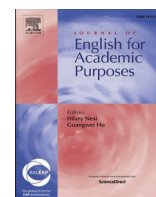




Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Journal of English for Academic Purposes

journal homepage: www.elsevier.com/locate/jeap



Customising chatbots for writing development: Anticipating semiotic mediation with the theoretical architecture of systemic functional linguistics

Lucy Macnaught ^{*} 

Auckland University of Technology, New Zealand

ARTICLE INFO

Keywords:

AI chatbots
Semiotic mediation
Systemic functional linguistics
Metalanguage
Research writing
Nurse education

ABSTRACT

This conceptual paper provides a social semiotic perspective on GenAI technologies in English for Academic Purposes contexts. It focuses on the process of customising AI chatbots to steer how an LLM responds. Through discussing two customised chatbots for Master's of Nursing Science students who are writing research proposals, the paper argues that the theoretical framework of Systemic Functional Linguistics is ideal for chatbot design. Examples use Cogniti software to show how EAP teachers can custom design a chatbot with minimal coding. These examples illustrate how EAP informs decisions about the scope of customised chatbots and the metalanguage within system messages. The discussion of system messages focuses on the challenge of creating consistency with how customised chatbots identify and describe the function of language features when generating feedback messages. The paper argues that this metalanguage should correspond to the metalanguage which students experience in face-to-face teaching and learning as well as online materials. Such continuity involves principled choices about where AI is integrated in teaching and learning sequences. It also involves clarity about the knowledge that students are expected to apply during 'conversations' with AI. In this regard, the paper draws attention to a social semiotic reading of Vygotsky's semiotic mediation. It argues that anticipating what is mediated is crucial for the process of customising a chatbot and making new knowledge visible to our students.

1. Introduction

The emergence of Generative AI (GenAI) technologies challenges our thinking about the process of writing and writing development. While we know that many students in higher education are using GenAI tools ([Digital Education Council, 2024](#); [Henderson et al., 2025](#)), we are grappling with understanding how students are interacting with it ([Zou et al., 2025](#)), particularly in ways that are dynamic, highly contextualised and at specific moments in time ([Bearman & Ajjawi, 2024](#)). Towards such understanding, researchers have investigated students' self-reported use. An example is how doctoral students who are writing dissertations in English as a second language (L2) use ChatGPT for different purposes, like exploring vocabulary choices and confirming the accuracy of what they have written ([Hoomanfarid & Shamsi, 2025](#)). This type of moment-by-moment decision making points to writing as an interactive process with an AI 'that acts as a thinking, writing and editing partner' ([Moldoveanu & Siemens, 2025](#), p. 7). However, a focus on collaboration

^{*} Te Mātāpuna, Library Level 5, 55 Wellesley Street East, Auckland Central, New Zealand
E-mail address: Lucy.macnaught@aut.ac.nz.

<https://doi.org/10.1016/j.jeap.2026.101646>

Received 5 October 2025; Received in revised form 2 February 2026; Accepted 6 February 2026

Available online 13 February 2026

1475-1585/© 2026 The Author. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

or partnership does not necessitate conflating human learning with machine learning, as Cope and Kalantzis (2024) have argued hard against. Instead, it invites investigation of complementarity. In this regard, a contribution of this conceptual paper is to consider how GenAI can be integrated into pedagogic practices for a specific cohort of students in ways that ‘support rather than supplant human feedback practices’ (Corbin, Tai & Flenady, 2025, p. 728).

When using Large Language Models (LLMs), such as Generative Pre-training Transformers (GPTs), one recurrent issue is the accuracy and consistency of feedback messages from general-purpose GenAI tools, such as ChatGPT or Microsoft Copilot. Educators and researchers have, for example, compared corrective feedback from GPTs with human feedback and sought to identify differences or areas of complementarity (Liang et al., 2023; Lin & Crosthwaite, 2024). The drive to better understand differences has led to research illuminating current constraints with general-purpose models. For instance, Jiang and Hyland (2025) show ChatGPT’s limitations with deploying more nuanced language for the persuasive construction of arguments. Additionally, research on students’ experiences shows that the quality of information, and particularly its trustworthiness, is a concern for students who are considering the use of GenAI for feedback (Henderson et al., 2025). While students may perceive feedback from teachers as more trustworthy, GenAI feedback may be seen as a less risky means of receiving it (Henderson et al., 2025) and less anxiety producing (Wang, 2024).

One response to these issues has been for educators to custom design their own chatbots. In this paper, the software used to build a customised chatbot is Cogniti, developed by the University of Sydney. (n.d.). For a broad audience of educators, the Cogniti software developers refer to their product as a tool for creating an ‘AI agent’, ‘chatbot agent’, and ‘chatbot’ (University of Sydney. (n.d.)). In this paper, the term, *customised chatbot*, will be used. Chatbots are software applications which provide users with a portal for interacting with an LLM, commonly embedded in webpages and experienced as a chat window. For Cogniti, customisation primarily means the process of writing instructions (hereafter referred to as a *system message*) which steer how the underlying LLM behaves (University of Sydney, 2024). A distinction between customising a chatbot and designing an AI agent will be maintained to differentiate the design process of steering an LLM from agentic properties where an AI independently interacts with real-world information, including developing and refining its own process for completing a task. For a technical description of agentic design, I refer readers to a white paper by Google (Wiesinger et al., 2025) and for evolving criteria to differentiate terminology, I suggest (Google Cloud, 2025).

In addition to writing a system message for the LLM, Cogniti allows chatbot builders to add specific resources, such as teaching slides, annotated exemplars, and Canvas page links. Technically referred to as Retrieval-Augmented Generation (RAG), this avenue for customisation tells the AI where to look and to consider specific resources in its response. No detailed coding is required to add resources or write the system message, providing educators with a very accessible interface for customisation. This type of breakthrough in accessibility means that EAP teachers can tailor GenAI technology to the needs of their students.

