourse functions: the object or the related step or procedure in methodology and stating the significance of the study. The first function refers to the object or real entity (See examples 17 and 18), and the second function is shown in examples 19 and 20.

(17) **Grammar** *is one of the language components essential to be mastered since it consists of a range of rules that are useful in establishing good communication* (Novice writer)

(18) **Self-access centers** *(SACs) have become an increasingly important means of promoting independent learning in educational settings for the past 30 years* (Professional writer)

(19) *Based on the information obtained by* **the researcher** *in SMA Negeri 1 Salo, the students faced problems in learning to speak* (Novice writer)

(20) **Some of the participants**' *overuse errors involved post-modification, as in I have heard of *the parents who don't give their children enough to eat.* (Professional writer)

The last subcategory was real-world event/process, only presenting the event from which the results are generated (Examples 21 and 22).

(21) **The development of IT helps the English teaching and learning process** *become more effective and efficient* (Novice writer)

(22) **With the increasing interest in learner-centered approaches in the field of foreign language education, how to promote autonomous language learners** *has become a prominent concern of many researchers (Benson, 2001).* (Professional writer)

Box Plot and Whisker Plot Analysis

The RAIs were then analysed to discover the normality of the data and the direction of the data distribution. The data were analyzed using SPSS 21 followed by a box plot and a whiskar plot. The box plot method divided the input data set into quartiles, consisting of the minimum, the lower, the median, the upper quartile, the outliers, and the maximum values (Thirumalai, et al, 2017). Lines that are extensions (up or down) of the box are called whiskers. Values that are above or below the whiskers are called outlier values and extreme values. The outliers are data values 1.5 times greater than the box length (IQR) as measured by the UQ (top of the box) or LQ (bottom of the box). Meanwhile, Streiner (2018) and Simbolon, et. al (2022) noted that extreme values are values three times greater than the box length (IQR) measured from UQ (top of the box) or LQ (bottom of box).
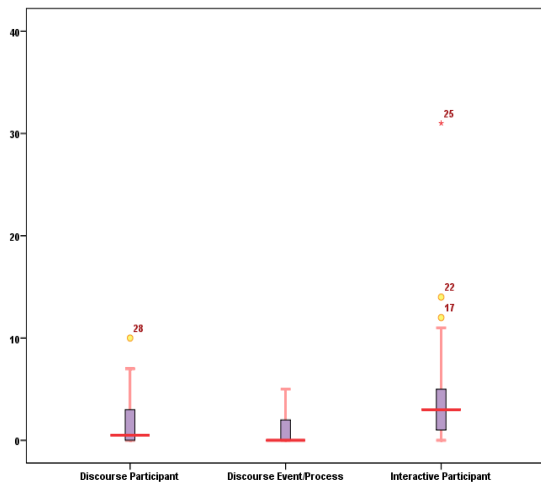
Participant Domain

Figure 6 shows that the variation of participant domain data on the Discourse Participant, the Interactive Participant and the Discourse Event/ Process. The position of the median line for the Discourse Participants and the Discourse Events is not in the middle position, which was in contrast to the Interactive Participant, whose median line position was in the middle. The three Whiskar groups were not in a symmetrical position, and they tended to be in the upper position so that the data distribution stretched to the right (positive skewness). The Discourse Participant had an outlier value of 28, while the Interactive Participant had outlier values of 17, 22 and an extreme value of 25.This situation illustrates the distribution of data that is not normally distributed.

