# A genre, scoring and authorship analysis of AI-generated and human-written business refusal emails: Is there still a case for business writing training?

Winny Wilson

MSc in Applied Linguistics for Language Teaching, 2024

# DECLARATION BY THE CANDIDATE

# AS AUTHOR OF THE DISSERTATION

DEPARTMENT OF EDUCATION

UNIVERSITY OF OXFORD

1. I understand that I am the owner of this dissertation and that the copyright rests with me unless I specifically transfer it to another person.

2. I allow the Department to deposit on my behalf a copy of this dissertation in the Oxford University Research Archive ('ORA') where it shall be freely available online for use in accordance with ORA's Terms and Conditions of Use [https://ora.ox.ac.uk/terms_of_use].

3. I understand that this dissertation should not contain material that can be used to personally identify individuals or specific groups of individuals (unless permission has been obtained from the individuals) and that such material should be removed before this dissertation is deposited in ORA.

4. I agree to be bound by the terms of the ORA Grant of Non-exclusive Licence [https://ora.ox.ac.uk/deposit_agreements] and I warrant that to the best of my knowledge, making my thesis available on the internet will not infringe copyright or any other rights of any other person or party, nor contain defamatory material.

5. I agree that my dissertation shall be available for download in ORA in accordance with paragraphs 2, 3 and 4 above.

| | |
|---|---|
| Signed [an electronic signature is sufficient]: | Winny Shum |
| Date: | 26/8/2024 |

## Abstract

In a business context, writing negative messages such as refusals to requests is particularly challenging, as the act of refusal can damage the relationship between writer and reader. Generative artificial intelligence (AI) tools have shown the capacity to assist with various writing tasks, but do these tools remove the need for business writing training? This study analysed the genre moves and steps that typify business refusal texts generated by ChatGPT ($n$ = 18) and Gemini ($n$ = 18) alongside human-written texts ($n$ = 18). To assess writing quality, experienced business communication instructors ($N$ = 36) scored the texts on an established rubric for evaluating business written professional communication. Assessors also made authorship judgements as to whether each text was AI-generated or human-written and provided rationales for their decisions. The views of the instructors regarding the use of AI tools for writing in the workplace were also captured.

The genre analysis showed that AI-generated texts were formulaic and largely followed the same moves and steps, regardless of text type or audience, whereas human-written texts were more nuanced and flexible in their adoption of genre conventions. On writing quality, human-written texts achieved the highest mean scores, followed by Gemini texts and finally ChatGPT. Inferential statistics revealed that Gemini texts did not differ significantly in the rubric evaluation scores compared to human-written texts, although Chat-GPT produced texts received significantly lower evaluations of writing quality. Participants were able to identify authorship of AI-generated texts with an accuracy rate of 68.1%, and human-written texts were correctly identified 86% of the time. Qualitative data revealed that assessors were most concerned with three broad characteristics of the texts: tone, relationship and formality; language choice and grammatical accuracy; content, detail and structure. Insights into key differences between AI-generated texts and human writing drawn from these results are used to inform five key areas of focus for teaching business writing in the AI age.

# Acknowledgements

Many people were involved in the process of completing this dissertation. First of all, I would like to thank Professor Heath Rose, my supervisor, who has been a supportive and encouraging sounding board for ideas throughout the entire two-year course. When the original plan fell through, Heath was there to help brainstorm ideas for the design of a new study and provide guidance. Working with Heath proved to me that completing a degree at Oxford while living in Asia is both possible and rewarding.

Next, I would like to acknowledge my employer, the British Council, and my colleagues who work there for their generous support. Thank you to trainers and teachers from different parts of their world who volunteered their time and expertise to act as participants and the human writers for this project. I am especially grateful to my team in Singapore for all their help. Thanks to lobbying by senior colleagues in Corporate English Solutions, I received financial support for my studies under the Professional Development Scheme. I hope many more will benefit from this scheme in future. In addition, it has been rewarding to have colleagues express interest in my research and to be able to contribute to shaping the organisation's response to artificial intelligence.

To the classmates in my ALLT cohort – thank you for keeping me company on this journey. It was wonderful to have met you. Special thanks are due to our course representative, Filip, for his energy and efficiency in keeping us updated and getting questions answered. To my "Spoons" group - I have treasured your WhatsApp messages for showing me scenes of life in your parts of the world, and for bringing calm during moments of panic.

I leave the people closest to me for last. My family in Singapore – Brad, Camille, Peter, Teddy and Rea – you have been involved in the everyday ups and downs and have been so patient and encouraging for two years. I'll be expressing my thanks to you in many ways for many years. Finally, I am lucky to have two sets of loving parents in Australia – mum and dad, Pat and Allan – who always cheer me on in whatever I do.

# Table of Contents

# List of tables

# List of figures

# 1 Introduction

The ideas and aims for this study were conceived in a corporate training department within a global organisation well known for providing English language courses. This department focuses on helping professionals develop the communication skills needed for success in the workplace. Training on written communication skills is in high demand, as companies seek to help their employees improve the quality and effectiveness of their writing. Many business writing courses include a large component on the writing of negative messages, such as when a writer needs to say no to a reader's request. These messages are a particularly challenging aspect of business correspondence, as considerable skill is required to ensure the act of refusal does not damage the relationship between writer and reader.

The motivation for this study began with the desire to conduct detailed analysis of business refusals, in order to add to the scholarship on genre approaches to teaching writing in English for Specific Purposes (Bhatia, 1991; Paltridge, 2014; Swales, 1990 among others). This remains an aim, as genre awareness, being an understanding of typical patterns in texts written with a communicative goal in mind, is a very powerful tool for learners. While initial ideas and sources of text for this research were being considered, OpenAI released ChatGPT. The capabilities of the technology, especially its ability to generate natural-sounding text in response to a prompt, elicited both excitement and concern in this researcher's context. The potential for generative artificial intelligence (AI) tools such as ChatGPT to be used for assisting in, or outright completing, writing tasks at work provoked concern among professional skills trainers: were we going to lose our jobs? How would we respond to client and learner questions on whether AI tools would make our writing workshops obsolete? What can human writers do that AI tools cannot?

Given the very recent emergence of AI tools and the rapid changes that have occurred since, research has struggled to keep up. At time of writing, there were no published studies comparing human-written and AI-generated texts in a business context. To fill the gap, this exploratory study investigates business refusal texts from three writer groups: ChatGPT, Gemini and Human. First, genre analysis is employed to understand the typical moves and steps of texts from each writer group and the differences between each. Next, human assessors are asked to blind evaluate the texts by scoring them based on established business writing criteria, with the aim of finding out whether AI tools generate texts that are considered higher-quality than human writing. To gather information on whether there are identifiable and characteristic qualities of AI-generated versus human-written texts, participants are asked to identify authorship of the texts and explain what influences their decisions. Finally, as developments in generative AI will continue to have effects on the work of business writing trainers, their views on using AI tools for writing in the workplace are captured.

This exploratory study is designed with practical applications for business communication teachers in mind. Chapter 2 considers previous studies on business genres and generative AI in the business writing context, although these are limited. Research questions are presented at the end of the literature review. Chapter 3 outlines this study's mixed-methods design, along with the processes for data collection and analysis. Chapter 4 presents results for each of the research questions, followed by detailed discussion of findings in Chapter 5. Finally, Chapter 6 uses insights drawn from the data to inform implications for teaching, with five key areas of focus for teachers of business writing in the AI age. The concluding chapter also considers the limitations of the current study, along with recommendations for future research.

# 2 Literature Review

This chapter provides an overview of the concepts and previous research relevant to the topics under study. First, background on genre research and its aims is summarised, with a focus on move structure analysis. Second, genre research in the context of business communication is considered further, particularly studies done on strategies and structures for business refusals. The third section of the literature review considers generative AI within the business context. A brief overview of AI tools is provided, along with a summary of research into the application of these tools in business writing. Previous research into the ability of humans to distinguish between AI-generated and human-written texts is outlined. Finally, studies on the views of business communication instructors towards AI tools are introduced. After exploring the existing literature, gaps in the research are considered and the research questions for this study outlined.

## 2.1 Overview of genre research

### 2.1.1 Background

Genre is an important construct in the study of discourse in general, particularly written discourse, and has gained much attention over the past four decades (Tardy & Swales, 2014). Genre studies can be broadly grouped under two approaches: rhetorical and linguistic (see Hyon, 1996 and Kong, 2014 for detailed reviews). Rhetorical approaches are interested in the connections between genres and how they facilitate social action in institutions and society at large, often employing ethnographic research methods (Miller, 1984; Bazerman, 1988). Miller suggests that learning a genre is not simply about learning linguistic forms and conventions, but also about learning how to "participate in the actions of a community" (1984, p. 165).

While also interested in the function of genres to carry out actions and purposes, linguistic approaches are known for their pedagogical focus. The English for Specific Purposes (ESP) framework uses descriptions of genres as discourse models for writing instructors to apply in their teaching (Bhatia, 1993; Flowerdew, 1993; Swales, 1990). This approach is interested in structures and how genres are employed in a discourse community, being a group of people sharing basic values and assumptions on how to use communication to achieve common goals. According to Swales (1990, p. 58):

*A genre comprises a class of communicative events, the members of which share some set of communicative purposes. These purposes are recognised by the expert members of the parent discourse community and thereby constitute the rationale for the genre. This rationale shapes the schematic structure of the discourse and influences and constrains choice of content and style.*

Under this definition, participants share a common focus and preference for structure, style and content for their communicative activity to typify a genre. In addition to textual knowledge, this requires awareness and understanding of the shared practices of the discourse community and their choice of genres in order to perform their everyday tasks (Bhatia, 2004).

Hyland (2003, 2007, 2015) has written extensively on genre-based pedagogy as a means of providing a contextual framework that offers an explicit understanding of how texts in target genres are structured to engage effectively with their readers. Hyland sees genre analysis as “probably the most important item in the ESP toolbox” (2022, p. 206), with few people having explicit knowledge of the rhetorical and formal features of everyday texts. This knowledge is important and genre awareness is an essential element of ESP’s core mission of preparing learners to use English in their target contexts: the situations in which they hope to study, work and live (Hyon, 2018). To develop genre awareness, ESP genre research identifies salient features of texts which allow writers to engage effectively with their readers and seeks to show how people accomplish tasks through language (Hyland, 2015). Genres are dynamic and open to change, reflecting norms and ideologies in discourse communities (Paltridge, 2000). Thus, while linguistic knowledge is necessary for effective communication, writers also need to understand the expectations of the context they are writing in.

Zhai and Razali’s (2023) recent systematic review of studies into genre-based approaches to teaching writing since 2003 found that between 2014 to 2021, there was a flourishing of research in the ESP field. This was driven by the popularity of genre-based writing instruction aiming to raise awareness of the conventions of specific writing contexts. However, existing studies focus mostly on academic writing in science and social science-based disciplines, revealing a notable gap in genre-based studies within the business disciplines, with only two articles falling under this subject area out of a total of 52 articles reviewed (Zhai & Razali, 2023). The present study argues that a genre-based approach is relevant for business writing training, as knowledge of the expectations, conventions and requirements of workplace communicative settings gives professionals the tools to achieve their communicative goals. More research into the business disciplines is thus needed and this study presents an analysis of an under-researched business genre and implications for teaching.

### 2.1.2 Move structure analysis

Multiple methods of genre analysis exist, including text analysis, move structure analysis, comparative genre analysis and critical genre analysis (Dudley-Evans, 1994; Tardy & Swales, 2014). Move structure analysis will be discussed in detail here, as this study makes use of the method to analyse texts of the same genre across three writer groups. Move analysis involves analysing a corpus of texts considered representative of the genre of interest to identify moves and steps. Moves are defined as carrying out distinct functions within the text, with steps being sub-categories within each move. John Swales (1990, 2004) conducted pioneering work on academic writing genres and is best known for the Create A Research Space (CARS) model, which described the organisational pattern of research article introductions. Swales saw research article introductions as especially challenging for research article writers, in part because they involved decisions about how to get the attention of their readers (Hyon, 2018). By analysing a sample of article introductions in several science and social science fields, Swales identified three common moves, each containing three or four steps within. Table 1 shows the CARS model, updated in 2004 to make it suitable to describe research article introductions of various disciplines.

**Table 1: CARS model for research article introductions (Swales, 2004, pp. 230-233)**

| |
|---|
| **Move 1: Establishing a territory** |
| Step 1: Topic generalisations of increasing specificity |
| |
| **Move 2: Establishing a niche** |
| Step 1A: Indicating a gap/or |
| Step 1B: Adding to what is known |
| Step 2* (optional): Presenting positive justification |
| |
| **Move 3: Presenting the present work** |
| Step 1 (obligatory): Announcing present research descriptively and/or purposively |
| Step 2* (optional): Presenting RQs or hypotheses |
| Step 3* (optional): Definitional clarifications |
| Step 4* (optional): Summarising methods |
| Step 5 (PISF): Announcing principal outcomes |
| Step 6 (PISF): Stating the value of the present research |
| Step 7 (PISF): Outlining the structure of the paper |

*These steps are not only optional but less fixed in their order of occurrence than others
PISF = probable in some fields, but unlikely in others

Swales defines the move as “a discoursal or rhetorical unit that performs a coherent communicative function in a written or spoken discourse” (2004, p. 228), which then serves the major communicative purpose of the genre. The linguistic realisation of the move can be as short as a clause or as long as a paragraph and sometimes repeated in cycles (Moreno & Swales, 2018). The function of the move is realised by steps, which can be regarded as sub-moves and have particular communicative purposes related to the move. Vergaro (2002) posits that analysing the move structure of a text thus means to assign a pragmatic function to a stretch of language, and to build the schematic structure through which its communicative purpose is achieved. Move analysis thus organises text boundaries to give insight into how writers achieve their purpose (Swales, 1990; Vergaro, 2004). In addition to consideration of whether moves are obligatory or optional, genre analysts have considered their importance by referring to the percentage of texts in which a move appears to inform discussion about how typical, predominant or conventional certain moves are in a genre (Biber et al., 2007; Henry & Roseberry, 2001). The information gained from move analysis can thus help writers make decisions on how they might structure their texts, taking into account the conventions of the discourse community in which their genre is situated.

A large body of work follows the Swalesian tradition, focusing on the academic and research space (e.g., Alharbi, 2021; Cotos et al., 2015; Kwan, 2006; Muangsamai, 2018; Peacock, 2002). These follow-up studies adopting the CARS model found that each specified move often occurs in research article introductions regardless of discipline and language, but preferred steps for the realisation of each move vary. Swales (1990) mentions that within the overall generic structure, cultural variation is possible and writers can make specific rhetorical choices. While early genre analyses were mostly carried out in the academic domain, the concept has also been adopted in other ESP contexts. Bhatia (1991, 1993, 2008) extended genre analysis into professional domains and, along with other researchers (e.g., Evans, 2012; Zhang, 2013), sought to equip learners with the ability and skills to engage with genres most likely to be utilised in the business world.

## 2.2 Genre research in business communication

### 2.2.1 The “business letter” as a genre

Taking Swales’ (1990) definition of genres presented above, the “business letter” as a broad term would not be classified as a genre. While a letter performs a communicative function, there are different indications of purpose depending on the writer’s intention (although it could be said that an underlying purpose of all business writing is to initiate or maintain relationships). Nevertheless, Van Nus (1999) contends that a business letter constitutes a genre because of its recurrent schematic structure comprising a subject; an opening salutation; prepropositional, propositional, and postpropositional sections; and a closing salutation. This characterisation is overly simplistic as most genres incorporate what amounts to a beginning, a middle and an end. However, specific move structures and the order in which they are made would be expected to differ depending on the contents and aims of the business letter.

Going some way to alleviate the debate, Swales (1990) identifies the concept of genre repertoire, or a group of genres, within a discourse community. Hence, within the genre of "business letter", it is possible to distinguish genre repertoires such as applying for jobs (Bhatia, 1993; Henry & Roseberry, 2001; López-Ferrero & Bach, 2016; Tongpoon-Patanasorn & Thumnong, 2020), requesting (Kong, 1998; Nguyen & Miller, 2012; Park et al., 2021), negotiating (Pinto dos Santos, 2002; Townley, 2021), promoting sales (Bhatia, 1993; Lima-Lopes, 2006; Vergaro, 2004; Zhu, 2000), chasing money (Vergaro, 2002) and so on, all within the business discourse community. By analysing genre-specific corpora, researchers can investigate the different types of writing skills and strategies needed for specific professional writing tasks (Upton & Connor, 2001). To date, much of the ESP genre literature has focused on letters, faxes and memoranda. However, the emergence of email and its widespread use has seen emails explored in terms of textual features (Gains, 1999), register (Gimenez, 2000), pragmatics (Ho, 2011), communicative purpose (Alafnan, 2015) and the design of email tasks for the Business English classroom (Evans, 2012). A few studies have compared letters and emails, finding differences in terms of formality (Perez Sabater et al., 2008) and other linguistic features (Baron, 1998). This review does not aim to compare letters and emails, but rather emphasises that the communicative purpose of each repertoire within the business letter genre remains the same, whether sent as a physical letter or electronically as an email.

### 2.2.2 Refusals and politeness strategies

There is a wide range of communicative purposes within the business letter genre repertoire, with requests featuring prominently in previous studies. A sensible next step would then be to consider the actions that follow after a request has been received. In general, there are two possible responses: the request is granted (good news), or the request is refused (bad news). Few studies in the professional writing literature have considered refusals, despite the importance of the genre in the business context. Writing "bad news" or negative messages such as refusing a request is difficult and anxiety-inducing (Schryer, 2000), but a skill critical to interpersonal effectiveness in the workplace (de Rycker, 2014). The act of refusal can potentially damage the relationship between writer and reader and requires considerable thought and skill in constructing the message. Decock et al. (2020) define a refusal as:

1. A speech act by which a speaker expresses his or her will not to engage in action proposed by the interlocutor
2. A face-threatening act
3. Characterised by linguistically softened, face-saving devices to maintain social harmony (e.g., apologies and explanations for the refusal)

In their work on politeness, Brown and Levinson (1987) explained the notion of "face", or public self-image, which can be lost, maintained or enhanced and must be constantly attended to in interaction. Face is mutually vulnerable and as people can be expected to defend their faces if threatened, it is in each participant's best interest to maintain each others' face by employing "politeness" strategies for lessening the threat (Brown & Levinson, 1987). These strategies include the use of explanations to maintain goodwill when composing negative messages. Campbell (1990) theorised that citing reasons for denying a request is the most difficult strategy to employ, as it requires effective audience analysis and inventive effort on the part of the writer. Students should hence be offered instruction on how best to formulate explanations, with the particular situation of each request dictating the strategies available for refusal. Similarly, Creelman (2012) emphasises a need for process over product in pedagogical practice, as a specific formula or pattern should not be used for every message without considering the process of audience and context analysis.