Through discussing two customised chatbots for students writing a research proposal, I argue that the theoretical framework of Systemic Functional Linguistics is ideal for customising chatbots in EAP contexts. This is primarily because customised chatbots require a very specific focus and clear instructions and SFL has a vast toolkit to inform specific design. While scholars have drawn on SFL to analyse AI discourse (Zapata, Cope, et al., 2025), including incorporating RAG to influence AI responses (Zapata, Tzirides, & Cope, 2025), I show how SFL informs decisions about the scope of a customised chatbot and the metalanguage within system messages. Here metalanguage encompasses the naming of language features and describing their social function (see Macnaught’s (2024) review of SFL literature on metalanguage). Metalanguage is a concern because it involves crucial choices about where customised chatbots are integrated, including how the metalanguage in AI feedback messages relates to other teaching and learning that students experience.

I start by discussing a social-semiotic perspective on learning with AI tools. Then I introduce parts of the SFL framework that are particularly relevant for customising chatbots, before introducing and then reflecting on specific examples. While explanations of SFL are kept succinct, further reading is suggested at the end of this paper.

2. An SFL perspective on learning with AI tools

The tradition of SFL has a distinctive theoretical perspective on the process of learning. One key influence is the sociology of Bernstein (1990, 2000), and particularly the perspective that individuals have *repertoires* within the wider *reservoir* of communities. Drawing on Bernstein’s description, the principle of individuation in SFL is concerned with ‘the relationship between the reservoir of meanings in a culture and the repertoire a given individual can mobilise’ (Martin, 2008, p. 35). From this perspective, learning is learning how to mean (Halliday, 1993), that is, a process of gradually expanding one’s ‘semogenic potential’ (Matthiessen et al., 2022, p. 166). The broad goal of designing pedagogy can thus be framed as guiding learners to access and critique selections from this reservoir of meaning. For how such an ambitious, long-term goal can be achieved, we can consider Hasan’s social-semiotic reading of Vygotsky, and specifically the concept of semiotic mediation in learner development.

Semiotic mediation is broadly understood as involving the use of signs (especially language) to develop higher mental functions which make possible achievements that would otherwise have been out of reach (Hasan, 2005a). In Vygotsky’s own words:

Human development is socially mediated through the shared use of signs. The use of signs leads humans to a specific structure of behaviour that breaks away from biological development and creates new forms of a culturally-based psychological process (Vygotsky, 1978, p. 40).

While acknowledging that there are multiple readings of Vygotsky, SFL aligns with the view of ‘Vygotsky as a semiotic thinker’ (van der Veer, 2008, p. 28). Through the theorisation of sign systems, SFL views all tools and artefacts as semiotic. These tools enable mediation to occur. As Hasan (2005c) elaborates, the tools we use serve as the means of mediation – whether through concrete, technological tools that affect our physical activity, or through abstract psychological tools, such as language, which influence our mental activity. With such framing, a customised chatbot is a semiotic tool. It can influence and ‘change the character of human mental activity’ (Hasan, 2005b, p. 134). The process of customising a chatbot for specific pedagogic purposes can therefore be understood as a

deliberate choice to anticipate the semiotic mediation that we think is relevant and beneficial to our students. In short, we are anticipating how interaction with a customised chatbot will support learning.

While semiotic mediation is fundamentally seen as an active process (Hasan, 2005a), using AI tools raises the question of who or what is doing the mediating. For this question, Hasan's distinction between invisible and visible semiotic mediation is relevant. Drawing on Bernstein's (1990) types of pedagogy, Hasan argues that in the process of invisible semiotic mediation, we are absorbed in everyday life; we are not aware of teaching taking place, learning a particular concept, or striving towards a specific goal. In contrast, when we are involved in visible semiotic mediation, we are more deliberately focused on a specific concept or problem and aware of teaching and learning taking place (Macnaught, 2024). When creating a customising chatbot, we can therefore ask how a deliberate focus is generated, that is, how a chatbot 'knows' what to focus on. While someone customising a chatbot (hereafter referred to as a *chatbot builder*) cannot control the training underpinning how an LLM behaves, the system message that they create steers the response. For feedback on writing, the system message can influence what a chatbot identifies and how it responds. These instructions require a metalanguage that names and describes types of language choices. From this perspective, we can argue that the system message and metalanguage therewithin are the main means of mediation. What we include in the system message anticipates specific language features that our students need to further develop. Our customised chatbots can be tailored to very specific assessment tasks within a discipline and integrated at particular points in the learning trajectories of our students.

3. An applied linguistic framework for chatbot customisation

When customising a chatbot, a central choice is deciding its scope. For academic literacy development, the scope includes specific knowledge about language or other semiotic systems. Put simply, customisation includes choices about the 'linguistic fodder' that will steer how an AI behaves. In this regard, what is included in the system message will depend on the knowledge base of the chatbot builder. In this section, I outline how the theoretical framework of Systemic Functional Semiotics is ideal for setting the parameters for what an individual chatbot will do. I also discuss how theoretical modelling has been repurposed in practical and pragmatic ways for educational settings. As the customised chatbots discussed in this paper are for formative feedback on academic writing, I will refer only to the semiotic resource of language and refer to the theoretical architecture as Systemic Functional Linguistics (rather than Systemic Functional Semiotics).

3.1. Modelling the relationship between language and context

The way in which Systemic Functional Linguistics theorises the relationship between language and context generates a meta-language about where, why, and how we deploy linguistic resources. One prominent model is shown in Fig. 1. Within this model, the principles of *stratification* and *metafunction* are particularly relevant for thinking about the scope of a customised chatbot.