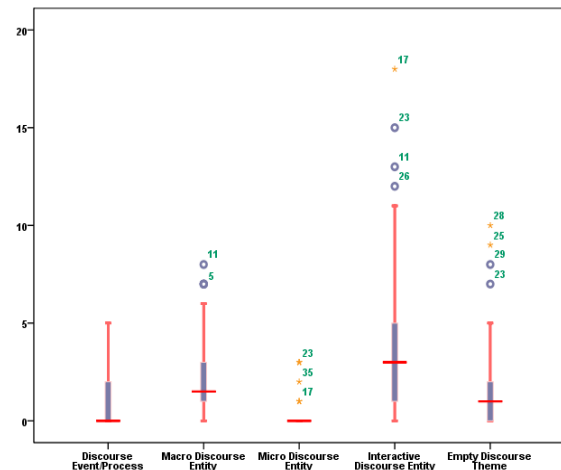**Figure 6**

Participant Domain

Figure 6 shows that the variation of participant domain data on the Discourse Participant, the Interactive Participant and the Discourse Event/ Process. The position of the median line for the

Discourse Participants and the Discourse Events is not in the middle position, which was in contrast to the Interactive Participant, whose median line position was in the middle. The three Whiskar groups were not in a symmetrical position, and they tended to be in the upper position so that the data distribution stretched to the right (positive skewness). The Discourse Participant had an outlier value of 28, while the Interactive Participant had outlier values of 17, 22 and an extreme value of 25.This situation illustrates the distribution of data that is not normally distributed.

Discourse Domain

The median line positions of the Interactive Discourse Entity and the Empty Discourse Theme in the Box Plot above was situated in the middle, which was different from the Discourse Event, the Macro Discourse Entity, and the Micro Discourse Entity. In the Box Plot above, only the Discourse Events and the Micro Discourse Entity were excluded as outliers. The extreme value of the Interactive Discourse was 17, and the Empty Discourse Theme had the extreme values of 28 and 25. Meanwhile micro discourse entity contained a lot of extreme values, i.e., 23, 35, and 17. The data distribution tended to be right-biased (positive skewness) because the one-sided Whiskers was situated at the top. This indicates that the entire dataset is not normally distributed.
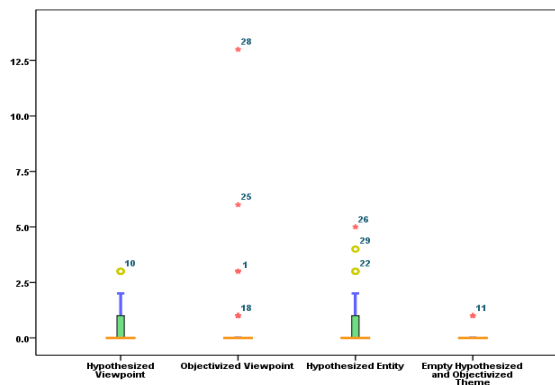
Figure 7

Discourse Domain

Hypothesized and Objectivized Domain

The four data groups in Figure 8 were not symmetrical (the data were not normally distributed). This can be seen from the position of the data median, the outlier values, the Whiskers, and the extreme data values. In all four data groups as a whole, the position of the media line was not in the middle position. There were extreme values in Hypothesized Entity reaching 26, the Empty Hypothesized and Objectivized Theme with 11. This was in contrast to the Objectivized Viewpoint,

which had more extreme values of 25, 1, and 18. The data distribution was more right-biased (positive skewness) because the Whiskers of the aggregated data tended to be on top.

Figure 8
Hypothesized and Objectivized Domain

Real World Domain
Figure 9 clearly shows that the data variation from the Real World Domain criteria for the Mental Process, the Real World Entity, the Real World Event, and the Empty Real World Theme. The median line position for all data groups was not centered. The data distribution was stretched to the right (positive skewness) because the Whiskers in the four groups were not in symmetrical positions and the top four whisker groups tended to be longer. In the Mental Process, there were outlier values of 11 with extreme values of 22 and 23. The Real World Entity had outlier values of 3 and 23 and extreme value of 42, while the Real World Event and the Empty Real World Theme only had outlier values of 26, 2 and 36. This situation illustrates the distribution of data that was not normally distributed.