In addition to explanations, scholars of linguistic politeness have proposed that "you-attitude", defined as viewing a situation from the other person's point of view, should be included in the teaching of business and technical communication (Rodman, 2001). Shelby and Reinsch (1995) contrast you-attitude with I-attitude, arguing that you-attitude speaks directly to the reader in order to establish a connection between the author's information and the reader's wants, whereas I-attitude focuses on the writer's perspective. However, you-attitude does not simply imply the dominance of the second pronoun "you" rather than the first person pronoun "I" in a text. Indeed, in negative situations, advice from some textbooks is to avoid the word "you" and use passive verbs to protect the reader's ego and avoid assigning blame (Rodman, 2001). Rather, you-attitude entails showing empathy, being sensitive to the reader's perspective and addressing their concerns.

### 2.2.3 Structuring refusals

Although Campbell (1990) and Creelman (2012) advocated a case-by-case approach to formulating refusal strategies, being able to see a broad structure of moves and steps is useful, particularly for learners who are new to the genre. This allows newcomers to the discourse community to familiarise themselves with typical performances of the genre, as well as observing differences that may occur in move frequency and step sequencing. Unfortunately, there is a scarcity of research on refusal messages in the workplace. Scholars attribute this to the difficulty of collecting data for analysis, given that organisations are often unwilling to share sensitive and negative communication with customers (Creelman, 2012; Limaye, 2001). This was also the case for this current study, which had planned to analyse a corpus of employee-written refusals to customers, before pivoting to AI-generated texts when the target organisation ultimately did not give permission.

Schryer's (2000) study of negative letters is notable for having been conducted "from the inside", as the researcher was granted access to the insurance company in which the correspondence was written. Schryer had been asked by the company to develop workshops on negative correspondence and requested the opportunity to first research the writing and production practices of refusal texts in the company. The study combined textual analysis of 26 insurance claim denial letters with writers' accounts of producing these letters. It drew on the analytical frameworks of genre theory, discourse analysis and sociocultural theories to describe the perspectives of managers, writers and the researcher on what constituted an effective negative letter. Schryer identified the rhetorical organisation of the letters as following a traditional structure: a neutral buffer opening, the policy explanation, the medical explanation, the decision and a closing section.

The existence of the explanation corresponds with Campbell's (1990) work, and there is further support in the literature for including an explanation as a component designed to justify the writer's decision and help persuade the reader to accept the refusal and its rationale (Creelman, 2012). Limaye (2001) notes that people can invent reasons for rejection that are worse and more damaging than they actually are, so organisations should seize the initiative to provide explicit explanation and shape reader perception. A little less clear is where the explanation should appear in the generic sequence, with debate on whether negative messages should (1) start with a buffer and an explanation before delivering bad news - an indirect approach; or (2) immediately begin with bad news, followed by an explanation - a direct approach (see Creelman, 2012 for a review).

Textbooks mostly recommend the indirect approach as a way to maintain the recipient's face needs and goodwill, through a sequence of rhetorical moves typically involving an explanation and rationale before presenting the negative news (Creelman, 2012; De Rycker, 2014). Locker (1999) provided a summary of traditional textbook advice on negative messages, comprising six principles:

1. Use a buffer (a neutral or positive sentence enabling the writer to delay the negative information).
2. Explain why you are refusing.
3. Place the reason before, not after, the refusal.
4. Phrase the refusal itself as positively as possible.
5. Offer an alternative or a compromise, if one is available.
6. End on a positive, forward-looking note.

However, rather than supporting this advice, Locker's (1999) study questioned the effectiveness of the indirect organisational approach by investigating reader responses. Testing of simulated credit and graduate school refusal letters with college students revealed that the presence or absence of buffer statements or positive endings had no effect, but responses were least favourable when the rejection was unexpected and other options were limited. Based on the findings, Locker recommended that the refusal be given directly, but that reasons should follow, as well as alternatives and compromises (e.g., a counter-offer) if they exist.

Other studies support the widely recommended indirect approach. Shelby and Reinsch (1995) studied reader reactions to business memoranda using factor analysis of message attributes. They found validation for the use of positive emphasis and concluded that recipients preferred the use of an indirect message organisation scheme in negative situations. In addition, the experiments of Jansen and Janssen tested the effects of adding and combining positive politeness strategies, finding that giving an explanation has a positive effect on the reader's evaluation of the letter (2010a) and that the explanation-first, indirect order was superior (2010b, confirmed in 2013). In reality, the degree of (in)directness depends on context, including factors such as the type of audience, whether the negative news concerns a routine business matter, the writer's anticipation of the reader's response, the writer's and the reader's cultural and personality characteristics, and so on (De Rycker, 2014). This supports the idea that while genre analysis can reveal useful insights into text patterning, a specific formula should not be used for every message without considering the audience and context.

Taken together, evidence on writing effective “bad news” messages suggests that it is context-specific, difficult for even trained professionals, anxiety-inducing, and potentially damaging of professional relationships. Refusals in particular are a critical part of business communication and thus often form part of the content of many ESP courses on business writing. However, existing research on the business refusal genre is now dated and there appears to have been little recent interest in the topic. In addition, the introduction of artificial intelligence (AI) tools to workplace communication (discussed next) is a new area and research has not yet caught up. It remains to be seen how effectively AI can be used to generate negative messages that achieve their intended communicative purpose within the genre.

## 2.3 Generative AI and the business context

### 2.3.1 Overview of generative AI tools

The term “artificial intelligence” (AI) was first coined in 1956, but has received renewed and feverish attention due to recent developments in natural language processing that have made it possible for users to generate new content using AI tools (Baron, 2023). The technology, known as generative AI, uses learning algorithms to generate human-like responses to text-based prompts. In November 2022, OpenAI released ChatGPT, an AI model based on GPT-3 (Generative Pretrained Transformer) technology. GPT-3 was founded on a large language model (LLM) trained on hundreds of billions of words and images from diverse domains, genres and sources, including books, articles, websites, and other online content (see Fraiwan & Khasawneh, 2023 for further information on GPT and LLMs). The corpus of data that the LLM draws upon is continually updated and expanded and the latest update, GPT-4, was released in March 2023. As of May 2024, ChatGPT 4.0 is a paid subscription service, and the free version uses the GPT-3.5 model.

The introduction of ChatGPT elicited a mix of enthusiastic and alarmist responses on the opportunities and threats presented by the technology (Rudolph et al., 2023). Users quickly found that, compared to other AI systems, ChatGPT was significantly better at understanding questions from people and responding in human-like language (Cardon et al., 2023a). It can answer questions in a comprehensive way; generate different types of text, such as but not limited to letters, poems, computer code or musical pieces; and translate languages. ChatGPT can mimic many aspects of human writing, but commentators (e.g., Barrot, 2023) have found that some qualities of human writing, such as emotional depth, voice and identity, and rhetorical flexibility remain a challenge. Berber Sardinha (2024) investigated ChatGPT-generated texts across different registers including academic, conversation, L2 essay and news, finding that AI's ability to capture the complex patterns of natural language remains limited. Despite the perceived shortcomings, ChatGPT set a record for the fastest growing user base of any consumer application in history (Hu, 2023). In response, competitor Google released Bard, based on the PaLM 2 LLM, to the wider public in May 2023. Unlike ChatGPT, Bard had real-time access to information from the web through Google Search. Comparisons of performance metrics between the two models have found ChatGPT outperforms Bard in terms of accuracy and relevance, but Bard has faster response time (Bhardwaz & Kumar, 2023). Bard was replaced by Gemini in February 2024, with similar functionality, but powered by a new LLM also named Gemini (Hsiao, 2024). While there are many AI writing tools available on the market, the present study focuses on ChatGPT and Gemini as they are the most popular choices for workplace usage (Coles, 2024). There is currently no published research evaluating business writing generated by these two AI tools alongside human writing.

### 2.3.2 Generative AI in business writing

Research on generative AI is concentrated in the academic field, particularly around the applications and implications of ChatGPT in education (e.g., Bhullar et al., 2024; Gill et al., 2024; Okonkwo & Ade-Ibijola, 2021; Rudolph et al., 2023). More research is emerging about AI use in the business context, but this is still a very new and rapidly-changing area. Only two published research articles were found that evaluate AI-generated business emails and these are reviewed below.

AlAfnan et al. (2023) assessed AI-generated texts based on prompts from composition, business writing and communication university courses. 30 texts across the three disciplines were analysed for plagiarism and task accomplishment as evaluated by professors in the respective disciplines using established answer keys and rubrics. The experiment related to producing business writing prompted ChatGPT to generate a 150-word email in reply to an angry customer. The generated emails were found to be formulaic and templated (as evidenced by Turnitin similarity scores) rather than personalised. The rubric on which the evaluation was based was not published, but the business emails were reported to have received average grades, being penalised for aspects such as inappropriate salutation and tone and for lacking in detail. Detail was an important point for the evaluators, as students taking the business writing course were taught to provide details of what went wrong and suggest follow-up mechanisms. Although an interesting initial study, several limitations are apparent. Firstly, a very small number of texts were analysed in each discipline, with only five texts generated for the same prompt in the business writing case. The researchers also did not disclose the wording of the prompt, which affects interpretation of the results as the quality of the prompt has an effect on the text that is generated (Giray, 2023; Lo, 2023; Teubner et al., 2023). Finally, the 150-word limit imposed runs counter to the provision of detail that was valued by the evaluators.

Similar to AlAfnan et al. (2023), Jovic and Mnasri (2024) found that AI-generated emails fell short in providing sufficient supporting evidence, information and detail. They conducted a wider comparative study of the performance of four freely-available LLMs (Bard, Bing Chat, ChatGPT 3.5 and Llama 2) in generating emails that are clear, concise, accurate and contextually relevant. Three email genres were chosen by the researchers: a routine complaint, a negative message (refusal) and a persuasive message. Prompts for each were relatively detailed in the instructions given. For example, the prompt for the refusal message gave specific directions to use an indirect approach in five paragraphs, including an expression of regret and offer of alternative solutions. ChatGPT 3.5 and Bing followed the prescribed format, whereas Llama 2 and Bard did not. The generated texts (presumably there were 12 texts, one for each prompt across the four LLMs) were analysed in terms of content, format, diction and tone and scored on a rubric created by the first author. In the overall analysis of all prompt types, Llama 2 achieved the highest performance score, followed by Bing, ChatGPT 3.5 and Bard. However, the robustness and generalisability of the scoring data are limited as the scoring was conducted by only one rater, the second author. Neither AlAfnan et al. nor Jovic and Mnasri evaluated texts generated by Gemini, as it had not yet been released at the time of their research.

In addition to the small number of texts rated by very few assessors, a major limitation of the studies reviewed is the lack of comparison between AI-generated and human-written emails. It is undoubtedly valuable to study the features of AI-generated texts and make comparisons between different LLMs. However, it is a glaring omission not to compare these to texts written by humans, whose writing the AI tools were trained to emulate. This current study aims to fill this gap.

### 2.3.3 Human and AI authorship

There has been wide interest in evaluating software designed to detect AI-generated writing, particularly in the context of academic plagiarism (e.g., Elkhatat et al., 2023; Walters, 2023). Fewer studies have explored the ability of humans to distinguish between AI and human writing, and at time of writing there were none in the business context. An early study by Gunser et al., (2021) investigated literary texts generated by GPT-2 (an earlier model preceding what now powers ChatGPT) and humans, with the human participants (*N* = 9) asked to evaluate which texts were AI-generated versus human-written. They found that even though the participants had literature-specific professional backgrounds, they experienced difficulty determining authorship, with over 26% of all texts misclassified. Another study found that human participants were incapable of reliably detecting AI-generated poetry, with an average accuracy rate of 50.21%, or no deviance from chance (Köbis & Mossink, 2021).

Moving to academic texts, Casal and Kessler (2023) investigated the extent to which reviewers from Applied Linguistics journals (*N* = 72) could distinguish between AI-generated and human-written research abstracts (eight texts in total). Findings suggested that the reviewers were not particularly effective in identifying AI versus human writing, with an overall positive identification rate of only 38.9%. However, the reviewers were more effective at identifying human authors of research abstracts (44.1% accuracy rate) than AI authors of abstracts (33.7%). To gain insight into the reviewers' rationales for their authorship decisions, semi-structured interviewers were conducted with seven of the participants. The top four rationales provided by seven reviewers in semi-structured interviews concerned (1) continuity and coherence, (2) specificity or vagueness of details, (3) familiarity and voice and (4) writing quality at the sentence-level. According to these rationales, reviewers determined abstracts that were connected and coherent; contained specific information about different aspects of the study; were written in unique style and voice; and used well-constructed phrases or sentences to be human-written. Of these rationales, it was found that writing quality at the sentence-level was the most reliable criterion, leading reviewers to correctly identify authorship 60% of the time. Reviewers who based their decisions on one of the other three rationales achieved accuracy rates of between 20-29%.

Still in the academic context, Yeadon et al.'s (2024) research into authorship decisions on 300 AI-generated and human-written university physics essays by human evaluators (*N* = 5) found an average accuracy rate of 62.4%, which the authors reported as only marginally better than random chance. This was compared with the performance of AI-detection software, with ZeroGPT achieving a 98% success rate in differentiating authorship. The use of AI-detection software is beyond the scope of this current study, as the focus is on human evaluation.

The studies discussed in this section on authorship suffer from the limitations of either small numbers of human evaluators or of texts evaluated. They were also conducted in specific contexts which may not generalise to the business context. As a very large number of texts from the academic realm are available on the internet, it is possible that LLMs have wider access to this type of data to inform their training. It may also be the case that academic texts are more technical and rigid in their use of genre structures and language, making them more easily replicated by AI tools and hence more likely to be judged as human-written. At present, fewer professional texts are likely to be available in the public domain to train LLMs, and the relationship-building focus of business texts could make them more difficult for AI tools to reliably replicate. Finally, the majority of existing studies on authorship do not provide details into what factors influence human assessors when deciding whether a text is AI-generated or human-written. As described above, only Casal and Kessler (2023) collected qualitative data to gain insight into rationales for authorship decisions.

No published studies have yet been conducted on authorship judgements of business writing. The current study is less concerned with accuracy rates, but hopes to understand the reasons behind authorship judgements. This may provide insights into the characteristics that lead humans to believe a business text is either human-written or AI-generated.

### 2.3.4 Perceptions of business communication instructors

Given the developments in generative AI, it is imperative for instructors to consider how these new technologies can change business communication and teach students about the uses, functions, and ethics of AI in the workplace (Getchell et al., 2022). Cardon et al. (2023b) surveyed 692 business practitioners and found that over 80 per cent have used generative AI for work purposes, with about 41 per cent being weekly users. They outlined evolving skill and competency needs, recommending that business communication instructors teach AI-assisted writing with a focus on a writing-as-learning approach. As instructors remain involved in the development of writing skills in the generative AI age, it is important to investigate their perceptions of the new technology.

To this end, Cardon et al. (2023a) conducted the first known and large-scale study of how business communication instructors in higher education view the emergence of AI tools. A survey completed by 343 instructors explored concerns and anxieties about, perceived benefits of and perceived acceptance of AI-assisted writing in the workplace. The survey consisted of closed questions using a 1 to 7 Likert scale and three open-ended questions for instructors to comment on perceived challenges and opportunities of AI-assisted writing. Overall, respondents believed that AI-assisted writing would be widely adopted in the workplace and would require significant changes to instruction. From qualitative analysis of the open-ended items, the researchers recommended a framework for developing AI literacy in students, focusing on the four capabilities in the table below.

**Table 2: AI literacy framework (Cardon et al., 2023a)**

| Capability | Description |
|---|---|
| Application | Understanding how to use AI tools for the task |
| Authenticity | Focusing on genuine communication and the human element: developing a voice, analysing the audience, building trust with sincere and empathetic communication |
| Accountability | Taking responsibility for the accuracy and appropriateness of AI-generated content and using generative AI in a fair and equitable manner |
| Agency | Professionals retaining control to make their own choices |

Cardon et al. (2023a) captured attitudes at an early stage of generative AI, when it is still new to many users, and these attitudes may shift as use of AI tools becomes more widespread. While the sample size was large, all participants were instructors working in higher education teaching undergraduate and graduate students, who are presumably younger and have lower levels of professional experience. Hence, the concerns of university instructors may differ from those teaching adults and practitioners already working in professional contexts. The present study investigates the perceptions of the latter group: trainers working with professionals who are already grappling with the introduction of AI tools in their workplaces.

## 2.4 Research questions

A review of the literature has shown gaps in the research on the business refusal genre and on AI-assisted writing in the business context. Research in these areas is important as business communications are undergoing rapid change with the increasing adoption of AI tools for workplace writing. This study aims to identify the genre structures and characteristics of human-written and AI-generated business refusal texts and investigate how they are evaluated by human assessors. It also explores whether humans can identify authorship of the texts and what influences these decisions. These insights, alongside the perceptions of professional skills teachers and trainers, will be used to inform pedagogical suggestions in the conclusion.

The following research questions (RQs) were posed to focus the study:

1. RQ1: What are the moves and steps that typify the business refusal genre? Are there differences between AI-generated and human-written texts?
2. RQ2: How are AI-generated and human-written texts evaluated by human assessors using an established business writing rubric?
3. RQ3: Can human assessors identify authorship of texts (AI-generated or human-written)? What influences their decisions?
4. RQ4: How do business communication trainers view the use of AI tools for writing in the workplace?

# 3 Method

This study adopted a mixed-method research design (Creswell and Plano Clark, 2018; Dörnyei, 2007), combining quantitative and qualitative approaches to answer the RQs. Texts collected from AI and human writer groups were used in the genre analysis to provide information on the move structures of business refusal texts (RQ1). Participants were recruited to perform an online task to provide quantitative and qualitative data to answer RQ2, RQ3 and RQ4.

## 3.1 Collecting the texts

The first step was to collect texts for both the genre analysis (RQ1) and evaluation by human assessors (RQ2 and RQ3). For this, two business scenarios were developed:

(a) An internal communication, instructing the writer to refuse a colleague's request to deliver a workshop on their behalf; and

(b) An external communication, instructing the writer to refuse a client's request to increase the number of participants on a workshop beyond the maximum class size.