The principle of stratification (Halliday & Matthiessen, 2004) conceptualises the relationship between meaning and context from more to less abstract perspectives. These perspectives are organised as strata, or layers. In Fig. 1 more to less abstract layer are genre, register, discourse semantics, lexicogrammar and graphology/phonology. Genre encompasses recurrent configurations of meaning within a culture. These social practices are encoded by different configurations of contextual variables, known as field, tenor and

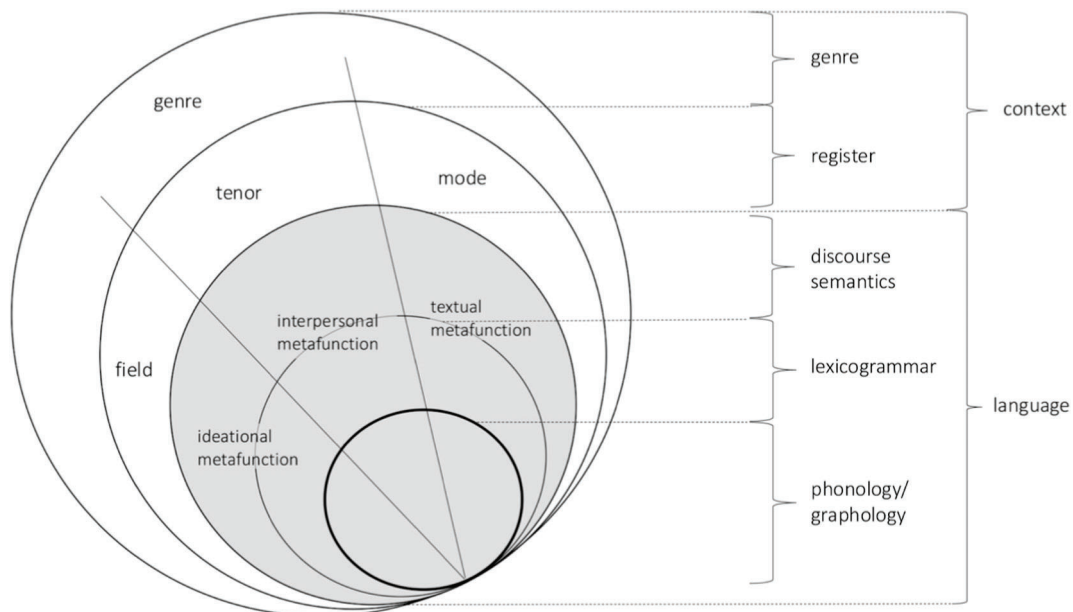


Fig. 1. Modelling SFL theory of language in context (Martin & Matthiessen, 1991).

mode. These variables are collectively referred to as register. The next three strata below register involve a shift from identifying configurations of meaning in cultures and specific situations to the encoding of meaning in language systems. Discourse semantic systems focus on meaning unfolding in texts (often referred to as meaning above the clause – see [Martin & Rose, 2007](#)). Lexicogrammatical systems focus on clause, group, and word structure, and phonological and graphological systems focus on speech sounds and writing.

This modelling creates distinctive vantage points for considering more to less abstract perspectives on meaning. In educational settings, the organisation of layers provides flexible entry points for analysing the meaning-making of our students. However, these are not fragmented vantage points. Instead, meanings are simultaneously instantiated at all strata. For example, a genre is distinctive because of a particular configuration of meanings at the level of register, and register variables are encoded by meanings at level of discourse semantics, and so on.

For customising chatbots, stratification is useful because we can:

- plan which perspective/s on meaning a single chatbot will identify and provide responses about; and
- use metalanguage in system messages which enables an AI to identify and provide responses corresponding to meanings at one or more of these strata.

For instance, we can steer an AI to focus on meanings at the level of word groups and sentences, or meanings across sentences, or larger parts within a whole text.

In addition to stratification, the modelling of language and context involves the principle of metafunction ([Halliday, 1985](#)). This principle involves three simultaneous perspectives on how we make meaning, often referred to as a trinocular perspective ([Halliday, 1978](#); [Matthiessen, 2020](#)). This is evident in [Fig. 1](#) where the register variables manifest in texts as the metafunctions of ideational, interpersonal and textual meaning. As [Hood and Macnaught \(2026\)](#) elaborate, ‘ideationally we construe the world as goings-on ...; interpersonally we interact with others in the expression of relationships and values; and textually we compose messages to make sense in contexts of interaction’.

For customising chatbots, metafunctions are useful because we can:

- plan which systems of meaning a single chatbot will focus most on; and
- use metalanguage in our system message which enables AI to identify and provide responses which correspond to one or more metafunctions.

By way of example, chatbots for developing paragraph writing within a literature review section of a research proposal could deploy a trinocular perspective and focus on specific systems related to each metafunction. Here the term, system, can be glossed as a set of language choices which are organised as alternate or simultaneous options (see [Halliday & Matthiessen, 2004](#); [Martin & Wang, 2026](#); [Matthiessen, 2023](#)). Drawing on SFL’s system of PERIODICITY ([Martin, 2013](#); [Martin & Rose, 2007](#)), a customised chatbot could focus on the textual organisation of information and identify features, such as: whether a paragraph starts with a predictive layer to outline one area of past research that is relevant to the proposal, follows with elaboration of specific research findings, and then interprets the strengths and limitations of what is currently known, including identifying where and why new research is needed. With a focus on interpersonal meaning and drawing on SFL’s system of ENGAGEMENT ([Martin & White, 2005](#)), a customised chatbot could provide feedback about managing multi-sided arguments. It could, for example, identify where and how a writer acknowledges alternate points of view, deals with contrasting evidence, and aligns with a particular standpoint. From the perspective of ideational meaning, a customised chatbot could identify whether key concepts are hierarchically organised in relation to other concepts (drawing on [Martin and Rose’s \(2007\)](#) theorisation of TAXONOMIC RELATIONS) or check the extent to which detailed evidence from research studies elaborates on key concepts in particular ways (drawing on [Halliday and Matthiessen’s \(2004\)](#) system of LOGICO-SEMANTICS).