Figure 9
Real World Domain
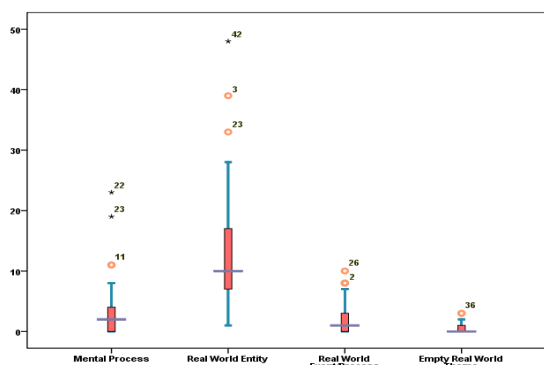
Bayesian Analysis
The normality test showed that the data were not normally distributed. The Bayesian-sample t-test (Bianchi & Hoe, 2021) was then conducted to provide solution related to the issues of the data.

The Bayesian analysis was appropriate when the data was not normally distributed. This view is also supported by Rigollet (2016) who argued that Bayesian analysis contained power procedures to the whole distribution, as the analysis provides some information with persistence accuracy. It emphasized the data proven transparently and accurately. The data can be seen in Figure 10:

Figure 10
Inferential Plots (Prior and Posterior)
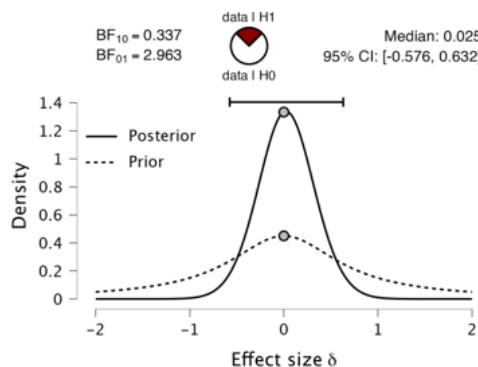
Variable t-tests with Bayesian analysis established the condition that there was a significant difference between professional and novice level variables with a percentage error of 3.5%. The test also gave an estimated posterior value of the difference between the two data sets, as described in the inference plot. The latter values indicated a significantly higher estimate of the difference compared to the empirical values of the former conditions.

Figure 11
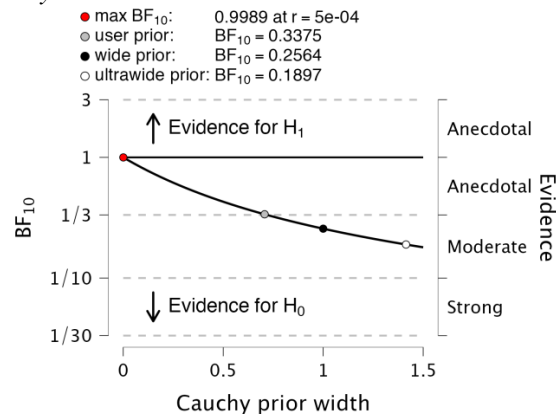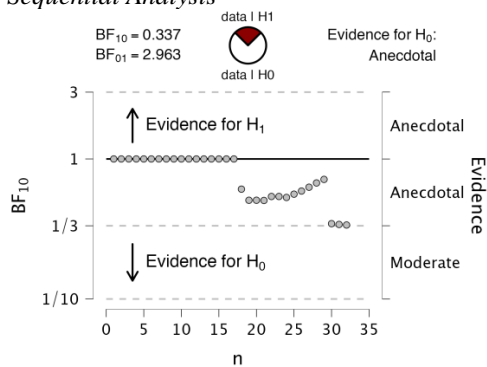*Bayes Factor Robustness Check*

Figure 12
Sequential Analysis

The sequence analysis showed a high probability of recurrence of changes between the professional and novice variables in the future. The plot illustrates that with n more than 15, there is still a tendency to accept H1, indicating a difference.