For each scenario, three prompts were created specifying different audiences, making a total of six prompts (Appendix A). There was no word limit for the output and writers were instructed to invent names and details. The same six prompts were presented to three writer groups to generate a corpus of 54 texts, as shown in Table 3 below.

**Table 3: Breakdown of writer groups and texts**

| **ChatGPT 3.5 (OpenAI)** | **Gemini (Google)** | **Human (*n* = 6)** |
|---|---|---|
| 18 texts<br>(3 versions per prompt) | 18 texts<br>(3 versions per prompt) | 18 texts<br>(3 versions per prompt) |

The six prompts were input into ChatGPT and Gemini one by one, with each tool generating three versions of texts for each prompt. For the six human writers, half were assigned to the internal scenario and half to the external scenario. Each writer wrote three emails: one for each prompt specifying a different audience for the same scenario. The six human writers were recruited from the researcher's network of colleagues working as professional skills trainers in Singapore and Spain, with between 7 to 30 years of experience in teaching business communication. In accordance with University of Oxford consent guidelines, the participants were provided with information on the study and gave written consent for their texts to be used for research purposes. Texts from the three writer groups were collected in March 2024. Samples of texts from each writer group and scenario type are provided in Appendix B.

## 3.2 Scoring instrument

A scoring instrument was needed by which human assessors would be able to evaluate the texts in answer to RQ2. Rubrics, which are tools that include criteria for rating important dimensions of performance, have been found to enhance reliable scoring by facilitating valid judgement of performance assessments (Jonsson & Svingby, 2007). In particular, analytic rubrics, in which texts are rated on several aspects of writing ("scales") rather than given a single score as in a holistic rubric, are more useful as assessors can more easily understand and apply criteria in separate scales (Weigle, 2002).

There are few existing rubrics for the assessment of business writing, and none that were suitable for this current study (although Fraser et al., 2005; Jovic & Msnari, 2024 provided inspiration). For the purposes of this study, a three-scale analytic scoring rubric (Appendix C) was created using the British Council's established principles of effective business writing. These are known as the 3Cs (Clarity, Credibility and Connection) and are used in professional skills writing courses globally, having been developed through many years of working with clients from different industries. Four descriptors were formulated under each of the Clarity, Credibility and Connection scales to guide assessors on what to look for. Descriptors were referenced against existing business writing rubrics (Fraser et al., 2005; Jovic & Mnasri, 2024), with the current rubric deliberately kept simple for ease of use and to prevent assessor fatigue. For each text, the assessor was asked to give a score out of 5 on each scale, to produce a total score out of 15. As with the researcher-created rubrics used in previous studies, there is an arbitrary aspect to the criteria and scoring mechanism in the 3Cs rubric. To check concurrent validity, an additional measure was introduced, with participants asked to also rate each text holistically by providing a score between 1 to 10 based on overall impression of the quality of the text. This would then be used to check whether raters were evaluating the texts on a holistic non-specific scale in a similar way as the analytic scale, adding evidence to whether the rubric could be considered a valid measure of the construct compared to alternative (concurrent) methods of scoring.

## 3.3 Participants

Participants for the main data collection phase were recruited through convenience sampling, with teachers and trainers from the British Council global network invited to participate in the study through an email distributed by the teaching leads in each region. Potential participants needed to have experience in teaching professional skills or Business English, including a writing component. Snowball sampling of this niche population (Bryman, 2016) helped reach the target number of participants for this study. Complete data was collected from a total of 36 participants, whose Business English-related teaching experience ranged from 1 to 25 years, with a median of 12.5 years. The majority of participants were located in Singapore (*n* = 21), with the remainder from Bulgaria, France, Hong Kong, India, Jordan, Malaysia, Mexico, Saudi Arabia, Spain, Taiwan, Thailand and the United Kingdom. The table below shows the profile of the study participants.

**Table 4: Profile of study participants**

| | |
|---|---|
| **Continent/region** | Asia: 81%<br>Europe: 11%<br>Middle East: 6%<br>South and Central America: 3% |
| **Gender** | Male: 36%<br>Female: 64% |
| **Years of experience teaching Business English** | 1-4 years: 17%<br>5-9 years: 11%<br>10-14 years: 31%<br>15-19 years: 22%<br>20-24 years: 17%<br>25 or more years: 3% |

## 3.4 Data collection procedure

Data collection was conducted online using JISC Online Surveys (Version 3) to create a data elicitation task (Appendix D). Before launch, the task was piloted with a colleague, a professional skills trainer with 25 years experience, to assess usability (Cohen et al., 2018). Refinements to the wording and sequencing of questions were made based on his feedback. He was then excluded from the main data collection phase and was not part of the final sample of 36 participants. Participants who had consented to participating in the research were emailed a link to the task to complete at a time and location of their choosing. The task comprised three parts:

- Part 1 was designed to elicit how human assessors evaluate texts from the three writer groups. Each participant was asked to score three texts from the total pool of 54, based on the scoring rubric provided (detailed in section 3.2 above). To reduce the influence of the "hawk-dove effect" (McManus et al., 2006), where individual assessors exhibit tendencies towards leniency or severity, one text each from the ChatGPT, Gemini and Human writer groups were presented in a randomised order. The researcher also acted as a third assessor and rated all 54 texts, in order to set a benchmark and reduce noise in the data due to potential differences between raters. While participants were aware that the research topic was human and AI-generated business texts, they were unaware of whether the texts they saw were written by humans or AI tools. In addition to scoring on the 3C rubric, participants were asked to give a score out of 10 to represent their overall impression of the quality of each text. In this way, each of the 54 texts was rated by two participants, in addition to the researcher's rating. Immediately after scoring each text, participants were asked to identify whether the text was written by a human or an AI tool. This was followed by an open-ended question asking participants to explain what influenced their decision.
- Part 2 explored participants' perceptions regarding the use of AI tools for business writing. Participants were asked three open-ended questions and encouraged to write freely about perceived concerns, benefits and challenges relating to the use of AI tools.
- Part 3 collected demographic data (years of teaching experience and work location). In addition, participants were asked whether they had used AI tools to help with writing and to rate their familiarity with using AI tools on a scale of 1 to 5.

## 3.5 Data analysis

Data analysis for this study was carried out in three stages. First, a genre analysis was conducted on texts from the three writer groups. Next, quantitative analyses were performed on scoring data and authorship judgements. Finally, thematic analysis was conducted on qualitative data.

### 3.5.1 Genre analysis

A genre analysis based on the full corpus of 54 texts was conducted based on Swales' (1990, 2004) move-step approach. There is scant guidance in the literature on the process as researchers generally do not describe the specific processes used to identify moves and steps. The method used in this study draws on the work of Upton and Cohen (2009) and Hyon (2018), who offered a more systematic description of the process for identifying and describing moves. Table 5 below describes each step in the move structure analysis process and how it was done in this study.

**Table 5: The move structure analysis process (adapted from Hyon, 2018)**

| Step | Description | How it was done in this study |
|---|---|---|
| 1 | Get an initial sense of the texts, their genre and the genre's purposes | The researcher read all of the texts to confirm alignment with the refusal genre and its purposes |
| 2 | Gather first impressions of the texts' moves and steps | A first pass was made at labelling moves (function) and steps (sub-parts of the moves) |
| 3 | Establish a working set of moves and steps categories | First impressions were refined, with categories combined or split as needed to develop the coding frame |
| 4 | Pilot code the moves and refine the categories as needed | A sample of the texts was coded using the coding frame developed in Step 3 and categories revised as necessary |
| 5 | Code the moves/steps in the full set of texts | The researcher carried out the coding for all texts |
| 6 | Note patterns in move/step frequency and organisation/sequencing | Frequencies were calculated to determine which moves/steps are obligatory and which are optional, as well as patterns in the order in which moves occur |

Following this procedure, a framework for moves and steps was produced. Move occurrence counts for each text were tabulated in a spreadsheet and frequencies calculated. Sequencing patterns were checked and deviations from the main framework analysed.

### 3.5.2 Quantitative data analysis

Parts 1 and 3 of the online task produced numerical data that were analysed using IBM SPSS Statistics (Version 29). The statistical tests in the following table were conducted to help answer RQ2 and RQ3.

**Table 6: Statistical tests for quantitative data analysis**

| RQ | Purpose | Statistical test |
|---|---|---|
| 2 | Establish reliability of rubric | Cronbach's alpha |
| | Establish concurrent validity | Pearson product-moment correlation |
| | Establish normality of data | Frequency statistics<br>Shapiro-Wilk test |
| | Compare scores between writer groups | One-way analysis of variance (ANOVA)<br>Descriptive statistics by writer group<br>Tukey HSD post-hoc test |
| 3 | Compare judged versus actual authorship | Crosstabulations<br>Pearson Chi-Square<br>Fleiss multirater kappa |

### 3.5.3 Qualitative data analysis

Parts 1 and 2 of the online task invited participants to give detailed comments through open-ended questions. The data contributed to answers for RQ3 and RQ4. Responses were analysed using thematic analysis, a qualitative method for identifying recurring themes within the data (Neuendorf, 2018). This approach was chosen to explore participants' perceptions in an inductive manner, allowing themes to emerge from the data itself. Thematic analysis involved an iterative process. First, the open-ended responses were read and re-read to gain a comprehensive understanding of the content. Initial codes were then assigned to capture interesting phrases, ideas, and potential themes. These codes were subsequently refined and grouped into broader thematic categories related to the research questions. Finally, the revised coding framework was systematically applied to all data to ensure consistency. Analysis was done using NVivo (Version 14).

## 3.6 Research ethics

This study received approval from the Central University Research Ethics Committee at the University of Oxford in November 2023 before data collection started (Appendix E). Written informed consent was obtained from participants following the University's template information sheet for online research. All participants were over 18 years old. They were informed that participation was voluntary and that they would not be identifiable from their data. Information was provided on use and storage of their data, as well as the process for reporting concerns. Participants were required to confirm they had read the participant information and gave consent to take part in the research, by selecting "Yes, I agree to take part", before being able to access the task.

# 4 Results

## 4.1 RQ1: What are the moves and steps that typify the business refusal genre? Are there differences between AI-generated and human-written texts?

### 4.1.1 Move and step framework

Following Swales (1990) and Hyon (2018) as described in Method, the 54 texts were analysed to produce a framework containing four major moves and segmentation of their associated steps (Table 7). This builds upon typical moves identified in previous studies to provide a richer description of the business refusal genre. Frequency and sequencing will be discussed in following subsections.

**Table 7: Move and step framework for business refusals with examples**

| **Moves and steps** | **Examples** |
|---|---|
| **MOVE 1: Open the correspondence** | |
| Step 1.1: Salutation* | Dear, Hi, Hello |
| Step 1.2: Polite remark | "I hope this email finds you well"<br>"How was your holiday?"<br>"Nice to hear from you" |
| Step 1.3: Express thanks | "Thank you for your email"<br>"Thanks for reaching out"<br>"I appreciate your prompt response" |
| Step 1.4: Refer to request | "…about your request to increase the participant limit"<br>"…regarding the upcoming Business Writing Skills 1 workshop"<br>"You would like to…" |
| Step 1.5: Empathy statement | "I understand that unexpected engagements can arise"<br>"I know this is important to you"<br>"I understand this may not be the answer you were hoping for" |

| | |
|---|---|
| **MOVE 2: Provide details** | |
| Step 2.1: Explanation, reason* | “My schedule is already full”<br>“Here are some key reasons for this limit”<br>“The trainer cannot give the individual attention each participant needs if the group is too large” |
| Step 2.2: Additional information | *Context-dependent and may include information on personal circumstances (internal texts) or policies and procedures (external texts)* |
| Step 2.3: Explicit refusal | “I regret to inform you that I am unable to…”<br>“We have decided to maintain the maximum class size”<br>“I can’t bail you out this time” |
| **MOVE 3: Call to action** | |
| Step 3.1: Specific alternatives or advice | “Another option would be to…”<br>“I’d like to offer the following alternative solutions”<br>“Have you tried checking with Karen or Julie…” |
| Step 3.2: Next step | “Please let me know which solution suits you best”<br>“Let me know if you would like to have a chat about this”<br>“Shall we schedule a meeting for next Tuesday at 10am” |
| **MOVE 4: Close the correspondence** | |
| Step 4.1: Closing comment | “I trust you will find a suitable arrangement”<br>“We are looking forward to working with you”<br>“Hope you manage to find someone” |
| Step 4.2: Sign-off | Sincerely, Regards, Best regards, Warm wishes, All my best |

*Obligatory step, occurring across all texts

### 4.1.2 Frequency of move and step occurrence

Move occurrence counts were tabulated by step and are shown as percentages in Table 8, separated by writer group and text type. ChatGPT and Gemini texts were largely consistent in performing all steps within the framework, whereas human-written texts were inconsistent and often omitted steps. Only Step 1.1 Salutation and Step 2.1 Explanation were obligatory steps that occurred across all texts.

**Table 8: Frequency of step occurrence by writer group, text type and total corpus (percentage)**

| | ChatGPT (*n* = 18) | | Gemini (*n* = 18) | | Human (*n* = 18) | | Total corpus (*N* = 54) |
|---|---|---|---|---|---|---|---|
| **Step** | **Internal** | **External** | **Internal** | **External** | **Internal** | **External** | |
| **1.1** | 100 | 100 | 100 | 100 | 100 | 100 | **100** |
| **1.2** | 100 | 100 | 0 | 0 | 56 | 33 | **48** |
| **1.3** | 189* | 200* | 133* | 200* | 11 | 56 | **131** |
| **1.4** | 100 | 100 | 100 | 100 | 78 | 78 | **93** |
| **1.5** | 178* | 167* | 122* | 156* | 44 | 11 | **113** |
| **2.1** | 100 | 100 | 100 | 100 | 100 | 100 | **100** |
| **2.2** | 0 | 100 | 100 | 100 | 44 | 100 | **74** |
| **2.3** | 100 | 100 | 100 | 100 | 78 | 33 | **85** |
| **3.1** | 100 | 100 | 100 | 100 | 100 | 78 | **96** |
| **3.2** | 100 | 100 | 100 | 100 | 44 | 78 | **87** |
| **4.1** | 100 | 100 | 100 | 100 | 78 | 78 | **93** |
| **4.2** | 100 | 100 | 100 | 100 | 78 | 100 | **96** |

*Percentage counts over 100 indicate that a move occurred more than once in a text

Analysis of the length of each move and step was beyond the scope of this study. However, on the measure of overall average word count, the AI tools produced longer texts (ChatGPT = 240 words per text; Gemini = 288; Human = 173). Both ChatGPT and Gemini generated more content for the external text types, generally by giving more additional information (Step 2.2) and, in the case of Gemini, multiple detailed alternatives (3.1).

### 4.1.3 Move sequencing

While frequency counts provided insight into how often moves occurred, further analysis was needed to clarify the sequence in which moves and steps are made. In the majority of cases, moves and steps occurred in the order listed in the framework. One notable deviation from the framework is in Move 2 (Provide details), namely how ChatGPT and Gemini texts present the refusal (Step 2.3). Both AI tools almost consistently delivered the refusal as the first step in Move 2, before the explanation (with the sequence being Step 2.3 → 2.1 → 2.2). Only two exceptions (Texts 19 and 29, both Gemini) gave reasons before the refusal, but these appeared to be the anomalies among 36 AI-generated texts. Conversely, all human-written texts that included an explicit refusal (2.3) followed the framework order and delivered the refusal as the final step of Move 2. This direct versus indirect structure will be addressed again later in the results for RQ3.

In addition to omitting steps, there was more variation in how human writers organised their texts. While most human-written texts followed the order listed in the framework, the examples in the table below are noted for deviation from the order.

**Table 9: Variations in move sequencing (human-written texts)**

| Text | Type and audience | Sequence |
|---|---|---|
| 37 | Internal, colleague at same level | Step 1.1 → 1.4 → 2.1 → 2.2 → 2.3 → 1.5 → 3.1<br>(7 of 12 steps occurred) |
| 38 | Internal, colleague at same level | Step 1.1 → 1.2 → 1.4 → 2.1 → 3.1 → 3.2 → 1.5 → 4.1 → 4.2<br>(9 of 12 steps occurred) |
| 54 | External, longstanding customer | Step 1.1 → 1.3 → 1.4 → 2.1 → 2.2 → 3.1 → 4.1 → 3.2 → 4.2<br>(9 of 12 steps occurred) |

Text 37 is notable for its omission of steps and variations in move sequencing, and a full example is presented in the next subsection. For Text 38, the empathy statement (1.5) occurs at the end of Move 3 Call to action, with the comment: "Always difficult when these things happen". Text 54 is written to an external client and is the only external text to deliver the next step (3.2) after the closing comment (4.1), asking: "Shall we schedule a meeting for next Tuesday?" before signing off. This may reflect the writer's desire to keep the conversation open and invite a response in a more explicit way.

### 4.1.4 Examples

Below are examples of the move and step analysis for one text from each writer group, according to the framework presented in Table 7. All three texts were written using the same prompt below, specifying an internal communication with a colleague in the same role:

> *You are a professional skills trainer at a corporate training provider. Your colleague, another experienced trainer, is scheduled to deliver the Business Writing Skills 1 workshop three days from now. They have just realised they have another engagement that day and have asked you to deliver the workshop on their behalf. Write an email refusing the request. You can invent names and details.*

Figure 1 below shows Text 1, which was generated by ChatGPT. The email uses 11 of the 12 steps contained in the framework, omitting only Step 2.2 Additional information. Two steps are repeated: Express thanks (Step 1.3) and the Empathy statement (1.5). Another deviation from the framework is the sequencing of Move 2, in which the refusal statement (2.3) is delivered before the explanation and reason (2.1). This sequencing, with an explicit refusal delivered as the start of Move 2, was seen in all of the ChatGPT-generated texts.