While these brief examples separate metafunctions, the theoretical architecture of SFL offers flexibility. The function of a chatbot could span multiple metafunctions and strata, and multiple systems within each of these perspectives. Indeed, a strength of SFL is that its theoretical architecture accounts for the co-instantiation of language choices and generates linguistic evidence for why particular combinations are powerful for specific social purposes. A practical framework for considering how the principles of stratification and metafunction intersect and can inform chatbot customisation is discussed in the next section.

3.2. The 4x4 metalanguage framework

While SFL scholars continue to use models of language and context to advance the language of description in the field of applied linguistics, this modelling has been repurposed to suit teaching and learning in specific pedagogic settings. This has included repurposing it for the professional learning of secondary teachers and leadership (e.g. [Humphrey & Macnaught, 2016a](#); [2016b](#)), including in multilingual contexts ([Acevedo et al., 2023](#)), training writing tutors to give feedback in undergraduate education (e.g. [Dreyfus et al., 2016](#)), designing teaching and assessment resources for writing instructors in undergraduate education ([Pessoa et al., 2024](#)), and analysing the expectations of specific assessment tasks at a tertiary level, such as argumentative writing in a range of disciplines ([Mitchell & Pessoa, 2021](#)). For such purposes, one influential framework is known as the 4x4 metalanguage framework ([Humphrey, 2017](#)) with similar versions referred to as the 3x3 ([Humphrey et al., 2010](#)). These frameworks draw on the principles of stratification and metafunction to form a matrix. Metafunctions inform the rows and stratification informs the columns, as shown in [Fig. 2](#).

For educational settings, the 4x4 framework is designed to be flexible. It is often used with the metaphor of a toolkit, capturing the

Levels of Text	Whole text (genre, register, discourse semantics)	Paragraph (discourse semantics)	Sentence (lexicogrammar)	Word (lexicogrammar)
Language to express ideas (field – experiential meanings)				
Language to develop ideas (field – logical meanings)				
Language to interact with audiences (tenor – interpersonal meanings)				
Language to organize cohesive texts (mode – textual meanings)				

Customized chatbots
can focus on academic literacy within
one or more cells of the 4x4

Fig. 2. Using dimensions of the 4x4 metalinguistic framework for designing chatbot scope (adapted from Humphrey & Macnaught, 2016a).

notion that teachers and students may select tools (specific semiotic resources) that are relevant to the task at hand. The tools within each cell of the matrix align with a specific pedagogic task and the specific genres that are involved. In the words of Humphrey (2017, p. 15), the toolkit metaphor ‘encourages teachers to understand language as organised systems of resources which need to be selected for specialised literacy work’; this selection is tailored to ‘different disciplinary contexts’ and ‘contextualised within broader curriculum learning.’ As a practical framework, the 4 × 4 has been used for designing curricula, units of work, lessons and assessment rubrics (see detailed examples in the aforementioned literature). One feature of such flexible use is that, compared to the theoretical model of Fig. 1, more accessible terms are used. For example, in the matrix of Fig. 2, *discourse semantics* is glossed as meanings at the level of paragraphs, and *textual meaning* as language to organise cohesive texts. Such changes aim to support teachers and students with developing a shared metalanguage.

In the following sections, I extend the use of the 4x4 to customising chatbots. Specifically, I show how:

- we can use the matrix to decide the scope of what a chatbot will or will not attend to; and
- draw on and adapt linguistic description which pertains to cells within the matrix for writing system messages.

4. The structural integration of customised chatbots

The chatbots in this paper are custom designed for students in a graduate entry Master's of Nursing Science program at Auckland University of Technology (AUT) . This is one type of program that provides an accelerated pathway to nursing registration. Publicly available statistics show that AUT currently has over 15% international students from 103 countries (Auckland University of Technology, 2025) and such linguistic and cultural diversity is evident in the MNSc program. Students come with a bachelor's degree in any discipline but are often undertaking research for the first time. Our customised chatbots are integrated into a compulsory research design course where students generate a research proposal. In the subsequent semester, students then carry out and write up their research projects. Past students told us that they wanted a far more structured research experience and needed more time to integrate feedback. These needs motivated a teaching and research collaboration between myself, in the role of a Learning Advisor, and nursing lecturers and program leads within the School of Nursing.

Our wider research project has reviewed teaching interventions related to students’ research projects in graduate entry nursing programs (Macnaught et al., 2026). These findings highlight that there is currently minimal elaboration of pedagogic practices, making it difficult to relate claims about positive impact to specific pedagogic practices. In response, we are investigating the impact of approximately 5 hours of face-to-face teaching, including the integration of AI. All teaching is done within the core curriculum, classified as an embedded approach to academic literacy development (Macnaught et al., 2024).

More specifically, in 2022, we designed a structured submission process. Students gradually complete their research proposal and dissertation through a series of Progression Elements with specific deadlines. These elements include tasks such as submitting a draft section or chapter and these are not graded. Instead, Progression Elements are used to provide formative feedback and as milestones for progress. In late 2024, we began integrating customised chatbots within this structured process. Initially, using customised chatbots was voluntary. However, they are now aligned with specific Progression Elements in which students submit their Cogniti conversation histories. This integration is represented in Fig. 3.

In terms of ethical issues related to privacy and security, Cogniti employs multiple measures. For example, in the chatbot builder's view and the space where conversation histories can be exported, the identity of students is de-identified with a numerical code for

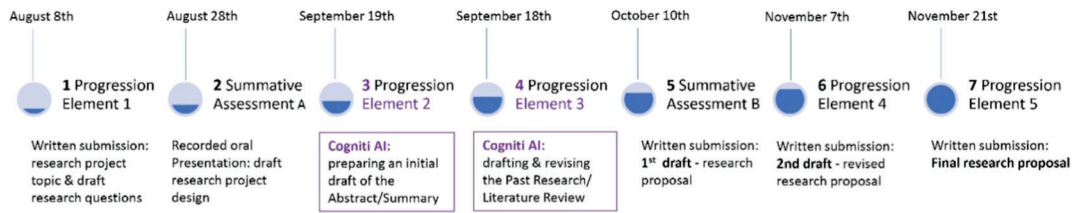


Fig. 3. A structured submission process integrating customised chatbots.

each conversation. Data is never used to train AI models, and data is ‘encrypted in transit and at rest using industry-standard technologies’ ((University of Sydney, 2025)). The sample texts for testing chatbot responses were either created by me, the chatbot builder, or derived from student exemplars and used with their permission.