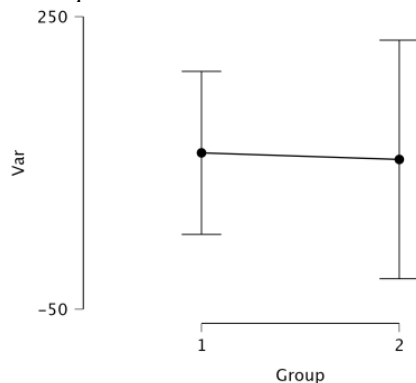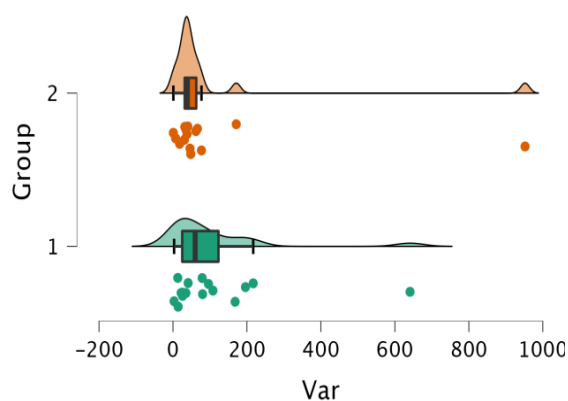**Figure 13**
Descriptive Plots

Figure 14
Raincloud Plots Var

A Raincloud plot visualizes the difference between a professional article (G1) and a novice article (G2). Data are commonly combined with bar charts, violet charts, and jitter charts to visualize the strength between the two data sets. Figures 13 and 14 showed that the G1 values were better than the G2 values. Moreover, the distribution of data in G1 appeared to be more stable across each measurement range. The purpose of the Participant Domain was to engage the readers in reading the research and present interactional writing to catch the point of view of the authors in the research. This is supported by Martin (1986) and Martinez (2001) that the use of the Discourse Participant could make the article more interactional. According to Almor and Nair (2007), the provision of pronouns in research articles is reflected in the subject form or first-mention antecedent. It emphasizes the authors' statement in conveying the ideas into the writing.

Professional writers tended to use the Interactive Discourse Entity to reference to previous studies conducted by other researchers and to analyse those studies in identifying the similarities and differences. This is in line with Mack's (2018) previous study, which is one of the criteria in researching because the writers can learn what has been done by other researchers and collect all the information before starting to conduct their own study. On the other hand, novice writers used the Empty Discourse Themes more frequently than professional writers, indicating the significance of the research. The novice writers showed the importance of their study in stating the reasons of conducting the study and what makes their study differ from the other previous studies. It is also supported by Ebrahimi (2017) who stated that writers needed to show the importance of their research in relation to the existing professional literature.

The use of the Hypothesized and Objectivized Domain demonstrated that comments written by professional writers were of higher quality than that by novice writers. The highest domain was the Objectivized Viewpoint used by professional writers, signifying the evaluations or comments of the writers in a research article. The provision of a decision or statement is elaborated to level up the understanding of the content of the study (Hinchliffe, 2020). Last is the distinction of the Real World Domain used by professional and novice writers. The use of the subcategories of the Real World Entity and the Empty Real World Theme by professional writers was less frequent than novice writers. Professional writers used the subcategories adequately by stating the object or related the entity that appears in writing. According to Kopche et al. (2010), the Real World Domain serves as a bridge to identify the set of data and the real object that can be found in a research article, and it is the basic element in delivering who, what, and where of the study.

The discussion above demonstrated that the Bayesian analysis provided advantages to this study. It proved that Bayesian analysis can adapt the data that are not normally distributed. This is supported by Garcia (2018) that Bayesian analysis for second language acquisition provided advantages which is not only for non-normal distribution but also for easy interpretation, which can be inferred to prior research. The ongoing study can reflect to the

previous study and visualized the data with the plot in parameter not the data themselves.