**Figure 1: Move analysis example of Text 1 (ChatGPT)**

| **Text 1 (ChatGPT)** | **Step sequence** |
|---|---|
| Dear [Colleague's Name], | 1.1 Salutation |
| I hope this message finds you well. Thank you for reaching out to me regarding the Business Writing Skills 1 workshop scheduled for three days from now. I appreciate your transparency about your scheduling conflict and understand that unexpected engagements can arise. | 1.2 Polite remark<br>1.3 Express thanks<br>1.4 Refer to request<br>1.5 Empathy statement |
| After careful consideration, I regret to inform you that I am unable to take on the responsibility of delivering the workshop on your behalf. Unfortunately, I already have prior commitments and would not be able to dedicate the time required for the effective delivery of the Business Writing Skills 1 workshop. | 2.3 Explicit refusal<br>2.1 Explanation and reason |
| I understand the importance of the workshop and the value it brings to our participants, and I sincerely apologize for any inconvenience this may cause. If there is any way I can assist you in finding an alternative solution or support you in preparing materials, please let me know. | 1.5 Empathy statement<br>3.1 Specific alternatives or advice<br>3.2 Next step |
| I trust that you will find a suitable arrangement for your conflicting engagement, and I wish you the best in delivering the workshop. | 4.1 Closing comment |
| Thank you for your understanding. | 1.3 Express thanks |
| Best regards, | 4.2 Sign-off |

| |
|---|
| [Your Name] |

Figure 2 shows Text 21, generated by Gemini on the same internal communication prompt. It follows all moves in the framework, omitting only the polite comment (Step 1.2). Similar to the ChatGPT text, it delivers the refusal statement (2.3) at the beginning of Move 2, before giving the explanation and additional information.

**Figure 2: Move analysis example of Text 21 (Gemini)**

| **Text 21 (Gemini)** | **Step sequence** |
|---|---|
| Hi [Colleague's Name], | 1.1 Salutation |
| Thanks for reaching out and letting me know about your scheduling conflict for the Business Writing Skills 1 workshop on [Date]. I understand how frustrating it can be when commitments overlap, and I appreciate you considering me as a potential replacement. | 1.3 Express thanks<br>1.4 Refer to request<br>1.5 Empathy statement |
| However, after careful consideration, I won't be able to deliver the workshop on your behalf. My current schedule is already quite packed, and taking on this additional training would require me to significantly adjust other commitments, which unfortunately isn't feasible right now. | 2.3 Explicit refusal<br>2.1 Explanation and reason |
| I wouldn't want to deliver the workshop without being fully prepared and confident in my ability to provide the participants with the best possible experience. Business writing is a nuanced topic, and it requires time for me to familiarize myself with the specific materials and tailor them to the audience. Given the short notice, I wouldn't be able to do this justice. | 2.2 Additional information |
| Would you like me to help you explore other options? Perhaps we can reach out to other trainers within the company who might be available, or even look into external resources if needed. I'm happy to assist in any way I can to ensure the workshop runs smoothly. | 3.1 Specific alternatives or advice |

| | |
|---|---|
| I apologize for any inconvenience this causes, and I understand this is a tight spot. Please let me know how I can further support you in resolving this issue. | 4.1 Closing comment<br>3.2 Next step |
| Best regards,<br>[Your Name] | 4.2 Sign-off |

Figure 3 shows the move analysis for Text 37, which was human-written. This text deviates from the framework considerably, performing only 7 of the full set of 12 steps. In terms of sequencing, the empathy statement (1.5) at the end of Move 2 rather than as part of Move 1. Notably, Move 4 is omitted, with no closing statement (4.1); it is also the only text in the corpus where there is no sign-off (4.2) before the writer's name. This, and the reduced number of steps applied overall, may signal an informal relationship between colleagues and less need for content when context-specific knowledge is already shared.

**Figure 3: Move analysis example of Text 37 (Human)**

| **Text 37 (Human)** | **Step sequence** |
|---|---|
| Hi Selina | 1.1 Salutation |
| I just saw your email about the BWS workshop this Thurs-Fri. | 1.4 Refer to request |
| I would have covered for you, but I have a workshop scheduled for Mon-Tues next week. This is the pilot run for the new client (XYZ) and I have to work on some contextualisation for the workshop. They are quite particular about what they are looking for and it does require some prep and thinking through. | 2.1 Explanation and reason<br>2.2 Additional information |
| I am really sorry, I can't bail you out this time. I have a lot to prep for the Mon-Tues workshop. But I do hope you find cover. | 2.3 Explicit refusal<br>1.5 Empathy statement |
| Have you tried checking with Kane or Jessie if this could be rescheduled or allocated to another trainer who is available? | 3.1 Specific alternatives or advice |

## 4.2 RQ2: How are AI-generated and human-written texts evaluated by human assessors using an established business writing rubric?

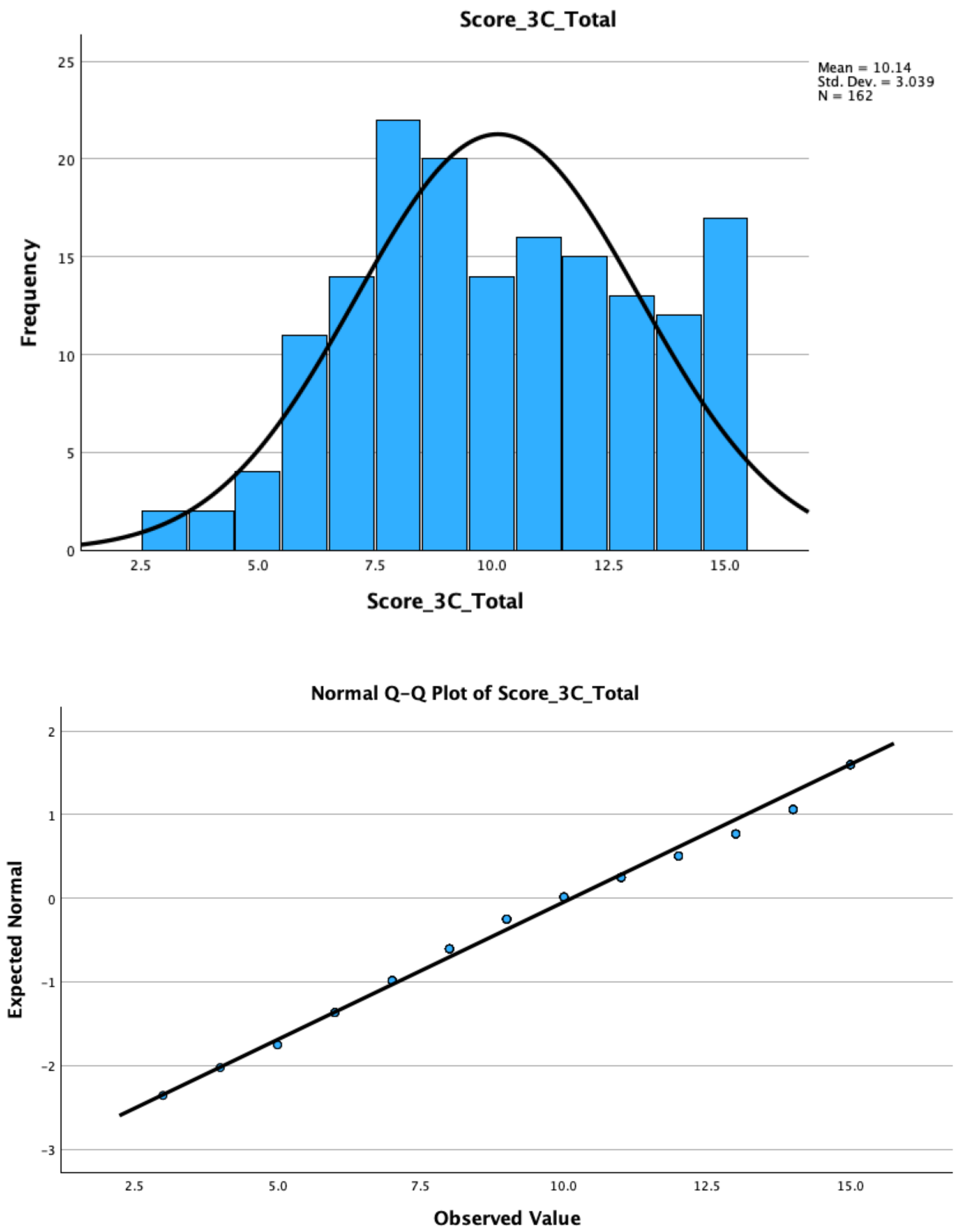
### 4.2.1 Reliability, correlation and normality

As discussed in Section 3.2, a three-scale analytic rubric was created for assessors to evaluate the texts. Reliability statistics produced a Cronbach's alpha of .819, indicating a high level of internal consistency for the scale items (Clarity, Credibility and Connection – hereafter referred to as the 3Cs). This showed that the three items were measuring the same construct of writing quality and further analysis could be done on the total score for the 3Cs.

In addition to scoring texts on the 3C rubric, participants were asked to express their overall impression of the quality of the text as a score out of 10 (1 = poor, 10 = excellent). This holistic measure was intended to strengthen the validity of the rubric by checking that scale items produced scores corresponding with another test of writing quality. A Pearson product-moment correlation was run to determine the relationship between the 3C total score and the overall impression score. There was a strong, positive correlation, which was statistically significant ($r = .811$, $n = 162$, $p < .001$). The high correlation suggests that concurrent validity is demonstrated between the two scoring methods and the 3C score can be used in further analyses as a reliable and valid measure of professional writing.

Next, the 3C score data were subjected to normality tests. Frequency statistics for the total 3C score showed very slight negative skewness (-0.027) and negative kurtosis (-0.856), with both values being within the acceptable range for normal distribution (Roever & Phakiti, 2017). While the numerical Shapiro-Wilk test suggested non-normal patterns ($p < .001$), this could be due to the sensitivity of the test to minor deviations (Mishra et al., 2019). Visual inspection of the Q-Q plot and histograms showed normal distribution of the data (Figure 4). It was therefore decided that parametric tests could be applied to answer the research question.

**Figure 4: Graphical tests of normality - histogram and Q-Q plot for 3C total score**



### 4.2.3 Comparison across writer groups and text types

The 3C scores were submitted to a one-way analysis of variance (ANOVA) across the three writer groups. Table 10 shows there was a statistically significant difference between groups with a medium to large effect size ($F(2,159) = 8.859$, $p < .001$, $\eta^2 = .100$). This indicates that the writer group (ChatGPT, Gemini, Human) has a moderate influence on the variance in total writing scores.

**Table 10: One-way ANOVA on 3C total score**

| | Sum of Squares | *df* | Mean Square | *F* | Sig. |
|---|---|---|---|---|---|
| Between Groups | 149.086 | 2 | 74.543 | 8.859 | <.001 |
| Within Groups | 1337.926 | 159 | 8.415 | | |
| Total | 1487.012 | 161 | | | |

To identify which specific groups differed, descriptive statistics were conducted for the three writer groups (Table 11), along with a Tukey post-hoc test (Table 12). The results showed that participants rated human-written texts ($M = 11.22$, $SD = 2.957$, $p < .001$) significantly higher than ChatGPT-generated texts ($M = 8.89$, $SD = 3.007$). There was also a significant difference in scores between ChatGPT-generated texts and Gemini-generated texts ($p = .034$). However, differences between Gemini-generated ($M = 10.30$, $SD = 2.731$) and human-written texts did not reach significance at the 0.05 level ($p = .224$).

**Table 11: Descriptive statistics by writer group**

| | | | | | 95% Confidence Interval for Mean | |
|---|---|---|---|---|---|---|
| Groups | *N* | Mean | Std. Deviation | Std. Error | Lower Bound | Upper Bound |
| ChatGPT | 54 | 8.89 | 3.007 | .409 | 8.07 | 9.71 |
| Gemini | 54 | 10.30 | 2.731 | .372 | 9.55 | 11.04 |
| Human | 54 | 11.22 | 2.957 | .402 | 10.42 | 12.03 |
| Total | 162 | 10.14 | 3.039 | .239 | 9.66 | 10.61 |

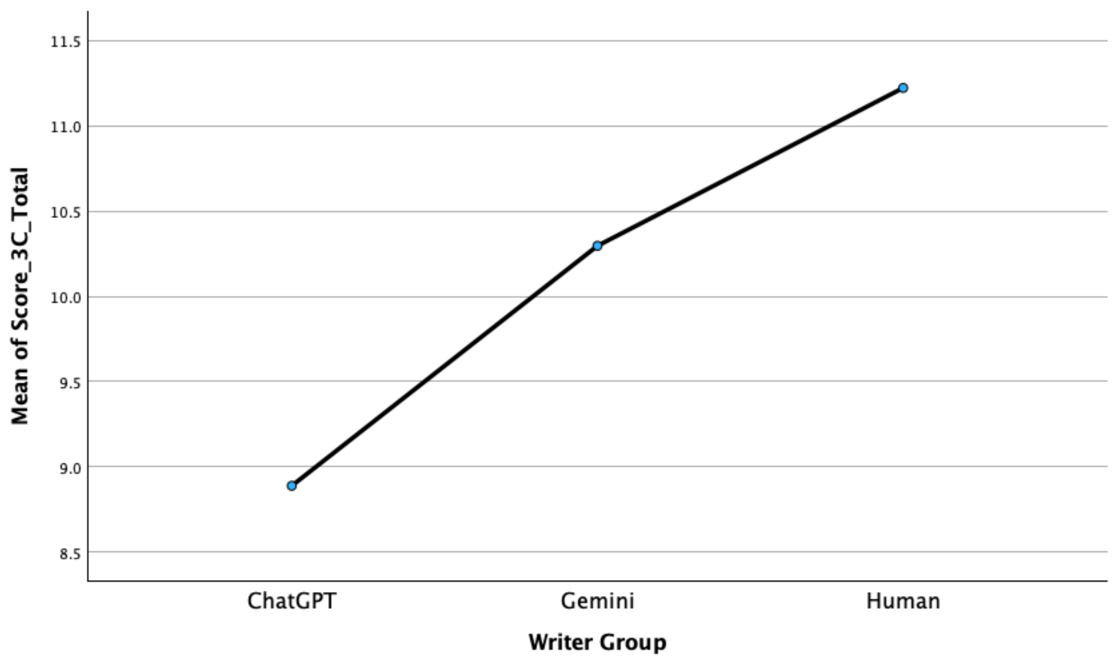
**Table 12: Tukey HSD post-hoc test – multiple comparisons on 3C total score by writer group**

| (I) Writer_Group | (J) Writer_Group | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval Lower Bound | 95% Confidence Interval Upper Bound |
|---|---|---|---|---|---|---|
| ChatGPT | Gemini | -1.407* | .558 | .034 | -2.73 | -.09 |
| | Human | -2.333* | .558 | <.001 | -3.65 | -1.01 |
| Gemini | ChatGPT | 1.407* | .558 | .034 | .09 | 2.73 |
| | Human | -.926 | .558 | .224 | -2.25 | .39 |
| Human | Gemini | .926 | .558 | .224 | -.39 | 2.25 |
| | ChatGPT | 2.333* | .558 | <.001 | 1.01 | 3.65 |

*The mean difference is significant at the 0.05 level

A means plot (Figure 5) confirmed that ChatGPT-generated texts had the lowest mean scores, while human-written texts scored the highest. Among the two AI tools investigated, raters preferred Gemini-generated texts over those generated by ChatGPT. Finally, to check for effects of text type (internal or external) on total 3C scores, a two-way ANOVA was conducted. No statistically significant interactions were found between writer group, text type and total 3C score ($p$ = .448).

**Figure 5: Means plot – mean of 3C total score by writer group**



## 4.3 RQ3: Can human assessors identify authorship of texts (AI-generated or human-written)? What influences their decisions?

### 4.3.1 Quantitative results

After completing scoring based on the rubric, participants were asked to judge the authorship of each text (whether it was written by a human or an AI tool). Table 13 shows crosstabulations for judged versus actual authorship. Overall accuracy of judgements was 74% (80 correct out of a total 108 attempts). When analysing by writer group, participants were most easily able to identify human-written texts, with an 86.1% accuracy rate (only 5 instances out of 36 where a human-written text was incorrectly identified as AI-generated). By contrast, there was more uncertainty around identifying the authorship of AI-generated texts, with a 68.1% accuracy rate. Within the AI writer groups, there was little difference between ChatGPT and Gemini, with accuracy rates of 66.6% and 69.4% respectively.

**Table 13: Crosstabulation - judged versus actual authorship**

| | | | Actual | | |
|---|---|---|---|---|---|
| | | | AI | Human | Total |
| Judgement | AI | Count | 49 | 5 | 54 |
| | | % within Judgement | 90.7% | 9.3% | 100.0% |
| | | % within Actual | 68.1% | 13.9% | 50.0% |
| | | % of Total | 45.4% | 4.6% | 50.0% |
| | Human | Count | 23 | 31 | 54 |
| | | % within Judgement | 42.6% | 57.4% | 100.0% |
| | | % within Actual | 31.9% | 86.1% | 50.0% |
| | | % of Total | 21.3% | 28.7% | 50.0% |
| Total | | Count | 72 | 36 | 108 |
| | | % within Judgement | 66.7% | 33.3% | 100.0% |
| | | % within Actual | 100.0% | 100.0% | 100.0% |
| | | % of Total | 66.7% | 33.3% | 100.0% |

Results of a Pearson Chi-Square test show $\chi(1) = 28.167$, $p$ <.001, suggesting a statistically significant association between judged and actual authorship, with a moderate effect size (Cramer's V = .511). Hence, participant judgements as to whether a text was written by a human or AI tool were not random. To further check that agreement between judged and actual authorship did not occur by chance, Fleiss' kappa was run. This test was chosen over Cohen's kappa as there were non-unique raters for each text (Laerd Statistics, 2019), with 36 participants independently judging three texts out of a possible 54. Per Table 14, Fleiss' kappa showed that there was moderate agreement between judged and actual authorship, $\kappa$=.467 (95% CI, .278 to .655), $p$ <.001.

**Table 14: Fleiss multirater kappa – overall agreement between judged and actual authorship**

| | | Asymptotic | | | Asymptotic 95% Confidence Interval | |
|---|---|---|---|---|---|---|
| | Kappa | Standard Error | z | Sig. | Lower Bound | Upper Bound |
| Overall Agreement[a] | .467 | .096 | 4.850 | <.001 | .278 | .655 |

a. Sample data contains 108 effective subjects and 2 raters.

### 4.3.2 Qualitative insights

After judging the authorship of the text, participants were presented with an open-ended question asking them to explain what influenced their decision as to whether the text was AI or human-written. Thematic analysis of the comments revealed that participants were most concerned with three broad characteristics of the texts: tone, relationship and formality; language choice and grammatical accuracy; content, detail and structure.