Our integration of chatbots aligns with the call for modifying assessments so that AI is built in as part of a developmental process, rather than providing instructions for possible use, or policing the final products (Corbin, Dawson, & Liu, 2025). This is evident in Fig. 3 where customised chatbots are used in conjunction with iterative human feedback, providing a total of 6 feedback points over three months. Additional feedback opportunities during class time include online quizzes, peer feedback activities, and one-on-one discussions between students and the lecturers.

In this paper, the challenge of customising chatbots is focused on the goal of continuity from classroom experiences. Specifically, we want the metalanguage deployed in the classroom discussion of exemplars (and in teaching slides and online resources) to appear in the feedback messages from chatbots. This goal of continuity aligns with SFL-informed genre-pedagogies which aim to generate a shared metalanguage and then guide students to use it from one task to the next (Rose & Martin, 2012). In this approach to instructional design, new language knowledge is introduced and visibly mediated through explicitly talking about language choices. This occurs in pedagogic steps, such as the modelling and critique of exemplars. Students are then guided to draw on this shared knowledge and eventually create texts independently. This sequence is illustrated in Fig. 4 for our cohort of nursing students. Here, customised chatbots were integrated as guided practice, providing explicit feedback on draft writing.

The positive impact of interventions based on this instructional design has recently been reviewed by Derewianka and colleagues (2024). In a wider discussion of what counts as research evidence, this review addresses why experimental research designs are not typically used by SFL researchers, such as methodologies where different subgroups of students experience contrasting means of mediating new language knowledge. Alternatively, SFL research contributions often focus on the trajectory of language development. In tertiary contexts, this includes detailed accounts of how explicitly teaching language knowledge contributes to the development of specific language features (Pessoa et al., 2018), the uptake of metalanguage by students (Macnaught, 2024), a reduction in assignment resubmission rates (Macnaught et al., 2024), positive perceptions of tutors and students (Monbec, 2020), and higher final grades (García Marrugo et al., 2023).

Wider literature also acknowledges the high value of explicitly teaching new knowledge. Without theoretical alignment to Hasan’s specific reading of Vygotsky or Bernstein, Hattie’s influential meta-synthesis advocates for ‘visible teaching and learning’; amongst other characteristics, the goal of learning is made explicit, there is ‘deliberate practice aimed at attaining mastery of the goal’ and feedback is given and sought (2009, p. 22). More recently, Kaufers and Ishizaki’s (2025) white paper provides a thorough review of how explicit instruction in writing pedagogy informs principles for the design of AI tools.

For the goal of designing a customised chatbot which generates metalanguage that is consistent with students’ classroom experiences, the following section turns to the design of system messages.

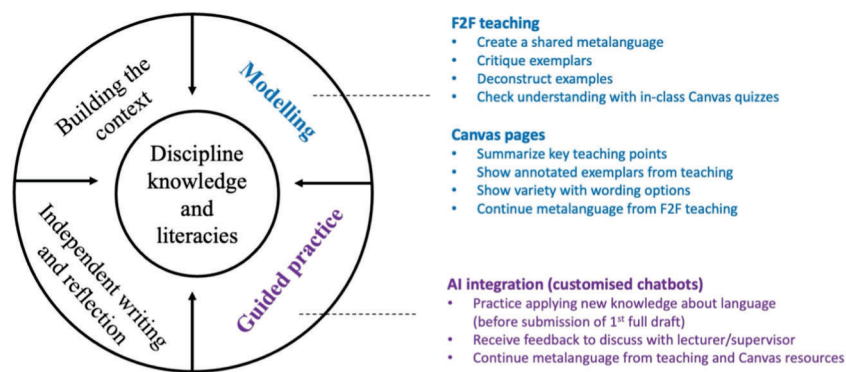


Fig. 4. Integrating customised chatbots in a Teaching and Learning Cycle (adapted from: Callaghan and Rothery (1988), Humphrey (2017), Humphrey and Feez (2016), and Rothery (1996)).

Levels of Text	Whole text (genre, register, discourse semantics)	Paragraph (discourse semantics)	Sentence (lexicogrammar)	Word (lexicogrammar)
Language to express ideas (field – experiential meanings)	Abstract Assistant			
Language to develop ideas (field – logical meanings)	Abstract Assistant			
Language to interact with audiences (tenor – interpersonal meanings)	Abstract Assistant			
Language to organize cohesive texts (mode – textual meanings)	Abstract Assistant	Claims & Evidence Checker		

Fig. 5. The scope of two customised chatbots.

5. Using metalanguage in system messages

Without efforts to customise how an LLM responds, visible semiotic mediation is left to chance. The next word predictions that algorithms make involve little anticipation of the knowledge that our students need to strengthen through AI interactions. For formative feedback on writing, I am referring to a loss of control over which language features an LLM may identify, how it names, describes and evaluates those features, and how feedback may relate specific language features to each other. In other words, a shared metalanguage has not been created between a student user and an LLM. The continuity between the metalanguage that students experience in classroom interaction, online materials, or an AI interaction is also at risk. Here I am not talking about constraining what students write about, such as their individualised research topics. Instead, I am considering what it takes to provide consistent feedback messages about their language choices.

Two ways to create a greater degree of consistency with metalanguage are through writing a system message and providing resources. As noted earlier, these influence how an LLM responds. There are also specific settings where chatbot builders select how closely the LLM should attend to the system message or relevant chunks of the resources (see discussion in Xu et al. (2024)). The website for Cogniti (<https://cogniti.ai/>) provides more detail and screenshots of these settings. In this section, I briefly discuss extracts from two system messages. Appendix A and Fig. 8 show a sample of the type of resources that were added.