Learning from Professional Writers

Having conducted the descriptive analysis, as well as the Box Plot, the Whiskar analysis, and Bayesian analysis, the findings revealed what novice writers can learn from professional writers. First, novice writers have to use more sentences or clauses with the use of the Participant Domain as GS, especially from the Participant Viewpoint. In research articles, the use of pronouns to demonstrate the writers' viewpoint to make the readers involved in the writers' ideas is allowed. Chen (2018) FThe development of a research article emphasizes certain personal pronouns for a given context, or even the presence or non-presence of a personal pronoun in journal articles, and particularly in scientific journal articles, can often reveal how writers view themselves, their relationship with the readers, and their relationship with the discourse community they belong to (Wang et al., 2021). This can be called the strategy of the writers to show their contribution to what is being investigated and to build a connection with the point of view of the readers. The writers also shape the language style by using pronouns to format their arguments or opinions in their research articles. This is also supported by Hyland (2003) who explains that an effective scientific paper actually depends on interactional elements that supplement propositional information in the text and alert readers to the writer's opinion. Öz (2022) further states that a research article will be influential when the writer provides interactive ways to engage the reader.

In addition, novice writers need to use more sentences or clauses with the Discourse Domain as GS. In term of the Interactive Discourse Entity and the Empty Discourse, the novice writers need to elaborate strong arguments with evaluative judgment. It means they can comment on any ideas in the research articles, which make it become readers' point of view in understanding the whole research articles. According to Xie (2020), producing these evaluative comments will help convince readers to understand the writers' opinions. In addition, the writers have to look back to the previous research articles as references in conducting the research articles. The novice writers must review previous studies to discover the critical point of the research articles and the gap between previous and current articles. They also need to build correlation to the previous research articles, providing space for the writers to identify the current research article with the review of some areas of previous articles (Nicholson et al., 2018). He also argued that previous study can be extended to the scope of knowledge and information that the earlier writers give. Thus, the writers can elaborate the content of articles vividly.

On the other hand, the novice writers need to reduce the use of the Hypothesized and Objectivized domains, especially for the subcategories of the Empty Hypothesized and Objectivized themes. It means the novice writers need to control the sentences with this domain as the GS. The domain refers to the clarification's design and formula by using the words "it is or there". This word can be ambiguous when used improperly and can be misinterpreted for readers in understanding the information in the research articles. According to Hyland (2003) and Hotaling (2020), writing a scientific paper considers the language structure to avoid misinterpretation and error of the content. It implies the readers can understand the meaning of the research articles and gain meaningful insight.

The last point to consider is the need for novice writers to reduce the reliance on the Real World Elements, especially when depicting the Mental Processes, the Real World Entities, and the Real World Events or Processes. The novice writers must learn how to strike a balance within their sentences or phrases when composing their statements. In terms of the use of the Real World Entities, it is advisable for writers to minimize the use of physical objects or entities in their writing in order to avoid unnecessary wordiness in research articles. The writers can enhance their writing by employing diverse vocabulary to captivate readers while perusing their articles. This approach aligns with Hyland's (2003) assertion that novice writers should train to develop and enhance their writing skills, aspiring to become proficient writers.

CONCLUSION

Both professional and novice writers utilized various grammatical subject (GS) types, with similarities and differences apparent within each subcategory of these GS types. They share common ground in the use of the Discourse Participants, Interactive Participants, Discourse Events/ Processes, Macro and Micro Discourse Entities, Hypothesized Viewpoints, and Empty Real World Themes. Nonetheless, distinctions arise in the employment of the Participant Viewpoints, Interactive Discourse Entities, Empty Discourse Themes, Objectivized Viewpoints, Hypothesized Entities, Empty H and O Themes, Mental Processes, Real World Entities, and Real World Events/ Processes.