**Theme 1: Tone, relationship and formality**

Participant comments frequently referred to "tone" as a characteristic that influenced their assessment of text authorship. Tone can be defined as the writer's attitude toward the reader and subject of the message and can be adjusted according to the context and audience (White, 1988). Despite general consensus on the importance of tone for effective workplace writing, it remains a challenging area as tone perception is subjective and co-created by the reader of the message (Coman & Cardon, 2024). The examples that follow show some of the ways that participants expressed opinions on tone:

> *My decision was mostly influenced by the fact that there was an excellent use of tone throughout the email - it moved from being firm and clear to empathetic and sincere as required.*

> *The text strikes the right balance between personalised and professional.*

*Uses a warmer tone, that sounds more apologetic and concerned about helping out with finding an alternative solution.*

*This text is constructed from so many business writing clichés so it comes across as incredibly mechanistic.*

*It feels overly sincere, with too many words and too many explanations, which makes it feel unnecessary and impersonal rather than sincere.*

*The email is wordy and the tone is mechanical. There are no grammar mistakes and it's coherent but it's lacking warmth and the human touch.*

Tone was commonly described using adjectives, categorised below according to attributed authorship (Table 15).

**Table 15: Descriptors of tone by attributed authorship**

| AI-generated texts | Both AI and human-written | Human-written texts |
|---|---|---|
| clear | blunt | apologetic |
| clinical | casual | concerned |
| cold | mechanical, mechanistic | conversational |
| disingenuous | polite | courteous |
| distant | | empathetic |
| flowery | | familiar |
| formal | | firm |
| impersonal, not personalised | | friendly |
| inappropriate | | genuine |
| not engaging | | human |
| not friendly | | less formal |
| lacks connection | | matter of fact |
| lacks empathy | | natural |
| lacks sincerity | | personal, personalised |
| mixed | | professional |
| overly sincere | | sincere |
| rigid | | supportive |
| | | warm |

In general, adjectives with positive connotations were used to describe texts perceived as human-written. AI-generated texts were generally seen as less empathetic and sincere. Although one human-written text in this study was described as mechanistic, this descriptor was overwhelmingly applied to texts judged to be AI-generated. Emotion was cited as a factor affecting authorship assessment of texts by this participant:

> *The only reason I can't definitively say it's a human is because of the consistent lack of emotion throughout. The text has somewhat of a monotone, which is typical of AI-generated text.*

Participants also showed a tendency to prioritise interpersonal relationships, commenting on the (in)appropriateness of tone and formality depending on the reader-writer relationship. For example, one participant noted that one of the texts was "completely inappropriate as a response to a boss" and another was described as "too formal for people who have been working together for some time". Further elements that contribute to tone are considered in the following themes.

**Theme 2: Language choice and grammatical accuracy**

Words and phrases found in the texts heavily influenced participants' decisions on authorship. Texts identified as AI-generated were frequently described as containing "jargon", "cliché phrases", "business writing clichés", "stock phrases", "corporate speak" and "platitudes". Language used in AI texts was also seen as "formal", "inappropriate", "too standardised", "scripted", "outdated", "mechanical" or "strange". Below are examples of such language, specifically cited by participants (in order of frequency of occurrence in the AI-generated texts):

- reaching out (19 occurrences)
- more than happy/willing (14)
- apologize for any inconvenience caused (13)
- regret to inform (13)
- don't hesitate (7)
- at your earliest convenience (1)
- fostering a culture of continuous learning (1)
- I truly appreciate (1)
- my reason is twofold (1)

- we look forward to continuing our successful collaboration (1)

In addition to finding phrases “typical of AI”, participants also judged texts as AI-generated on account of their complexity, including long-winded sentences and unnecessary or exaggerated use of vocabulary. By contrast, texts judged to be human-written were described as using simpler, clearer and more accessible language (“Plain English”, “less jargony”). Overuse of the first-person singular pronoun “I” was identified as a feature of AI-generated texts, while human-written texts exhibited features such as subject pronoun drop (“Hope you are well”) showing informality and connection, as well as being noted for use of inclusive language (“We could explore options”). Rigorous statistical analysis of linguistic features was outside the scope of the present study, but a simple word frequency count showed that occurrences of the word “I” were far higher in the AI-generated texts (ChatGPT = 145 instances; Gemini = 126; Human = 86).

Texts assessed as human-written also stood out for the use of friendly and colloquial language, such as:

> *The style is very chatty and informal with phrases like ‘bail you out’.*
>
> *The phrase 'I wouldn't be able to do it justice' is too lively and idiosyncratic to be AI.*
>
> *It's also written using less formal language, it reads as though that is how the writer speaks in real life.*

Related to this are remarks on the imperfections that characterise human writing, such as mistakes in grammar and punctuation or not writing out words fully (e.g., using “Thurs” instead of Thursday). Such errors, or lack of errors, guided participants to make their authorship assessments as expressed below:

> *Its flaws made it seem more human to me.*
>
> *It sounds a little too perfect to me for it to be written by a human.*

**Theme 3: Content, detail and structure**

AI-generated texts were judged to be long and heavy on content. Participants commented that AI texts were “verbose”, “long-winded” and contained “too many words and too many explanations”. Content tended to be repetitive, sometimes not making sense, and needed editing for clarity and conciseness. These comments are consistent with the average word count per text for each writer group (ChatGPT = 240; Gemini = 288; Human = 173). Human-written texts were described as “succinct”, “concise” and “to the point”.

A lack of contextualisation contributed to observations that AI-generated texts were templated and generic, as expressed in the comments below:

> *The fact that this is a template suggests to me that it is AI.*
>
> *The early opening section uses standard and logical text that comes across as ‘template style’.*
>
> *The closing is very generic and impersonal, which makes it sound insincere.*
>
> *Sounds a bit generic and formulaic and potentially OTT in details. It's almost too consistent.*
>
> *Too many of the reasonings and offers to be helpful sound pre-scripted. This is something I would expect an AI engine to draft for me.*
>
> *Repetition (eg “regretfully” and “unfortunately” in the same paragraph) suggests a cut-and-paste approach.*

Personalisation was seen as a key factor in achieving tone and connection, as described by one participant:

> *Opening, closing and details are personalised and not generic. This makes the piece of writing sound friendly and sincere, creating connection with the reader.*

In terms of structure, AI-generated texts were widely judged to be clearly organised and coherent. This is likely related to the observation that AI texts follow a template, with one participant commenting:

> *AI is usually good at presenting information in a formulaic and structured way.*

This corresponds with the findings for RQ1, in which AI texts applied moves and steps with the same frequency and sequencing. Some participants remarked that AI texts tended towards being “too balanced”, resulting in overly long responses lacking connection with the reader. The use of bullet points was also mentioned by several participants, who correctly associated them with AI texts (Gemini tends to use bullet points). Where participants judged a text to be human-written, there were comments about a “lack of structure”, particularly in relation to texts written for an internal audience. This could also be due to the relative informality of internal texts, which tend to be shorter and more direct in style. One participant noted:

> *The issues with organisation and clarity indicates the lack of planning that often comes with human writing.*

Several participants mentioned message organisation, specifically the use of direct or indirect structures discussed in the literature review and the genre analysis for RQ1. Participants negatively evaluated the tendency of AI tools to use the direct structure. For example:

> *Firstly it started with a refusal, rather than giving an explanation and then the refusal. This would create a negative impact on how the reader was feeling when reading through the explanation. They would be less likely to take on board the issues raised, or would do so in a negative frame of mind.*

There was also observation of the politeness strategy in implied rather than explicit refusals, as noted by a participant:

> *The text omits the direction to refuse the request. This seems like a human decision made based on cultural norms, specifically a desire to avoid saying 'no' directly.*

Overall, the three broad themes that emerged from the qualitative data show a preference for texts that consider the relationship between writer and reader. While AI-generated texts were notable for their grammatical correctness, they were also assessed to be formulaic and lacking in contextual awareness. By contrast, less clichéd language choice, concise messaging, specific rather than general content, and attention to interpersonal relationships were cited as the main influences when identifying texts as human-written.

### 4.3.3 Summary of text characteristics

The table below summarises the characteristics of AI-generated and human-written texts as found through the move structure analysis and qualitative data from participants.

**Table 16: Characteristics of AI-generated and human-written texts**

| | AI-generated | Human-written |
|---|---|---|
| **Genre structure** | Rigid and formulaic<br>Some steps repeated<br>Explicit refusals before reason | Flexible (steps omitted or different sequencing depending on context)<br>Refusal after reason (or implied) |
| **Tone, relationship, formality** | Mechanical and impersonal<br>Very formal regardless of writer-reader relationship | Human, friendly, sincere<br>Formality level adjusted for audience |
| **Language choice** | Jargon and clichés<br>Overuse of personal pronoun "I" | Plain style<br>Subject pronoun drop<br>More use of "we" |
| **Grammatical accuracy** | No errors | Minor mistakes<br>Short forms used |
| **Content and detail** | Long and repetitive<br>Generic content | Concise<br>Specific details for writer/reader's situation |

## 4.4 RQ4: How do business communication trainers view the use of AI tools for writing in the workplace?

RQ4 aimed to address trainer perceptions of AI tool usage in the workplace, given that developments in generative AI have already begun to have effects on the work of business communication training. This study asked three open-ended questions:

1. What are your concerns about using AI tools in business writing?
2. What do you think are the benefits of using AI tools in business writing?
3. What challenges does the emergence of AI tools present for the training of business writing skills?

Questions 1 and 2 are adapted from Cardon et al.'s (2023a) exploratory study, with Question 3 added to try to elicit context-specific challenges. Questions were compulsory and were responded to by all participants. Percentages shown in Tables 17-19 below are based on the number of participants who mentioned the theme (total *N* = 36).

### 4.4.1 Concerns

Data analysis revealed six broad themes describing participants' concerns about using AI tools in business writing, as shown in Table 17 below.

**Table 17: Thematic analysis of concerns**

| **Theme** | **Count (%)** | **Example comments** |
|---|---|---|
| Loss of the human element | 26 (72) | "Getting an AI to do your writing can also mean that you do not have the opportunity to show your personality in your correspondence and create a connection with the recipient"<br>"[AI tools] seem to make written communication less of a connection between people"<br>"I think for more nuanced, complex issues, human interpretation is required"<br>"AI does not understand human complexities and emotions" |
| Loss of writing and communication skills | 11 (31) | "I worry that people won't develop their writing skills and as a result their ability to communicate effectively will diminish"<br>"Overreliance on AI to write business documents could limit development of writing skills in general" |
| Loss of uniqueness and context-specificity | 9 (25) | "If everyone uses the same tools they may produce identical or similar content which diminishes individuality"<br>"Inability to tailor a message specifically for a target audience" |
| Concerns about accuracy and credibility | 6 (17) | "Could destroy your reputation without careful editing"<br>"Might contain hallucinations or factual untruths" |

| | | |
|---|---|---|
| Concerns about ethics, plagiarism, data security | 6 (17) | “AI is only as good as its creators and may replicate unconscious biases”<br><br>“Potentially employees inputting confidential, copyrighted or sensitive workplace information into generative AI” |
| Loss of jobs | 4 (11) | “There's also the loss of jobs. Lots of people spend a good chunk of their working hours writing. If that’s cut, many of those jobs can be cut - and there doesn’t appear to be new jobs springing up in the wake of AI”<br><br>“Another concern is the easy dismissal of teachers who help professionals with their business writing” |

### 4.4.2 Benefits

Five themes emerged in the perceptions of the benefits of AI tools in business writing, presented in Table 18 below.

**Table 18: Thematic analysis of benefits**

| Theme | Count (%) | Example comments |
|---|---|---|
| Time savings and efficiency | 22 (61) | “Streamlining tasks and making it quicker to write”<br><br>“They obviously save time and cost if used correctly”<br><br>“Speed and convenience when the message isn't too specific or important” |
| Idea and content generation | 21 (58) | “Can support creativity by generating/ brainstorming ideas”<br><br>“It's an excellent way to get ideas and see models before starting to write”<br><br>“AI provides ideas to get started which can then be refined by humans” |

| | | |
|---|---|---|
| Helps with grammar and language | 10 (28) | "It can help rephrase sentences that don't sound quite right"<br><br>"AI has better grammar and spelling than many humans!" |
| Helps organise and structure ideas | 4 (11) | "AI generated documents are also useful to plan the structure and flow of a business document"<br><br>"These could support with organising written information/ layout for writing longer documents" |
| Standardises communication | 4 (11) | "Standardisation of content and tone"<br><br>"AI tools might be good for generating 'template' emails" |

In addition to the above, one participant stated that there were no benefits, calling AI "extremely damaging". A further two participants mentioned that AI tools reduce cognitive load and allow people to focus their energy on other things.

### 4.4.3 Challenges

Finally, data were analysed in terms of challenges associated with applications of AI for teaching professional writing. Some participants responded about challenges in general, while others more closely related AI tools to business writing training. Table 19 identifies four frequently mentioned themes in the comments.

**Table 19: Thematic analysis of challenges**

| **Theme** | **Count (%)** | **Example comments** |
|---|---|---|
| Difficulties for trainers | 18 (50) | *There was variation within this code, with comments about trainer lack of knowledge and confidence; needing to keep up-to-date with new developments; managing diverse attitudes of learners towards AI tools; differences in organisational contexts and guidance on usage* |
| Justifying the need for trainers and training | 16 (44) | "Business writing skills is being increasingly seen as an obsolete skill"<br><br>"Some organisations may feel that there is no need to train staff in business writing, if they can just rely on AI"<br><br>"Participants may not see the necessity of learning these skills" |

| | | |
|---|---|---|
| Overreliance on AI | 12 (33) | “The ease of use and availability of AI tools have meant that participants use it often, even when they shouldn't”<br><br>“Relying on AI means not relying on self to come up with content”<br><br>“It's easy to use and trainees are tempted to resort to it” |
| Integrating AI into training content | 7 (19) | “I think the challenge is how best to integrate AI in training so that it completes current human training models”<br><br>“To make AI fit in the contents of workshops and re-think most of the activities to include it”<br><br>“Training in writing skills should include AI and look at how to appropriately leverage AI” |

Five responses were coded as “other” and included comments on flaws in AI tools and costs involved, with one response stating there are no challenges, only opportunities.

# 5 Discussion

Through the investigation of four research questions, this study used the business refusal genre as a starting point to investigate differences between AI-generated and human-written texts. In this section, results for each RQ are discussed and situated within the existing literature.

## 5.1 Genre move structure

RQ1 contributed a genre framework for business refusals that did not exist in previous literature. The move structure analysis showed there are typical moves and steps employed across writer groups for business refusals. The salutation (Step 1.1) and explanation (2.1) were obligatory steps, used across all texts and writer groups. Texts generated by ChatGPT and Gemini applied moves and steps in a largely consistent manner across text type (internal, external) and did not adjust their approaches, despite the prompts specifying different writer-reader relationships. Only two notable differences between the tools were observed: Gemini did not include polite opening remarks (1.2) for either text type, and ChatGPT omitted additional information (2.2) for internal communications. Some steps appeared more than once in the same text. There was a tendency for the AI tools to express thanks (1.3) twice: once in the opening move to thank the sender for their email and again in the closing, generally a generic comment such as “Thank you for your understanding”. Additionally, for the majority of AI-generated texts, the empathy statement (1.5) occurred as part of Move 1 and then again at the end of Move 2. Steps occurring on two different occasions within a single text (resulting in percentages of over 100 in Table 2) were seen only in AI-generated texts and likely contributed to their perceived repetitive nature. The consistency with which AI texts applied genre moves corresponds with the findings of AlAfnan et al. (2023) and Jovic and Mnasri (2024), who described AI-generated emails as formulaic and templated.

On the other hand, the human-written texts varied in the moves they employed, with differences seen between internal and external texts. Human writers were much more likely to omit steps, depending on the context and writer-reader relationship specified in the prompt. For example, only one of nine (11%) human-written internal texts to colleagues included an expression of thanks (Step 1.3), with this being the more informal comment: “Thanks for thinking of me”. Particularly for external texts written to clients, it was expected that overt signals of politeness would be more consistently included; however, only 33% and 11% of human-written texts included a polite remark (1.2) and empathy statement (1.5) respectively. Instead, politeness may be inferred from the fact that only one-third of external texts included an explicit refusal statement (2.3). The remaining majority contained an implied refusal, which the reader would infer from the performance of other steps. For example, the writer would first offer an explanation to the customer before providing specific alternatives (in line with politeness strategies proposed by Campbell, 1990 and Locker, 1999) and attempting to persuade the customer to accept one of these. Implied refusals can be interpreted as an indicator of human writing skill and a desire to maintain face (Brown & Levinson, 1987) not seen in the AI-generated texts, all of which delivered an explicit refusal statement, such as "I'm unable to approve your request". Overall, human-written texts were more nuanced in their adoption of genre conventions and appeared to consider the text and audience type when choosing whether to include or omit moves, and how to sequence them.

## 5.2 Scoring

RQ2 was answered by collecting quantitative data on how the 54 texts were scored using a three-scale rubric assessing Clarity, Credibility and Connection (3Cs). While further reliability and validity testing should be done before broader use, the 3Cs rubric and its descriptors were shown to be a viable instrument for assessing business writing. Based on the rubric, assessors showed a preference for human-written over AI-generated texts, with Human texts achieving the highest mean score, followed by Gemini and finally ChatGPT. The significant difference in scores between Gemini and ChatGPT texts suggests that Gemini may be the more effective AI tool for business writers. Interestingly, score differences between Gemini-generated and human-written texts did not reach significance. This may signal that Human and Gemini texts might be perceived similarly, but further research with a larger sample size would be needed to explore this aspect. Given the recent release of these AI tools, there were no previous studies comparing the applications of ChatGPT and Gemini in business writing, nor were there studies comparing human-written with AI-generated business texts.

Another area of interest was whether scores were dependent on the scenario type. No statistically significant interactions were found, suggesting that perceived quality based on writer group was not dependent on whether texts were written for an internal or external audience. This is not surprising for the AI-generated texts, as the move analysis showed that ChatGPT and Gemini applied an almost identical approach to structuring all texts, and would therefore be evaluated in a similar manner. Score differences between internal and external texts might have been expected for human-written texts, as move structures varied greatly for the two text types. However, differences did not reach significance in this study. Overall, human-written texts were judged to be of higher quality regardless of whether they were written for an internal or an external audience. It is difficult at present to situate these results within the literature, given that no previous studies have compared scoring of AI-generated and human-written business texts. As summarised in the literature review, AlAfnan et al. (2023) and Jovic and Mnasri (2024) found that assessors gave low scores to AI-generated business texts, but human-written texts were not part of either study. Complicating the comparability and generalisability of results between studies is that each study has used a different scoring instrument incorporating different criteria for assessment.