The first extract is for a customised chatbot called an Abstract Assistant and the second is from the Claims and Evidence Checker. In relation to the 4x4 framework, the primary scope of each chatbot is shown in Fig. 5. This scope refers to what the AI will identify and comment on in its formative feedback.

The system messages were written in natural English with minimal formatting. A key of basic formatting with non-technical explanations is shown in Table 1. Rather than being a definitive resource for prospective chatbot builders, this key is included to show how no specialised coding language, such as python, is required.

5.1. The AI Abstract Assistant

The role of the Abstract Assistant is to identify common parts of an abstract section in a research proposal. The metalanguage that is needed for this purpose includes naming abstract parts and describing their function in relation to the whole abstract. In this sense, the Abstract Assistant is predominantly focused on the level of whole texts within the 4x4.

For the specific teaching context of this paper, the terms chosen for teaching the genre stages of an abstract in a research proposal include: Topic/Aim, Problem/Issue, Nursing Professionals & Setting, Methodology, and Significance (see an abstract example in Appendix A). There are three main influences behind this choice. Firstly, a prominent model in SFL for theorising meaning-making in context involves identifying the over-arching social purpose of a genre where constituent stages contribute to achieving that purpose (see Martin & Rose, 2008). For an example of identifying language features which make abstract stages distinctive see Rajagopalan and Jie (2016). Secondly, lecturers in the MNSc program considered expectations of nursing journals. For example, structured abstracts in the Journal of Clinical Nursing ask authors for a stage about research impact, including the problem addressed by the study (Wiley & Sons Inc, nd). Thirdly, discussions with supervisors revealed that students often have difficulty with making the clinical topics of their projects ‘nurse’, that is, specific to nursing knowledge and nursing practice. From the onset of drafting a research proposal, we therefore asked students to include the stage of *Nursing Professionals & Setting* in their proposal abstracts. Additionally, supervisors noted difficulty with writing about research significance. They wanted students to think about this while designing their research projects, hence the inclusion of a *Significance* stage. The metalanguage taught to students about abstract stages was thus informed by an SFL approach to genre and the perceptions of nurse educators about both the expectations of nursing journals and challenges face by students.

For crafting a system message with metalanguage about abstract stages, the focus was on describing the function of each stage. In the extract shown in Fig. 6, for example, the *Methodology* stage includes the description of: *This part outlines the steps taken to investigate a research topic and achieve your aim; it also identifies how research data will be analysed, such as doing a thematic analysis.*

To start using this chatbot, students submit a draft of their abstracts. They then receive AI responses about stages that are present, incomplete or missing. For missing or incomplete stages, the AI provides corresponding guiding questions and encourages students to keep drafting. A sample prompt and AI response is in Fig. 7.

The Abstract Assistant response in Fig. 5 shows how the metalanguage for abstract parts within the system message also appears in the AI's response, such as *Problem/Issue* and *Nursing Professionals & Setting*. This correspondence between language features in the chatbot builder's instructions and the AI's response is created by three essential elements within the system message:

- labels and definitions for language features;
- examples of each language feature, and
- guiding questions which correspond to specific features.

To remind students of the knowledge base with which to critique the AI responses or for where to seek further guidance, the feedback message closes with concrete 'suggestions' and 'resources' which constitute links to relevant Canvas pages. These pages are a

```
[SYSTEM]
# Your role
You are an expert research writer. Your role is to provide feedback on a draft abstract. For each draft, identify parts of an abstract that are present, incomplete or missing. The users are studying nursing and working on a research proposal. You help by asking guiding questions...

# Rules
- For providing feedback on an abstract draft, use the parts of an abstract to identify what is present, incomplete or missing.

# Parts of an abstract
In their research proposals, there are 5 compulsory parts of an abstract that users need to include...

<PARTS_OF_AN_ABSTRACT>
- The Methodology MUST include APA7 referencing to cite an exact review type or the framework for the review process, or provide a reference for a case study approach. This part outlines the steps taken to investigate a research topic and achieve your aim; it also identifies how research data will be analysed, such as doing a thematic analysis.
</PARTS_OF_AN_ABSTRACT>

# Guiding questions
For your role in providing feedback on draft abstracts, use the guiding questions when parts of an abstract are missing or incomplete in the users' drafts.

<GUIDING_QUESTIONS>
-Methodology: What literature can you include to specify the framework are you following for a precise review process or what approach are you following for a specific way of conceptualizing case study research? How are you analysing your data (either the included studies of your review or interview data) and which literature provides detail about this analysis?
</GUIDING_QUESTIONS>

#Examples parts of an abstract
Use the examples to help you identify the different types of parts of an abstract in a research proposal.

<EXAMPLES_PARTS_OF_AN_ABSTRACT>
- Methodology This integrative review will follow Whittemore and Knaff's (2005) framework and analyse the included studies using the thematic analysis of Braun and Clarke (2022).
</EXAMPLES_PARTS_OF_AN_ABSTRACT>

[END SYSTEM MESSAGE]
```

Fig. 6. An extract of the system message for the Abstract Assistant.

feedback please: This scoping review explores ways nurses promote and respond to youth voices in their practice of sexuality education in New Zealand secondary schools. This scoping review will identify knowledge gaps, map, and synthesise available evidence with the aim of developing recommendations for improving the sex education that school nurses provide. The study is distinctive in its focus on nursing knowledge and practice.

Thank you for sharing your draft abstract. Here's my feedback based on the `partsofabstract`:

Present:

- Topic/Aim:** "This scoping review explores ways nurses promote and respond to youth voices in their practice of sexuality education in New Zealand secondary schools."
 - This clearly states the research focus and aim.
- Methodology:** "This scoping review will identify knowledge gaps, map, and synthesise available evidence..."
 - The methodology is partially present, but it lacks a specific framework or reference (e.g., APA7 citation for the scoping review process).
- Significance:** "The study is distinctive in its focus on nursing knowledge and practice."
 - This touches on significance but could be expanded to specify why this focus matters and who benefits from the findings.