In light of these findings, novice writers can derive valuable insights from professional writers when it comes to crafting Research Article Introductions (RAIs). First and foremost, the novice writers should consider incorporating more sentences or clauses that emphasize the use of the Participant Domain, particularly from the perspective of the participants themselves. Secondly, they should focus on the expansion of

sentences or clauses within the Discourse Domain. Regarding the Interactive Discourse Entities and the Empty Discourse Themes, the writers can benefit from developing robust arguments enriched with evaluative judgments. Thirdly, it is essential for them to minimize their reliance on the Hypothesized and Objectivized Domains, particularly in the subcategories of the Empty Hypothesized and Objectivized Themes. Lastly, the novice writers should also reduce the dependence on the Real World domains, especially when incorporating the Mental Processes, the Real World Entities, and the Real World Events/ Processes.

The implications of these findings suggest that the novice writers can greatly enhance their skills in crafting RAIs by learning from professional writers. They should prioritize incorporating more sentences or clauses that highlight the perspective of the participants, expanding sentences or clauses within the Discourse Domain, and strengthening arguments with evaluative judgments, especially in relation to the use of the Interactive Discourse Entities and the Empty Discourse Themes.

REFERENCES

Abdi, J., & Sadeghi, K. (2018). Promotion through claiming centrality in L1 and L2 English Research Article Introductions. *International Journal of English Studies*, *18*(1), 53–70. https://doi.org/10.6018/ijes/2018/1/297381

Ahmed, I., & Afzal, M. T. (2020). A systematic approach to map the research articles' sections to IMRAD. *IEEE Access*, *8*(July), 129359–129371. https://doi.org/10.1109/ACCESS.2020.3009021

Akoto, O. Y., & Afful, J. B. A. (2020). Variations in metadiscourse use in english language introduction and literature review thesis chapters. *Language Literacy: Journal of Linguistics, Literature, and Language Teaching*, *4*(2), 390–408. https://doi.org/10.30743/ll.v4i2.2601

Arsyad, S., Zaim, M., & Susyla, D. (2018). Review and citation style in research article introductions: A comparative study between national and international English-medium journals in medical sciences. *Discourse and Interaction*, *11*(1), 28–51. https://doi.org/10.5817/DI2018-1-28

Bajwa, N. u. H., König, C. J., & Kunze, T. (2020). Evidence-based understanding of introductions of research articles. *Scientometrics*, *124*(1), 195–217. https://doi.org/10.1007/s11192-020-03475-9

Chen, C. Y. (2015). A Study Showing Research has been Valued over Teaching in Higher Education. *Journal of the Scholarship of Teaching and Learning,* 15 (3), 15-32. doi:

10.14434/ josotl.v15i3.13319

Cheng, A. (2018). *Genre and graduate-level research writing*. University of Michigan Press.

Chou, C. P., Wu, R. F., & Hu, J. (2011). The impact of SSCI on Taiwan's academic evaluation in humanities and social sciences: A case study of NCCU. *Journal of Comparative Education, 70*, 31-56. https://ah.lib.nccu.edu.tw/bitstream/140.119/71407/1/31-56a.pdf

Ebrahimi, S. F. (2014). *Thematicity in English academic research articles across disciplines in hard and soft sciences* [Unpublished Doctoral Thesis]. UPM Malaysia.

Ebrahimi, S. F. (2017). *"The overall aim of this work is…": Functional analysis of grammatical subjects in research article introductions across four disciplines. Discourse and interaction*, *10*(1), 5-30. https://www.ceeol.com/search/article-detail?id=762276

Ebrahimi, S. F., & Farnia, M. (2022). *"We examine the relationship …": Grammatical subject in research article abstracts: Disciplinary and functional analysis. Journal of Applied Linguistics and Applied Literature: Dynamics and Advances 10*(1), 99–118. https://doi.org/10.22049/jalda.2022.27278.1321

Gosden, H. (1993). Discourse functions of subject in scientific research articles. *Applied Linguistics*, *14*(1), 56–75. https://doi.org/10.1093/applin/14.1.56

Gray, J. A. (2018). Introduction sections: where are we going and why should I care? *AME Medical Journal*, *3*, 112–112. https://doi.org/10.21037/amj.2018.12.03