Studies in academic domains have compared AI-generated and human-written texts. Yeadon et al.'s (2024) study observed no statistically significant differences in scores between physics essays authored by human students and those generated by ChatGPT ($p = 0.107$, $\alpha = 0.05$). However, Herbold et al. (2023) found that ChatGPT-generated argumentative essays were rated higher in quality than human-written essays on seven categories of essay assessment covering content and language items. A complicating factor of Herbold et al.'s study is that although the human essays were written in English by German high school students, neither the students nor the raters were native speakers of English. Overall, the existing literature comparing AI-generated and human-written texts is extremely limited and further studies need to be done in both academic and business domains. Although initial studies in the academic realm show that AI models may be as good or better than humans at writing scientific and argumentative essays, the current study shows they fall short in business writing. One explanation may be that business writing is heavily judged on the ability to build relationships with the reader, an area in which AI-generated texts do not perform as well compared to human-written ones.

## 5.3 Authorship of texts and factors influencing decisions

RQ3 asked whether humans could identify the authorship of texts and what influenced their decisions. Results showed that participants could correctly identify human or AI authorship of business refusal texts with a higher accuracy rate than by chance. These results diverge from the findings of Casal and Kessler (2023) and Yeadon et al. (2024), who found low accuracy rates in human raters' ability to identify authorship. This may be attributable to the context in which these two previous studies were conducted: AI tools generate academic texts that can fool even experienced assessors, but fare less well in a business writing context which prioritises relationships and connection to the reader. Similar to the present study, Casal and Kessler's participants were more likely to correctly identify human-written texts as human. This suggests there are characteristics of human writing that make it readily identifiable as such and more positively perceived. These characteristics were probed with open-ended questioning on how authorship decisions were made.

Qualitative data for RQ3 revealed four main themes that influenced decisions on authorship: tone, relationship and formality; language choice and grammatical accuracy; content, detail and structure. Findings on tone corresponded with research on AI-generated messages conducted by Coman and Cardon (2024), in which professionals rated ChatGPT-generated messages as less sincere and caring. Similar to participant comments in this study, Jovic and Mnasri (2024) found that the tone in LLM-generated emails was emotionally neutral and overly formal, pointing to difficulty in expressing sentiments. In terms of language choice, Campbell et al. (2021) found that a plain style avoiding jargon and non-requisite words contributed positively to the perceived professionalism of business writers, and these perceptions may have influenced authorship judgements in this study. The tendency for AI-generated texts to be convoluted, verbose and jargonistic was an indicator of AI authorship for many participants. Human writers paid attention to the writer-reader relationship and wrote messages that were concise and specific in a plain style, avoiding the use of clichéd language. Several participants also commented on pronoun use, with AI-generated texts tending to overuse the first-person singular pronoun "I", corroborating AlAfnan and MohdZuki's (2023) study. While it was noted earlier in the literature review that you-attitude (Shelby & Reinsch, 1995) is not simply about replacing "I" with "you", the much lower occurrence of "I" in the human-written texts indicates a more reader-focused and inclusive approach based on this rough metric.

Grammatical accuracy was identified as a feature of AI-generated texts. However, despite observed inaccuracies in human-written texts, they achieved higher scores overall in this study (although it is not possible to parcel out the contribution of spelling and grammar). The existing literature continues to debate whether grammatical errors negatively affect perceptions of professionalism (Gubala et al., 2020 say yes; Campbell et al., 2021 say no). Further research could consider if such flaws characterise a humanness in writing that is in fact positively perceived. Regarding content, detail and structure, AI-generated texts were seen as formulaic and long-winded, corroborating what was found in RQ1 on the tendency for AI texts to follow all genre moves and steps in the same sequence (and sometimes repeat steps). In addition, participants noticed a lack of specific details in AI-generated texts, as they often did not contain dates, names, information on prior commitments or realistic alternative solutions. This was despite the fact that all writing prompts included the direction to invent names and details. Ma et al. (2023) and Amirjalili et al. (2024) also found that AI-generated texts, in scientific and academic contexts respectively, contain mainly generalities and are deficient in specificity and depth. In the present study, it was noted that human-written texts demonstrated context-specific knowledge and were tailored to the reader's needs. They contained specific names, reference to prior activities, credible reasons for decisions, benefits of suggested solutions and personal touches (e.g., referring to the client's holiday). Participants also commented that instead of a "harsh refusal", human-written texts offered "a choice of options with benefits". These comments signal a preference for the indirect structure recommended in previous literature (Creelman, 2012; Shelby & Reinsch, 1995; Jansen & Janssen, 2010a, 2010b, 2013). In summary, human assessors preferred an indirect refusal featuring concise, personalised writing in a plain style and sincere tone, containing details that show consideration for the reader's specific situation.

## 5.4 Views on the use of AI tools for writing in the workplace

Finally, RQ4 revealed how business communication trainers perceive the use of AI tools for writing in the workplace: their concerns, the benefits of the tools and the challenges they face. By far the biggest concern was the human element. The majority of participants mentioned this theme and showed a preoccupation with personal relationships, empathy and connection. Cardon et al.'s (2023a) findings of concerns around ethics, loss of skills and loss of jobs were also echoed in this study. In terms of benefits, 61% of respondents mentioned efficiency as a benefit, which is well above the 22% in Cardon et al.'s (2023a) study. However, mentions of the benefit of using AI tools to save time and generate content often came with the caveat that humans would then need to refine the AI output. A majority of participants see AI tools as capable of assisting in the writing process, with specific roles at stages such as brainstorming and organising ideas, as well as editing for grammar and language. This openness suggests that changes to business communication curricula to accommodate AI-assisted writing would not be strongly opposed. The question on challenges invited a variety of answers, including justifying the continued need for business writing training to sceptical learners and organisations who are becoming reliant on AI tools, as well the need to adapt the training curriculum in the face of increased AI usage. These challenges, along with the concerns and benefits raised by the participants, will be revisited when considering implications for teaching in the final section.

# 6 Conclusion

## 6.1 Implications for teaching

The title of this study included a provocative question for the profession: “Is there still a case for business writing training?” The short answer is yes. Business writing training and the teachers and trainers who deliver it are still needed in the age of AI. Effective business writing, especially in difficult situations, requires a human touch, understanding of the reader’s needs and of the context itself. However, there will undoubtedly be changes to how professional writing is taught. Below are five key areas of focus for business writing in the AI age.

Teachers should:

- **Raise awareness of the complex nature of genres**. The broad and simplified genre framework presented here for business refusals is useful as a teaching tool for learners new to the business discourse community, as well as professionals looking to refine their skills. However, it should be seen as a flexible framework and understood as description rather than prescription (Pinto dos Santos, 2002) and process over product (Creelman, 2012). Learners still need to be guided on how they can use and adapt generic conventions to best achieve their goals. Rather than simply providing a writing “formula”, teachers need to emphasise that genres are flexible. Each writer-reader interaction should be handled in a context-specific manner, taking into account factors such as role relationships, tasks, industry-specific practices or concerns (Zhang, 2013). Analysis of AI-generated texts in this study showed they were structured but also formulaic, largely applying the same move and step patterns for every message. Contextualisation remains important and learners should be introduced to a genre’s conventions through analysis, critique, reflection, practice and feedback.

- **Focus on involvement in the writing process, with emphasis on discussion, practice and performance.** AI tools cannot replace the experience of attending training workshops and interacting with other learners and a trainer. The discussion process is hugely important for helping writers develop knowledge in how to persuasively structure arguments, build relationships and embrace nuance (Ying, 2020). Business writing training should continue to focus on giving learners the opportunity to share authentic workplace experiences and engage in collaborative planning, writing and editing. Practice is essential, as each performance of a genre is unique (Devitt, 2009) and involves consideration of how to perform rhetorical moves in ways to fulfil specific tasks in a specific discourse community. Learners then receive feedback on the textual performance from peers and their trainer, reflect on it before performing again. They might also compare their human-written texts with texts generated by different AI tools, with discussion focusing on where AI can be used effectively to help and where human judgement remains most important. Fostering a sense of involvement will help retain text ownership and satisfaction in the writing process (Dhillon et al., 2024), as learners also hone their thinking skills.

- **Consider where AI tools can play a role and re-evaluate business communication curriculum needs**. Curricula should focus on leveraging the strengths of AI tools while compensating for its weaknesses. For example, the ability of AI tools to produce grammatically correct text reduces the need for intense grammar instruction in business writing courses, although learners should still be made aware of common errors. Instead, emphasis should be on using human judgement, creativity, empathy and contextual knowledge to evaluate AI output to ensure a text achieves its communicative purpose. AI tools can be used to support parts of the writing process, such as in finding relevant arguments, suggesting organisational structure and providing examples of styles that may be best suited to the writer's audience and goals (DeJeu, 2024).

- **Develop AI literacy.** As part of the curriculum redesign, guidance on AI literacy should be provided to all learners (Cardon et al., 2023a and Getchell et al., 2022 provide starting frameworks). This should include aspects such as understanding how to use AI tools for specific tasks and for what tasks they might be appropriate; considering authenticity and the human element in communication; taking responsibility for AI-generated content and using it fairly; and discussing issues around governance, ethics and safety. Learners must be able to negotiate what AI-human collaboration should look like. A discussion around the benefits, limitations and rules around use would be a good starting point in most courses. Analysis of the concerns and challenges mentioned in RQ4 also shows a need to upskill teachers and trainers in order to be able to deliver a new curriculum that incorporates AI.

- **Help learners develop aspects of human writing voice that feel human**. Given that AI tools can produce grammatically correct writing, it becomes even more important to focus on what dimensions humans can contribute beyond the efficiency of machines (McKnight, 2021). The results of this exploratory study suggest that having a relationship-orientation, emotional understanding and contextual knowledge are main differentiating factors between human-written and AI-generated texts. Hence, business writing training should focus on developing the ability to analyse the purpose, audience and context so that communication can be adapted to the specific situation at hand. Genuine empathy for the reader's needs and expectations, along with a human voice, personality and style (including non-standard and idiosyncratic language), is what sets human writing apart. It was also found in this study that the ability to write an implied refusal rather than an explicit one is a human skill not replicated by the AI tools. As the bulk of business writing converges on formulaic structures and clichéd phrases, which AI tools perpetuate, human writers should aim to engage and surprise their readers.

## 6.2 Limitations and recommendations for future research

Research on AI-generated writing is still in its early stages. This study is the first known investigation of both human-written and AI-generated business refusal texts by both humans and AI tools. It is hence exploratory and limited, with ample opportunities for continued study. In this final section, several limitations of this study will be discussed along with recommendations for future research.

### 6.2.1 Genre analysis is "fuzzy-edged" and multi-layered

Genre analysis is not an exact science. As moves and their boundaries may be "fuzzy-edged" (Paltridge, 1994) and difficult to delineate (Samraj, 2014), some researchers have recommended checking with a specialised informant acting as a second coder, to agree on classification with a sufficiently high level of agreement (Dudley-Evans, 1994; Hyland, 2015; Moreno & Swales, 2018). Second coding was considered for this study but could not be conducted because a specialist informant was not available within the required time period.

Secondly, this study was limited to a move and step analysis. It established that ChatGPT, Gemini and human-written refusal texts all broadly followed the same structure and included typical steps. However, further linguistic analysis, including with computerised corpus techniques (Flowerdew, 1998, 2005, 2012; Upton & Connor, 2001; Upton & Cohen, 2009), could elaborate on differences in lexicogrammatical features between texts. This would add a quantitative dimension to the perceptions around language choice reported by participants in RQ3. Future genre research would benefit from a combination of move and step analysis with corpus linguistic techniques.

### 6.2.2 Generative AI is constantly evolving and prompt strategies should be considered

The ChatGPT and Gemini texts were generated in March 2024. AI technology is constantly evolving and as LLMs are trained on updated data, output generated from using the same prompt may change. For consistency in this study, each text was generated with a single prompt. Prompts were deliberately kept simple and open to allow each writer group to produce what was natural to them. However, the prompt process is usually iterative in general usage of AI tools, meaning that often more than one prompt is used to fine-tune the desired output. Changes to the prompts (and including more than one) may change the output and the findings of the study.

Future studies on AI and writing in the business context should consider (1) using newer versions of the platforms; (2) evaluating AI models beyond the two studied here; and (3) testing different prompt strategies (Knoth et al., 2024). These steps are necessary to assess whether AI-generated texts are becoming more human-like and positively perceived over time. Continued comparative evaluation would also help users to assess which tool is most suitable for their intended writing task.

### 6.2.3 A wider range of participants and human-written texts is needed

In addition to evaluating texts generated by different AI models, future studies could consider broadening the representativeness of the participant sample. All participants in this study were teachers and trainers with a professional background in Business English. Resulting data hence reflects the opinions and perceptions of this particular group, but does not capture the views of learners or business practitioners. Future research should expand the diversity of the sample, such as including L2 English users to introduce a Business English as a Lingua Franca (BELF) perspective. This might also allow exploration of other pragmatic perspectives, such as the effectiveness of AI-generated texts for intercultural communication. Additionally, in this study, data on participants' self-reported familiarity with AI tools were collected, but this was ultimately not included in the analysis. There is scope in future studies for exploration into how reactions to texts and judgements of authorship may be influenced by the reader's familiarity with AI tools.

Finally, the contributors of the human-written texts in this study are specialists in professional skills training. They have knowledge of effective business writing practices and may not be representative of writers in all industries. Texts were written to a specific scenario for the purpose of this study and it is not clear whether writers would produce different texts under more authentic conditions.

### 6.2.4 Possibilities for hybrid AI-human collaboration

This study investigated AI-generated and human-written texts side by side, in isolation of one another. However, actual use of the technology is more likely to involve a combination of AI technology and human judgement. Further study on how professionals are using AI tools in their writing would be helpful for understanding points in the writing process where technology can assist. This might be achieved through experimental or ethnographic research techniques to track when and where humans are choosing to use or not use AI, as well as decisions made on modifying AI-generated texts to suit a particular communicative purpose. Insights gained would help foster a clearer understanding of best practices in integrating AI tools in business writing. This may include guidelines for ethical use of AI technology, an area which is outside the scope of the current study.

AI tools can enhance productivity and efficiency, and play a valuable role in supporting and augmenting human work. However, they cannot replace humans, who possess moral reasoning, ethical judgement and emotional intelligence. This study showed some of the human aspects of writing that AI-generated texts do not fully replicate. Further research will be needed as the technology continues to evolve, so that teachers can help learners make use of the strengths of AI tools to assist with writing, while knowing how to compensate for weaknesses in the technology.

# References


AlAfnan, M. A. (2015). Language use in computer-mediated communication: An investigation into the genre of workplace emails. *International Journal of Education & Literacy Studies*, *3*(1), 1–11. https://doi.org/10.7575/aiac.ijels.v.3n.1p.1

AlAfnan, M. A., Dishari, S., Jovic, M., & Lomidze, K. (2023). ChatGPT as an educational tool: Opportunities, challenges, and recommendations for communication, business writing, and composition courses. *Journal of Artificial Intelligence and Technology*, *3*(2), 60–68. https://doi.org/10.37965/jait.2023.0184

AlAfnan, M. A., & MohdZuki, S. F. (2023). Do artificial intelligence chatbots have a writing style? An investigation into the stylistic features of ChatGPT-4. *Journal of Artificial Intelligence and Technology*, *3*(3), 85–94. https://doi.org/10.37965/jait.2023.0267

Alharbi, S. H. (2021). A comparative genre-based analysis of move-step structure of RAIs in two different publication contexts. *English Language Teaching*, *14*(3), 12. https://doi.org/10.5539/elt.v14n3p12

Amirjalili, F., Neysani, M., & Nikbakht, A. (2024). Exploring the boundaries of authorship: a comparative analysis of AI-generated text and human academic writing in English literature. *Frontiers in Education*, *9*, 1347421. https://doi.org/10.3389/feduc.2024.1347421

Baron, N. S. (1998). Letters by phone or speech by other means: the linguistics of email. *Language & Communication*, *18*(2), 133–170. https://doi.org/10.1016/S0271-5309(98)00005-6

Baron, N. (2023). *Who wrote this? How AI and the lure of efficiency threaten human writing*. Stanford University Press.

Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, *57*, 100745. https://doi.org/10.1016/J.ASW.2023.100745

Bazerman, C. (1988). *Shaping written knowledge: The genre and activity of the experimental article in science*. University of Wisconsin Press.

Berber Sardinha, T. (2024). AI-generated vs human-authored texts: A multidimensional comparison. *Applied Corpus Linguistics*, *4*(1), 100083. https://doi.org/10.1016/J.ACORP.2023.100083

Bhardwaz, S., & Kumar, J. (2023). An extensive comparative analysis of chatbot technologies - ChatGPT, Google BARD and Microsoft Bing. *Proceedings of the 2nd International Conference on Applied Artificial Intelligence and Computing, ICAAIC 2023*, 673–679. https://doi.org/10.1109/ICAAIC56838.2023.10140214

Bhatia, V. K. (1991). A genre-based approach to ESP materials. *World Englishes*, *10*(2), 153–166. https://doi.org/10.1111/j.1467-971X.1991.tb00148.x

Bhatia, V. K. (1993). *Analysing genre: Language use in professional settings*. Longman.

Bhatia, V. (2004). *Worlds of written discourse*. Bloomsbury Publishing Plc.

Bhatia, V. K. (2008). Genre analysis, ESP and professional practice. *English for Specific Purposes*, *27*, 161–174. https://doi.org/10.1016/j.esp.2007.07.005

Bhullar, P. S., Joshi, M., & Chugh, R. (2024). ChatGPT in higher education - a synthesis of the literature and a future research agenda. *Education and Information Technologies*, 1–22. https://doi.org/10.1007/S10639-024-12723-X/TABLES/5

Biber, D., Connor, U., & Upton, T. A. (2007). *Discourse on the move: Using corpus analysis to describe discourse structure*. John Benjamins Publishing Company.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge University Press.

Bryman, A. (2016). *Social research methods*. Oxford University Press.

Campbell, K. S. (1990). Explanations in negative messages: More insights from speech act theory. *Journal of Business Communication*, *27*(4), 357–375.