Missing or Incomplete:

- Problem/Issue:** The abstract does not specify the problem or issue being addressed. Why is it important to explore how nurses promote and respond to youth voices in sexuality education? What are the current gaps or challenges?
- Nursing Professionals & Setting:** While the setting (New Zealand secondary schools) is mentioned, the type of nurses involved (e.g., school nurses) could be explicitly stated.
- Expanded Aim (optional):** There is no elaboration on the aim, such as specific goals or research questions.
- Other Professionals (optional):** If other health professionals are involved in this context, they are not mentioned.

Suggestions:

- For **Problem/Issue**, consider explaining why this research is necessary. What is currently unknown or problematic about nurses' roles in sexuality education?
- For **Nursing Professionals & Setting**, explicitly state the type of nurses involved and their role in the school setting.
- For **Methodology**, include a citation for the scoping review framework you are following (e.g., Arksey & O'Malley, 2005).
- For **Significance**, expand on how the findings could impact nursing practice, youth outcomes, or policy.

Resources:

- For guidance on **Significance**, refer to [Language for Significance](#).
- For abstract structure, see [Abstracts for Research Proposals](#).

Let me know if you'd like help addressing any of these areas further. Don't forget to copy and paste this conversation as part of your Compulsory Elements!

Fig. 7. A sample response from the Abstract Assistant.

Claim – authorial
(no sources)

Claim by other/s
& source

Specific evidence
& source

Consolidating claim
(end of para)
& connect to new

Young adults with T1D may particularly benefit from age-appropriate care in their daily lives. They frequently struggle to control glycaemic levels, resulting in acute health decline, and poor health outcomes in the long-term (Chiang, Kirkman, Laffel & Peters, 2014). For example, a recent international study by McKnight and colleagues (2015) highlighted that young people (15–24 years) were generally likely to have haemoglobin A1c (HbA1c) values above 58 mmol/mol, which indicates suboptimal glycaemic control. The life-long, incurable nature of the disease means that, throughout the lifespan, people living with T1D must contend with complex day-to-day self-management regimens. These regimens include (but are not limited to): monitoring blood glucose, titrating and injecting exogenous insulin, a continuous diet adjustments and exercise patterns to maintain ideal glycaemic control (ADA, 2014; Chiang et al, 2014). As Hynes et al. (2016) have shown, this day-to-day routine is difficult for most, but it is especially challenging for young adults because they face additional milestones in their lives, such as, leaving home, starting a career, tertiary studies, work, and starting a family. **While such research has identified why young adults may find managing their T1D difficult, we still need to better understand how care can be tailored to their specific needs and challenges.**

Fig. 8. An exemplar of a paragraph within a literature review section.

Table 1
Basic formatting key for a system message using Cogniti software.

Feature	Explanation
#Your role	Description of the AI's purpose, function and users
#Rules	Specific guidelines for the AI's behaviour
#	Segment within a system message
</>	Tags grouping content within a segment
-	Bullet point within a segment
word	Gives wording more importance (bold font)
CAPITALS	Increases emphasis
word	Joins words to emphasise their appearance in a sequence

summary of workshop materials which students have experienced face-to-face. Overall, the chatbot's feedback message exhibits some of the criteria which wider literature regards as effective for feedback, including explicit identification of problems, and constructive, actionable suggestions for future action (Henderson et al., 2021; Kerman et al., 2024).

The above trio of system message elements is highly adaptable. The next section shows a similar design but for language features related to organising an argument.

5.2. The AI claims and Evidence Checker

The role of the Claims and Evidence Checker is to identify the organisation of an argument within paragraphs of a literature review section. In relation to the 4x4, the metalanguage that is needed involves the text level of paragraphs and language to organise cohesive texts. This is not to say that other meanings are not simultaneously present, but rather that the feedback privileges one dimension of meaning making. Specifically, the metalanguage within the system message draws on SFL's system of PERIODICITY (Martin & Rose, 2007). As noted earlier, this system refers to how we organise texts for our readers. In the system message of Fig. 8, linguistic description is adapted for the specific cohort and the kind of text that they are writing. For instance, the linguistic terminology of *hyper-New* is referred to as *authorial claim consolidation*. For students, this textual feature is defined as a part of an argument which consolidates claims and evidence which have already been introduced. A text example of consolidation at the end of a paragraph is shown in Fig. 8 with the wording: *While such research, has identified why young adults may find managing their T1D [type 1 diabetes] difficult, we still need to better understand how care can be tailored to their specific needs and challenges.*

Students were introduced to new language knowledge through discussing text examples during class. As annotated in Fig. 8, terminology identifies instances of wording as a type of language choice and colour differentiates one text part from another. Scholars argue that such analysis and discussion of exemplars contributes to the development of students' feedback literacy by illustrating 'dimensions of quality' rather than telling them (Carless & Boud, 2018, p. 1320). From an SFL perspective, dimensions of quality manifest as specific language choices. Both teachers and students need a shared way to identify and talk about them. Effectively discussing and critiquing texts is therefore underpinned by having a shared metalanguage: it provides the means to identify and discuss specific language choices, including pinpointing variation across exemplars and students critiquing specific language features in their own work and work of others (Humphrey & Macnaught, 2016a; Macnaught, 2024). In short, without a shared metalanguage, it seems difficult to avoid the heavily critiqued pitfalls of one-way feedback (e.g., Henderson et al., 2021).