Green, R. G. (2008). Tenure and promotion decisions: The relative importance of teaching, scholarship, and service. *Journal of Social Work Education,* 44(2). 117-127. https://doi.org/10.5175/JSWE.2008.200700003

Hotaling, S. (2020). Simple rules for concise scientific writing. *Limnology And Oceanography Letters*, *5*(6), 379–383. https://doi.org/10.1002/lol2.10165

Kafes, H. (2018). A genre analysis of English and Turkish research article introductions. *Research on Youth and Language*, *12*(1), 66–79. https://eric.ed.gov/?id=ej1177684

Lin, J. Y., & Yang, S. W. (2011). The study of human resource management of faculty in public Universities in Taiwan. *Journal of National Taichung University, 25*(2),85-110. https://www.airitilibrary.com/Article/Detail/18176402-201112-201205050006-201205050006-86-111

Nicholson, J. D., LaPlaca, P., Al-Abdin, A., Breese,

R., & Khan, Z. (2018). What do introduction sections tell us about the intent of scholarly work: A contribution on contributions. *Industrial Marketing Management*, *73*(February), 206–219. https://doi.org/10.1016/j.indmarman.2018.02.014

Öz, G. (2022). A Comparative corpus-based analysis of boosters in NESs and NNESs' academic articles. *IJELTAL (Indonesian Journal of English Language Teaching and Applied Linguistics)*, *7*(1), 1. https://doi.org/10.21093/ijeltal.v7i1.1128

Ozturk, I. (2018). Rhetorical organization of the sub-section of research article introductions in applied linguistics. *Novitas-ROYAL (Research on Youth and Language)*, *12*(1), 52-65. https://eric.ed.gov/?id=EJ1177689

Schilhan, L., Kaier, C., & Lackner, K. (2021). Increasing visibility and discoverability of scholarly publications with academic search engine optimization. *Insights: The UKSG Journal*, *34*, 1–16. https://doi.org/10.1629/UKSG.534

Stensaker, B., Bilbow, G. T., Breslow, L., & van der Vaart, R. (2017). *Strengthening teaching and learning in research universities: Strategies and initiatives for institutional change*. Springer.

Suryani, I., Ahmad, N., Zubir, F., Ghazali, N., & Aclan, E. M. (2018). Description on research article introduction section by computer science researchers in Malaysian Private Universities. *Journal of Human Development and Communication*, *7*, 1–8. https://johdec.unimap.edu.my/images/PDF/JO HDEC%20DIS%202018/Vol_7_2018_1-8.pdf

Swales, J. M. (1990). *Genre analysis: English in academic and research setting*. Cambridge University Press.

Wang, S. P., Tseng, W. T., & Johanson, R. (2021). To we or not to we: Corpus-Based research on first-person pronoun use in abstracts and conclusions. *SAGE*, *11*(2). https://doi.org/10.1177/21582440211008893

Piromkij, W., & Phoocharoensil, S. (2022). A corpus-based study of move analysis and lexical bundles of chemical engineering research article introductions. *Journal of Positive School Psychology*, *6*(6), 10656–10672. https://journalppw.com/index.php/jpsp/article/view/9692

Wilang, J. D., Jantori, P., & Chutataweesawas, S. (2018). Worries of novice researchers in writing research papers. *Proceeding of International and National Conference on Learning Innovation in Science and Technology 2018* (pp. 171-180).

Xie, J. (2020). Chinese mas' evaluation in english academic writing: A student-oriented perspective. *Journal of Language Teaching and Research*, *11*(5), 739–748. https://doi.org/10.17507/jltr.1105.08

Zein, T. T., Sinar, T. S., Nurlela, N., Lubis, T., & Rosa, R. N. (2023). Theme structure in an introduction section of articles published in Indonesian national journals. *Studies in English Language and Education*, *10*(1), 251–265. https://doi.org/10.24815/siele.v10i1.25181