Campbell, K. S., Naidoo, J. S., & Smith, J. (2023). When your boss says, "You need to sound more professional": Writing style and writer attributions. *International Journal of Business Communication*, *60*(4), 1071–1094. https://doi.org/10.1177/23294884211025735

Cardon, P., Fleischmann, C., Aritz, J., Logemann, M., & Heidewald, J. (2023a). The challenges and opportunities of AI-Assisted writing: Developing AI literacy for the AI age. *Business and Professional Communication Quarterly*, *86*(3), 257–295. https://doi.org/10.1177/23294906231176517

Cardon, P., Fleischmann, C., Logemann, M., Heidewald, J., Aritz, J., & Swartz, S. (2023b). Competencies needed by business professionals in the AI age: Character and communication lead the way. *Business and Professional Communication Quarterly*, *87*(2), 1–24. https://doi.org/10.1177/23294906231208166

Casal, J. E., & Kessler, M. (2023). Can linguists distinguish between ChatGPT/AI and human writing?: A study of research ethics and academic publishing. *Research Methods in Applied Linguistics*, *2*(3), 100068. https://doi.org/10.1016/J.RMAL.2023.100068

Cohen, L., Manion, L., & Harrison, K. (2018). *Research methods in education* (8th ed.). Routledge.

Coles, C. (2024, May 21). *Shadow AI: how employees are leading the charge in AI adoption and putting company data at risk*. https://www.cyberhaven.com/blog/shadow-ai-how-employees-are-leading-the-charge-in-ai-adoption-and-putting-company-data-at-risk

Coman, A. W., & Cardon, P. (2024). Perceptions of professionalism and authenticity in AI-assisted writing. *Business and Professional Communication Quarterly*. https://doi.org/10.1177/23294906241233224

Cotos, E., Huffman, S., & Link, S. (2015). Furthering and applying move/step constructs: Technology-driven marshalling of Swalesian genre theory for EAP pedagogy. *Journal of English for Academic Purposes*, *19*, 52–72. https://doi.org/10.1016/J.JEAP.2015.05.004

Creelman, V. (2012). The case for “living” models. *Business and Professional Communication Quarterly*, *75*(2), 176–191. https://doi.org/10.1177/1080569912442981

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications Inc.

de Rycker, A. G. H. (2014). Mitigation in turning down business proposals across cultures: The case for pragmatic competence instruction. *3L: Language, Linguistics, Literature*, *20*(1), 87–100. https://doi.org/10.17576/3L-2014-2001-07

Decock, S., de Clerck, B., & van Herck, R. (2020). Interpersonal strategies in e-complaint refusals: Textbook advice versus actual situated practice. *Business and Professional Communication Quarterly*, *83*(3), 285–308. https://doi.org/10.1177/2329490620904952

DeJeu, E. (2024). A rhetorical approach for reimagining business writing instruction in the AI age. *Business and Professional Communication Quarterly*. https://doi.org/10.1177/23294906241255301

Devitt, A. J. (2015). Genre performances: John Swales’ Genre Analysis and rhetorical-linguistic genre studies. *Journal of English for Academic Purposes*, *19*, 44–51. https://doi.org/10.1016/J.JEAP.2015.05.008

Dhillon, P. S., Molaei, S., Li, J., Golub, M., Zheng, S., & Robert, L. P. (2024). Shaping human-AI collaboration: Varied scaffolding levels in co-writing with language models. *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–18. https://doi.org/10.1145/3613904.3642134

Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford University Press.

Dudley-Evans, T. (1994). Genre analysis: An approach to text analysis for ESP. In M. Coulthard (Ed.), *Advances in Written Text Analysis* (pp. 219–228). Routledge.

Elkhatat, A. M., Elsaid, K., & Almeer, S. (2023). Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text. *International Journal for Educational Integrity*, *19*(1), 1–16. https://doi.org/10.1007/S40979-023-00140-5

Flowerdew, J. (1993). An educational, or process, approach to the teaching of professional genres. *ELT Journal*, *47*(4), 305–316.

Flowerdew, L. (1998). Corpus linguistic techniques applied to textlinguistics. *System*, *26*(4), 541–552. https://doi.org/10.1016/S0346-251X(98)00039-6

Flowerdew, L. (2005). An integration of corpus-based and genre-based approaches to text analysis in EAP/ESP: countering criticisms against corpus-based methodologies. *English for Specific Purposes*, *24*(3), 321–332. https://doi.org/10.1016/J.ESP.2004.09.002

Flowerdew, L. (2012). Exploiting a corpus of business letters from a phraseological, functional perspective. *ReCALL*, *24*(2), 152–168. https://doi.org/10.1017/S0958344012000043

Fraiwan, M., & Khasawneh, N. (2023). A review of ChatGPT applications in education, marketing, software engineering, and healthcare: Benefits, drawbacks, and research directions. *ArXiv Preprint ArXiv:2305.00237*.

Fraser, L., Harich, K., Norby, J., Brzovic, K., Rizkallah, T., & Loewy, D. (2005). Diagnostic and value-added assessment of business writing. *Business and Professional Communication Quarterly*, *68*(3), 290–305. https://doi.org/10.1177/1080569905279405

Gains, J. (1999). Electronic mail - A new style of communication or just a new medium?: An investigation into the text features of e-mail. *English for Specific Purposes*, *18*(1), 81–101. https://doi.org/10.1016/S0889-4906(97)00051-3

Getchell, K. M., Carradini, S., Cardon, P. W., Fleischmann, C., Ma, H., Aritz, J., & Stapp, J. (2022). Artificial intelligence in business communication: The changing landscape of research and teaching. *Business and Professional Communication Quarterly*, *85*(1), 7–33. https://doi.org/10.1177/23294906221074311

Gill, S. S., Xu, M., Patros, P., Wu, H., Kaur, R., Kaur, K., Fuller, S., Singh, M., Arora, P., Parlikad, A. K., Stankovski, V., Abraham, A., Ghosh, S. K., Lutfiyya, H., Kanhere, S. S., Bahsoon, R., Rana, O., Dustdar, S., Sakellariou, R., … Buyya, R. (2024). Transformative effects of ChatGPT on modern education: Emerging era of AI chatbots. *Internet of Things and Cyber-Physical Systems*, *4*, 19–23. https://doi.org/10.1016/J.IOTCPS.2023.06.002

Gimenez, J. C. (2000). Business e-mail communication: some emerging tendencies in register. *English for Specific Purposes*, *19*(3), 237–251. https://doi.org/10.1016/S0889-4906(98)00030-1

Giray, L. (2023). Prompt engineering with ChatGPT: A guide for academic writers. *Annals of Biomedical Engineering*, *1*, 1–5. https://doi.org/10.1007/S10439-023-03272-4

Gubala, C., Larson, K., & Melonçon, L. (2020). Do writing errors bother professionals? An analysis of the most bothersome errors and how the writer's ethos is affected. *Journal of Business and Technical Communication*, *34*(3), 250–286. https://doi.org/10.1177/1050651920910205

Gunser, V. E., Gottschling, S., Brucker, B., Richter, S., & Gerjets, P. (2021). Can users distinguish narrative texts written by an artificial intelligence writing tool from purely human text? *Communications in Computer and Information Science*, *1419*, 520–527. https://doi.org/10.1007/978-3-030-78635-9_67

Henry, A., & Roseberry, R. L. (2001). A narrow-angled corpus analysis of moves and strategies of the genre: 'Letter of Application'. *English for Specific Purposes*, *20*, 153–167.

Herbold, S., Hautli-Janisz, A., Heuer, U., Kikteva, Z., & Trautsch, A. (2023). A large-scale comparison of human-written versus ChatGPT-generated essays. Scientific Reports 2023, 13(1), 1–11. https://doi.org/10.1038/s41598-023-45644-9

Ho, V. (2011). What functions do intertextuality and interdiscursivity serve in request e-mail discourse? *Journal of Pragmatics*, *43*(10), 2534–2547. https://doi.org/10.1016/J.PRAGMA.2011.04.002

Hsiao, S. (2024, February 8). *Google Bard is now Gemini: How to try Ultra 1.0 and new mobile app*. https://blog.google/products/gemini/bard-gemini-advanced-app/

Hu, K. (2023, February 2). *ChatGPT sets record for fastest-growing user base - analyst note*. Reuters. https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01

Hyland, K. (2003). Genre-based pedagogies: A social response to process. *Journal of Second Language Writing*, *12*(1), 17–29. https://doi.org/10.1016/S1060-3743(02)00124-8

Hyland, K. (2007). Genre pedagogy: Language, literacy and L2 writing instruction. *Journal of Second Language Writing*, *16*(3), 148–164. https://doi.org/10.1016/J.JSLW.2007.07.005

Hyland, K. (2015). Researching Writing. In B. Paltridge & A. Phakiti (Eds.), *Research methods in applied linguistics: A practical resource* (pp. 335–348). Bloomsbury Publishing Plc.

Hyland, K. (2022). English for Specific Purposes: What is it and where is it taking us? *ESP Today*, *10*(2), 202–220. https://doi.org/10.18485/esptoday.2022.10.2.1

Hyon, S. (1996). Genre in three traditions: Implications for ESL. *TESOL Quarterly*, *30*(4), 722. https://doi.org/10.2307/3587930

Hyon, S. (2018). *Introducing genre and English for Specific Purposes*. Routledge.

Jansen, F., & Janssen, D. (2010a). Effects of positive politeness strategies in business letters. *Journal of Pragmatics*, *42*(9), 2531–2548. https://doi.org/10.1016/J.PRAGMA.2010.02.013

Jansen, F., & Janssen, D. (2010b). Explanations first: A case for presenting explanations before the decision in Dutch bad-news messages. *Journal of Business and Technical Communication*, *25*(1), 36–67. https://doi.org/10.1177/1050651910380372

Jansen, F., & Janssen, D. (2013). Effects of directness in bad-news e-mails and voice mails. *Journal of Business Communication*, *50*(4), 362–382. https://doi.org/https://doi.org/10.1177/0021943613497053

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, *2*, 130–144. https://doi.org/10.1016/j.edurev.2007.05.002

Jovic, M., & Mnasri, S. (2024). Evaluating AI-generated emails: A comparative efficiency analysis. *World Journal of English Language*, *14*(2). https://doi.org/10.5430/wjel.v14n2p502

Knoth, N., Tolzin, A., Janson, A., & Leimeister, J. M. (2024). AI literacy and its implications for prompt engineering strategies. *Computers and Education: Artificial Intelligence*, *6*, 100225. https://doi.org/10.1016/J.CAEAI.2024.100225

Köbis, N., & Mossink, L. D. (2021). Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry. *Computers in Human Behavior*, *114*, 106553. https://doi.org/10.1016/J.CHB.2020.106553

Kong, K. C. C. (1998). Are simple business request letters really simple? A comparison of Chinese and English business request letters. *Text & Talk*, *18*(1), 103–141.

Kong, K. (2014). *Professional Discourse*. Cambridge University Press.

Kwan, B. S. C. (2006). The schematic structure of literature reviews in doctoral theses of applied linguistics. *English for Specific Purposes*, *25*(1), 30–55. https://doi.org/10.1016/J.ESP.2005.06.001

Laerd Statistics. (2019). *Fleiss' kappa using SPSS Statistics*. Statistical Tutorials and Software Guides. https://statistics.laerd.com/spss-tutorials/fleiss-kappa-in-spss-statistics.php

Lima-Lopes, R. E. de. (2006). Genre structure in sales letters. *Trabalhos em lingüística aplicada*, *45*(2), 293–309.

Limaye, M. R. (2001). Some reflections on explanation in negative messages. *Journal of Business and Technical Communication*, *15*(1), 100–110. https://doi.org/10.1177/105065190101500105

Lo, L. S. (2023). The CLEAR path: A framework for enhancing information literacy through prompt engineering. *The Journal of Academic Librarianship*, *49*(4), 102720. https://doi.org/10.1016/J.ACALIB.2023.102720

Locker, K. O. (1999). Factors in reader responses to negative letters: Experimental evidence for changing what we teach. *Journal of Business and Technical Communication*, *13*(1), 5–48. https://doi.org/10.1177/105065199901300101

López-Ferrero, C., & Bach, C. (2016). Discourse analysis of statements of purpose: Connecting academic and professional genres. *Discourse Studies*, *18*(3), 286–310. https://doi.org/10.1177/1461445616634553

Ma, Y., Liu, J., Yi, F., Cheng, Q., Huang, Y., Lu, W., & Liu, X. (2023). AI vs. human - Differentiation analysis of scientific content generation. *ArXiv Preprint ArXiv:2301*.

McKnight, L. (2021). Electric sheep? Humans, robots, artificial intelligence, and the future of writing. *Changing English*, *28*(4), 442–455. https://doi.org/10.1080/1358684X.2021.1941768

McManus, I. C., Thompson, M., & Mollon, J. (2006). Assessment of examiner leniency and stringency ('hawk-dove effect') in the MRCP(UK) clinical examination (PACES) using multi-facet Rasch modelling. *BMC Medical Education*, *6*, 42. https://doi.org/10.1186/1472-6920-6-42

Miller, C. R. (1984). Genre as social action. *Quarterly Journal of Speech*, *70*(2), 151–167. https://doi.org/10.1080/00335638409383686

Moreno, A. I., & Swales, J. M. (2018). Strengthening move analysis methodology towards bridging the function-form gap. *English for Specific Purposes*, *50*, 40–63. https://doi.org/10.1016/j.esp.2017.11.006

Muangsamai, P. (2018). Analysis of moves, rhetorical patterns and linguistic features in New Scientist articles. *Kasetsart Journal of Social Sciences*, *39*(2), 236–243. https://doi.org/10.1016/J.KJSS.2018.03.006

Neuendorf, K. A. (2018). Content analysis and thematic analysis. In P. Brough (Ed.), *Advanced Research Methods for Applied Psychology* (pp. 211–223). Routledge. https://doi.org/10.4324/9781315517971-21

Nguyen, H., & Miller, J. (2012). Exploring business request genres: Students' rhetorical choices. *Business Communication Quarterly*, *75*(1), 5–28. https://doi.org/10.1177/1080569911430379

Okonkwo, C. W., & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. *Computers and Education: Artificial Intelligence*, *2*, 100033. https://doi.org/10.1016/J.CAEAI.2021.100033

Paltridge, B. (1994). Genre analysis and the identification of textual boundaries. *Applied Linguistics*, *15*(3), 288–299. https://doi.org/10.1093/APPLIN/15.3.288

Paltridge, B. (2000). Genre knowledge and teaching professional communication. *IEEE Transactions on Professional Communication*, *43*(4), 397–401.

Park, S., Jeon, J., & Shim, E. (2021). Exploring request emails in English for business purposes: A move analysis. *English for Specific Purposes*, *63*, 137–150. https://doi.org/10.1016/J.ESP.2021.03.006

Peacock, M. (2002). Communicative moves in the discussion section of research articles. *System*, *30*(4), 479–497. https://doi.org/10.1016/S0346-251X(02)00050-7

Perez Sabater, C., Turney, E., & Montero Fleta, B. (2008). Orality and literacy, formality and informality in email communication. *Ibérica*, *15*, 71–88.

Pinto Dos Santos, V. B. M. (2002). Genre analysis of business letters of negotiation. *English for Specific Purposes*, *21*(2), 167–199. https://doi.org/10.1016/S0889-4906(00)00028-4

Rodman, L. (2001). You-attitude: A linguistic perspective. *Business Communication Quarterly*, *64*(4), 9–25.

Roever, C., & Phakiti, A. (2017). *Quantitative Methods for Second Language Research: A Problem-Solving Approach*. Taylor and Francis. https://doi.org/10.4324/9780203067659

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning & Teaching*, *6*(1). https://doi.org/10.37074/jalt.2023.6.1.9

Samraj, B. (2014). Move structure. In K. P. Schneider & A. Barron (Eds.), *Pragmatics of Discourse* (pp. 385–406). De Gruyter Mouton. https://doi.org/10.1515/9783110214406-015

Schryer, C. F. (2000). Walking a fine line: Writing negative letters in an insurance company. *Journal of Business and Technical Communication*, *14*(4), 445–497. https://doi.org/10.1177/105065190001400402

Shelby, A. N., & Reinsch, N. L. (1995). Positive emphasis and you-attitude: An empirical study. *International Journal of Business Communication*, *32*(4), 303–326. https://doi.org/10.1177/002194369503200401

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

Swales, J. M. (2004). *Research genres: Explorations and applications*. Cambridge University Press.

Tardy, C. M., & Swales, J. M. (2014). Genre analysis. In K. P. Schneider & A. Barron (Eds.), *Pragmatics of Discourse* (pp. 165–188). De Gruyter Mouton. https://doi.org/10.1515/9783110214406-007

Teubner, T., Flath, C. M., Weinhardt, C., van der Aalst, W., & Hinz, O. (2023). Welcome to the era of ChatGPT et al.: The prospects of large language models. *Business and Information Systems Engineering*, *65*(2), 95–101. https://doi.org/10.1007/S12599-023-00795-X

Tongpoon-Patanasorn, A., & Thumnong, P. (2020). Move and politeness strategies in job application letters in ASEAN contexts. *LEARN Journal: Language Education and Acquisition Research Network*, *13*(2), 105–125.

Townley, A. (2021). The use of discourse maps to teach contract negotiation communicative practices. *Business and Professional Communication Quarterly*, *84*(1), 59–84. https://doi.org/10.1177/2329490621994217

Upton, T. A., & Cohen, M. A. (2009). An approach to corpus-based discourse analysis: The move analysis as example. *Discourse Studies*, *11*(5), 585–605.

Upton, T. A., & Connor, U. (2001). Using computerized corpus analysis to investigate the textlinguistic discourse moves of a genre. *English for Specific Purposes*, *20*(4), 313–329. https://doi.org/10.1016/S0889-4906(00)00022-3

van Nus, M. (1999). "Can we count on your bookings of potatoes to Madeira?" Corporate context and discourse practices in direct sales letters. In F. Bargiela-Chiappini & C. Nickerson (Eds.), *Writing business: Genres, media and discourses* (pp. 181–205). Longman.

Vergaro, C. (2002). "Dear Sirs, what would you do if you were in our position?". Discourse strategies in Italian and English money chasing letters. *Journal of Pragmatics*, *34*(9), 1211–1233. https://doi.org/10.1016/S0378-2166(02)00046-2

Vergaro, C. (2004). Discourse strategies of Italian and English sales promotion letters. *English for Specific Purposes*, *23*(2), 181–207. https://doi.org/10.1016/S0889-4906(03)00003-6

Walters, W. H. (2023). The effectiveness of software designed to detect AI-generated writing: A comparison of 16 AI text detectors. *Open Information Science*, *7*(1). https://doi.org/10.1515/OPIS-2022-0158

Weigle, S. C. (2002). *Assessing Writing*. Cambridge University Press.

White, M. C. (1988). The tone scale: A five-part activity for measuring and adjusting tone in business correspondence. *Business Communication Quarterly*, *51*(1), 21–23. https://doi.org/10.1177/108056998805100107

Yeadon, W., Agra, E., Inyang, O., Mackay, P., & Mizouri, A. (2024). Evaluating AI and human authorship quality in academic writing through physics essays. *ArXiv Preprint ArXiv:2403.05458*.

Ying, J. (2020). The importance of the discussion method in the undergraduate business classroom. *Humanistic Management Journal*, *5*(2), 251–278. https://doi.org/10.1007/s41463-020-00099-2

Zhai, X., & Razali, A. B. (2023). Triple method approach to development of a genre-based approach to teaching ESL/EFL writing: A systematic literature review by bibliometric, content, and scientometric analyses. *SAGE Open*, *13*(1). https://doi.org/10.1177/21582440221147255

Zhang, Z. (2013). Business English students learning to write for international business: What do international business practitioners have to say about their texts? *English for Specific Purposes*, *32*(3), 144–156. https://doi.org/10.1016/J.ESP.2013.01.002

Zhu, Y. (2000). Structural moves reflected in English and Chinese sales letters. *Discourse Studies*, *2*(4), 473–496. https://doi.org/10.1177/1461445600002004004

# Appendices

The following supplementary materials are included in this section:

Appendix A: Writing prompts

Appendix B: Samples of AI-generated and human-written texts

Appendix C: Scoring instrument (3Cs rubric)

Appendix D: Data elicitation task (JISC Online Surveys)

Appendix E: Research ethics approval

## Appendix A: Writing prompts

| Scenario type | Prompt |
|---|---|
| **Internal** | **A:** You are a professional skills trainer at a corporate training provider. Your colleague, another experienced trainer, is scheduled to deliver the Business Writing Skills 1 workshop three days from now. They have just realised they have another engagement that day and have asked you to deliver the workshop on their behalf. Write an email refusing the request. You can invent names and details. |
| | **B:** You are a professional skills trainer at a corporate training provider. Your line manager is scheduled to deliver the Business Writing Skills 1 workshop three days from now. They have just realised they have another engagement that day and have asked you to deliver the workshop on their behalf. Write an email refusing the request. You can invent names and details. |
| | **C:** You are a professional skills trainer at a corporate training provider. Your colleague, a trainer who has just joined the team, is scheduled to deliver the Business Writing Skills 1 workshop three days from now. They have just realised they have another engagement that day and have asked you to deliver the workshop on their behalf. Write an email refusing the request. You can invent names and details. |
| **External** | **D:** You are a client manager for a corporate training provider. Your customer, who you have worked with for 5 years, has asked to increase the number of participants for an upcoming Business Writing Skills 1 workshop to a total of 30 people. Class sizes are usually a maximum of 20 people. Write an email refusing the request. You can invent names and details. |
| | **E:** You are a client manager for a corporate training provider. Your customer, who you have worked with for 10 years, has asked to increase the number of participants for an upcoming Business Writing Skills 1 workshop to a total of 30 people. Class sizes are usually a maximum of 20 people. Write an email refusing the request. You can invent names and details. |
| | **F:** You are a client manager for a corporate training provider. Your customer, who you have just started working with, has asked to increase the number of participants for an upcoming Business Writing Skills 1 workshop to a total of 30 people. Class sizes are usually a maximum of 20 people. Write an email refusing the request. You can invent names and details. |

## Appendix B: Samples of AI-generated and human-written texts

**Sample 1 – ChatGPT, internal, prompt A (Text 3)**

Dear [Colleague's Name],

I hope this email finds you well. Thank you for reaching out to me regarding the Business Writing Skills 1 workshop scheduled for three days from now. I understand that you have another engagement on that day and are unable to deliver the workshop yourself.

I appreciate your confidence in my abilities to step in and facilitate the session. However, I regret to inform you that I won't be able to take on this request. Unfortunately, I already have prior commitments and cannot rearrange my schedule at such short notice.

I understand the importance of the workshop and the impact it can have on our participants. If there's any way I can support you in preparing materials or providing insights for your session, please let me know. Additionally, if you require assistance in finding an alternative solution, I'm more than happy to collaborate on identifying another qualified facilitator.

I apologize for any inconvenience this may cause and appreciate your understanding in this matter. I trust you'll be able to find a suitable solution for the workshop.

Thank you again for considering me, and I wish you the best for your other engagement.

Best regards,

[Your Name]

**Sample 2 – ChatGPT, external, prompt D (Text 11)**

Dear [Client's Name],

I hope this message finds you well. Thank you for reaching out regarding the upcoming Business Writing Skills 1 workshop.

I truly appreciate your enthusiasm for expanding the scope of the workshop to accommodate more participants. Over the past five years, it's been a pleasure working with you and witnessing the dedication you have toward enhancing the skills of your team.

After careful consideration, however, we have decided to maintain the maximum class size at 20 participants for the upcoming session. Our experience has shown that smaller class sizes allow for more personalized attention and deeper engagement, ensuring that each participant receives the full benefit of the training.

While I understand the desire to include additional team members, I am confident that keeping the class size limited will ultimately lead to a more effective learning experience for everyone involved.

That said, I am more than happy to discuss alternative options to accommodate the needs of your team. Whether it's scheduling additional sessions or exploring customized training solutions, we are committed to finding the best approach to support your objectives.

Thank you for your understanding and cooperation in this matter. Please don't hesitate to reach out if you have any further questions or if there's anything else I can assist you with.

Warm regards,

[Your Name]

---

**Sample 3 – Gemini, internal, prompt B (Text 24)**

---

Dear [Line Manager Name],

Thank you for reaching out about the Business Writing Skills 1 workshop on [Date]. I understand you have another commitment that day and are looking for someone to deliver the training on your behalf.

While I appreciate your trust in my abilities, I'm unable to take on this workshop at such short notice. Unfortunately, I haven't had the opportunity to fully familiarize myself with the workshop materials and wouldn't be confident in delivering it to the same high standard you've set.

Additionally, [Explain your reason for being unavailable, e.g., existing commitments with other workshops, scheduled time off, or preparation for upcoming training].

I wouldn't want to compromise the quality of the training or the learning experience of the participants. However, I'm happy to help in other ways, depending on what you need. Perhaps I could:

- Assist you in finding a suitable replacement trainer with the relevant expertise and experience.
- Provide you with any additional resources or materials you might need to deliver the workshop yourself.
- Offer to support you with the logistics or follow-up after the workshop.

Please let me know if there's anything else I can do to help ensure the smooth delivery of the Business Writing Skills 1 workshop.

Thank you for your understanding.

Sincerely,

---

[Your Name]

**Sample 4 – Gemini, external, prompt E (Text 32)**

Dear Sarah,

Thank you for your email regarding the upcoming Business Writing Skills 1 workshop. I appreciate your continued trust in Learning Solutions as your corporate training provider for the past ten years.

I understand that you'd like to increase the number of participants to 30 for the upcoming session. However, after careful consideration, I'm unable to accommodate this request. Our Business Writing Skills 1 workshops are intentionally designed for a maximum of 20 participants to ensure optimal learning outcomes.

Here are some key reasons for this limit:

- **Interactive Learning Environment:** A smaller class size allows for personalized feedback, active participation, and individual attention, which are crucial for effective writing skill development. With 30 participants, facilitating meaningful discussions and ensuring individual progress becomes challenging.
- **Material Coverage:** The workshop curriculum is carefully crafted for a specific group size. Exceeding this limit could lead to rushed content delivery and potentially incomplete learning for all participants.
- **Trainer Capacity:** Our trainers are highly skilled and experienced, but their effectiveness depends on a manageable class size. Having 30 participants could negatively impact their ability to provide individual support and guidance.

I understand this might not be the answer you were hoping for, but I truly believe that maintaining the current class size is in the best interest of your team's learning and development.

However, I'm happy to discuss alternative solutions! We can:

- **Offer two additional workshops:** Splitting the group into two separate sessions would allow for the desired participant number while maintaining the optimal learning environment.
- **Explore customized training options:** We can tailor a writing program specifically designed for your team's needs and desired number of participants.
- **Consider alternative workshops:** We offer a range of writing-related workshops with varying participant capacities. Perhaps another option could better suit your current requirements.

I'm confident that together we can find a solution that meets your team's writing development goals while ensuring a positive and productive learning experience.

Please let me know your preferred course of action, and I'll be happy to discuss it further.

Thank you for your understanding.

Sincerely,

[Your Name]

**Sample 5 – Human, internal, prompt A (Text 39)**

Hi Abeer,

Nice to hear from you, hope all's good with you too 😊

I'm ever so sorry, but I'm going to be on a training course all day on Friday, so I wouldn't be able to cover the workshop. Have you tried Sheila? I think she was looking for some more hours..

I do hope you find someone, and sorry I can't be of more help!

Catch up soon!

**Sample 6 – Human, external, prompt E (Text 51)**

Hello Tran

Thank you so much for your email. How was your recent holiday back home for Tet?

It's great to be offering Business Writing Skills 1 again for your colleagues. It is amazing to hear that the April run has been oversubscribed and demand is so high. This clearly reflects the course's popularity and high feedback scores.

Given the overwhelming interest in the April course, here are 3 solutions:

1. Use the same trainer to deliver 2 courses during the same week (this can be at your premises again)
2. Run 2 courses in parallel (we have extra training space here in Toa Payoh if your office meeting space is busy)
3. Increase the scheduled Business Writing Skills 1 course to be quarterly

As you know, the first solution offers continuity and uses the trainer's knowledge of Vinfast and previous runs. The second solution reduces time out of the office and could even be presented as a fun team building activity. The final solution allows participants more flexibility each quarter to sign up for this popular course, while expanding your overall Learning and Development offerings.

Please let me know which solution suits you and Vinfast best. All solutions allow you to increase participants to a maximum of 20 on the course. As there is an additional and unplanned cost, I can give you a special 10% discount on the additional course.

I look forward to confirming the dates and details with you in the next week.

All my best

Sarka

## Appendix C: Scoring instrument (3Cs rubric)

*This was provided in Part 1 of the online data elicitation task for participants to rate each of their allocated texts. It was accessed by clicking on a link, which opened the rubric below as a PDF in a new window.*

# Scoring rubric

*Give one score for each category (Clarity, Credibility, Connection) from 0 to 5 in the online form.*
0 = very poor; 5 = excellent

| **CLARITY** |
|---|
| • The **language** used is modern, jargon-free and easy to understand<br>• **Organisation** and **layout** are logical and easy to follow<br>• Sentences are **concise**<br>• Explanations are precise and **easy to understand** |
| **CREDIBILITY** |
| • The content is **meaningful and useful** for the reader<br>• **Alternatives** are offered to provide a solution<br>• The language is **positive** and focuses on what can be done<br>• **Grammar, spelling and punctuation** are correct |
| **CONNECTION** |
| • The **tone** is professional and empathetic, not mechanistic<br>• The writer shows **awareness of the reader's needs**<br>• The **opening** section is specific and reader-friendly<br>• The **closing** section looks to the future and is personalised |

# Appendix D: Data elicitation task (JISC Online Surveys)

## Screen 1 - Participant information and consent

### Participant Information and Consent

This page provides information about the study. Please read the information and answer two questions at the bottom.

**Genre and scoring analysis of business refusals written by humans and artificial intelligence (AI) tools: Is there still a case for business writing training?**

Central University Research Ethics Committee Approval Reference: EDUC_C1A_23_306

**General Information**

The aim of this research is to investigate the features of human and AI-generated business refusal texts. It will also explore how professional skills trainers perceive the use of AI in business writing.

We appreciate your interest in participating in this online task. You have been invited to participate as you are a professional skills trainer or business English teacher. Please read through this information before agreeing to participate by ticking the 'yes' box below.

You can ask any questions before deciding to take part by contacting the researcher (details below). The Principal Researcher is Winny Wilson, who is attached to the Department of Education at the University of Oxford. This research is being completed under the supervision of Professor Heath Rose. The British Council is supporting this research through the Professional Development Scheme.

This online task is in two parts and should take about 25-30 minutes. In Part 1, you will be asked to score three business texts based on the rubric provided. You will also judge whether each text is AI-generated or written by a human writer and explain your assessment. In Part 2, you will be asked three general questions about your perceptions of AI tools in business contexts.

**Do I have to take part?**

No. Participation is voluntary. If you do decide to take part, you may withdraw at any point for any reason before submitting your answers by pressing the 'Exit' button/ closing the browser.

**How will my data be used?**

We will not collect any data that could directly identify you.

Your IP address will not be stored. We will take all reasonable measures to ensure that data remain confidential.

The responses you provide will be stored in a password-protected electronic file on University of Oxford secure servers and may be used in academic publications and conference presentations. Research data will be stored for three years after publication or public release of the work of the research.

The results will be written up for a Master's dissertation.

**Who has reviewed this research?**

This research has been reviewed by, and received ethics clearance through, a subcommittee of the University of Oxford Central University Research Ethics Committee [EDUC_C1A_23_306].

**Who do I contact if I have a concern or I wish to complain?**

If you have a concern about any aspect of this research, please contact Winny Wilson (winny.wilson@education.ox.ac.uk) or dissertation supervisor Heath Rose (heath.rose@education.ox.ac.uk). We will do our best to answer your question.

Please note that you may only participate in this survey if you are 18 years of age or over. *

○ I certify that I am 18 years of age or over

If you have read the information above and agree to participate with the understanding that the data you submit will be processed accordingly, please select the box below to start. *

○ Yes, I agree to take part

Next

**Screen 2 - Task information**

*Participants were randomly allocated one text from each writer group (ChatGPT, Gemini and Human), generated or written from one of the six prompts (Appendix A). The order in which texts appeared was randomised. Participants were not aware of the authorship of texts they were shown. As mentioned on this screen, salutations and sign-offs were removed from all texts for the scoring task, so that participants would not be biased in making their authorship decisions.*

### Task information

PART 1: You will score three business refusal texts based on the rubric provided. You will also judge whether each text is AI-generated or written by a human writer and explain your assessment.

PART 2: You will answer three general questions about your perceptions of AI tools in business contexts.

**For the three texts you will see, the same prompt below was given to both the human writers and AI tools:**

> *"You are a professional skills trainer at a corporate training provider. Your colleague, another experienced trainer, is scheduled to deliver the Business Writing Skills 1 workshop three days from now. They have just realised they have another engagement that day and have asked you to deliver the workshop on their behalf. Write an email refusing the request. You can invent names and details."*

(Salutations and sign-offs have been removed from all emails.)

Previous

Next

**Screens 3, 4, 5 – Part 1**

*The same questions appeared three times, with different texts inserted into the "Xx" field.*

**PART 1: Scoring the texts - Text 1**

This is the first scoring task. You need to:
1. Score the text based on the rubric - click on the link to open in a new tab
2. Answer the questions about the text

Click here to open the rubric (PDF)

Xx

Refer to the rubric. What is your score on each category? *

Clear selected

| | 0 = poor | 1 | 2 | 3 | 4 | 5 = excellent |
|---|---|---|---|---|---|---|
| Clarity | ○ | ○ | ○ | ○ | ○ | ○ |
| Credibility | ○ | ○ | ○ | ○ | ○ | ○ |
| Connection | ○ | ○ | ○ | ○ | ○ | ○ |

What is your overall impression of the quality of this text? Give it a score out of 10.

Clear selected

| | 1 = poor | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Overall score | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

This text was written by... *

- ○ a human
- ○ probably a human
- ○ probably an AI tool
- ○ an AI tool (e.g., ChatGPT)

What influenced your decision on whether the text was written by a human/AI tool? Please explain why. *

Previous

Next

**Screen 6 – Part 2**

PART 2: Your perceptions of AI in business writing

Please write freely about your thoughts in the three open-ended questions below.

What are your concerns about using AI tools in business writing? *

What do you think are the benefits of using AI tools in business writing? *

What challenges does the emergence of AI tools present for the training of business writing skills? *

Previous

Next

## Screen 7 – Demographic data

### Final questions: About you

This data will not be used to identify you as an individual. It will be aggregated and used to provide background context.

How many years have you been working in professional skills training and/or business English teaching? (Enter closest number of whole years.)

What country are you currently based in?

Have you used AI tools before to help with writing?

- ◯ Yes
- ◯ No

How would you rate your familiarity with AI tools? (1=not familiar at all; 5=very familiar)

Clear selected

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Familiarity with AI tools | ◯ | ◯ | ◯ | ◯ | ◯ |

Previous Next

## Screen 8 – Closing

### THANK YOU!

This study aims to fill a gap in the research literature on business writing, AI tools and trainer/teacher perceptions of AI.

You have just contributed to filling this gap. A very big thank you for your time and effort!

Remember to click the SUBMIT button :)

Previous Submit

# Appendix E: Research ethics approval

**SOCIAL SCIENCES & HUMANITIES**
**INTERDIVISIONAL RESEARCH ETHICS COMMITTEE**
**DEPARTMENTAL RESEARCH ETHICS COMMITTEE**

Department of Education
15 Norham Gardens, Oxford OX2 6PY
student.curec@education.ox.ac.uk staff.curec@education.ox.ac.uk

UNIVERSITY OF OXFORD

Winny Wilson
Department of Education, Social Sciences Division
University of Oxford

14 November 2023

Dear Winny,

**Research ethics approval**

**Research title:** Genre and scoring analysis of business refusals written by humans and AI tools: Is there still a case for business writing training?

**Research ethics reference: EDUC_C1A_23_306**

The above application has been considered on behalf of the Education Departmental Research Ethics Committee (DREC) in accordance with the University's procedures for ethical approval of all research involving human participants.

I am pleased to confirm that, on the basis of the information provided to the DREC, ethics approval has now been granted for this study.

Please note the following:

**Personal data**: It is the responsibility of the PI to ensure that all personal data collected during the project is managed in accordance with the University's guidance and legal requirements.

**In-person activities**: Any data collection involving in-person interactions with participants must have an up-to-date fieldwork risk assessment in place; further guidance is available from the Safety Office's website.

**Amendments**: Please notify the committee if you intend to make any amendments to the information in your ethics application as submitted at date of this approval, as all changes must receive ethical approval prior to implementation. The amendment form is available on the SSH IDREC webpage.

We welcome feedback on your experience of the ethical review process and suggestions for improvement. Please email any comments to staff.curec@education.ox.ac.uk / student.curec@education.ox.ac.uk or ethics@socsci.ox.ac.uk.

Yours sincerely

DREC Dr Jenny A. Wynn

cc: Professor Heath Rose