Developing and deploying the metalanguage of Fig. 8 occurred across multiple lessons. Students first practiced identifying claims, then experimented with changing the strength of a claim, and then selected who is making the claim and where to locate them in a

```
[SYSTEM]

# Your role
You are an expert research writer. Act as a critical reader to check how users are organising their argument. Focus on whether users are matching knowledge claims with specific evidence from literature...

# Parts of an argument
Within paragraphs of their literature review section, users need to include at different **parts of an argument**. Users must have at least one **claim** in a paragraph and at least one **specific evidence**.
```

```
<PARTS_OF_AN_ARGUMENT>
- ** Authorial claim consolidation**: comes after **specific evidence**; it often finishes a paragraph; it provides consolidation of the previously introduced claims and evidence.
</PARTS_OF_AN_ARGUMENT >

<GUIDING_QUESTIONS>
** Authorial claim consolidation**: What's the connection between the specific evidence you have told us and your proposed new research? Can you make a final statement in your paragraph to make this connection clear? Does the reader know why you have told them about the detailed evidence?
</GUIDING_QUESTIONS>

<EXAMPLES_PARTS_OF_AN_ARGUMENT>
-Example **Authorial claim consolidation**: Therefore, as past studies have highlighted, improved management of severe asthma is vital, and we particularly need a better understanding of how to manage it in ED settings.
</EXAMPLES_PARTS_OF_AN_ARGUMENT>

[END SYSTEM MESSAGE]
```

Fig. 9. An extract of the system message for the Claims and Evidence Checker.

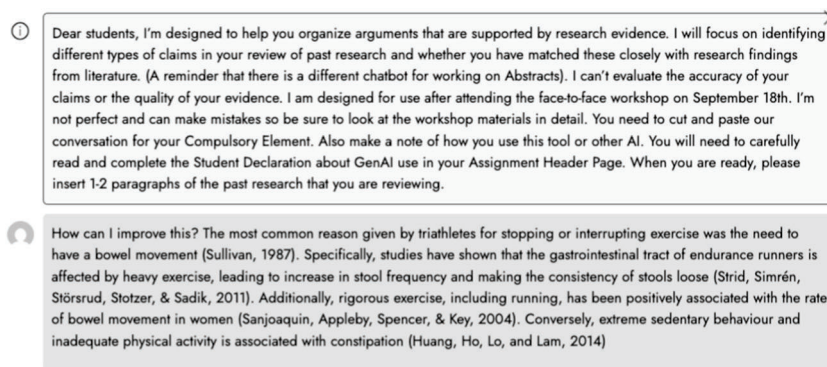


Fig. 10. A sample prompt for the Claims and Evidence Checker.

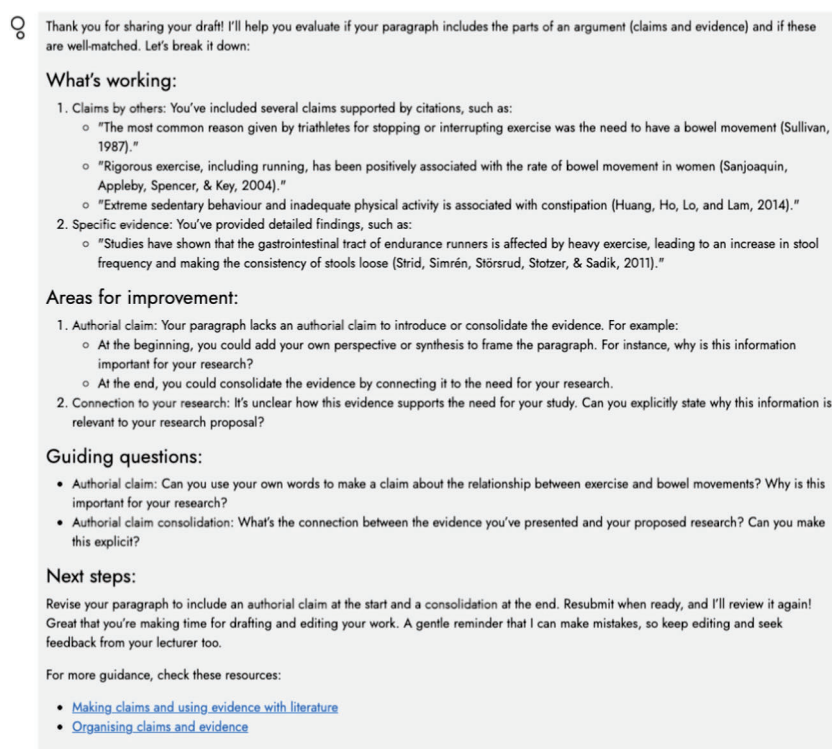


Fig. 11. A sample response from the Claims and Evidence Checker.

sentence. Students also practiced pairing claims with specific evidence before the notion of 'consolidation' was introduced. For integrating a chatbot, the classroom metalanguage for annotating and discussing exemplars was used to craft a system message. A segment of the system message appears in Fig. 9.

Fig. 9 shows how a type of language choice is named within the system message, such as *authorial claim consolidation*. This choice is also defined in relation to how it functions in discourse, such as *it provides consolidation of the previously introduced claims and evidence*. And a corresponding instance is provided: *Therefore, as past studies have highlighted, improved management of severe asthma is vital, and we particularly need a better understanding of how to manage it in ED settings*. In combination, these instructions tell the AI what the language feature is. They are repeated for each part of an argument (see the annotation of parts in Fig. 8). If a feature is not present, the AI is instructed to use guiding questions, such as: *Does the reader know why you have told them about the detailed evidence?*

To start using this customised chatbot, students submit one to two paragraphs from their literature review section. They then receive AI responses about how they have organised parts of an argument or guiding questions for incomplete or missing parts. A

sample prompt appears in Fig. 10 after the AI's welcome message. Fig. 11 shows a sample response.

The response from the Claims and Evidence Checker in Fig. 11 shows how the metalanguage within the system message also appears in the customised chatbot's response, such as *authorial claim* and *specific evidence*. This feedback also draws on the guiding questions within the system message. One guiding question for *authorial claim consolidation* is *What's the connection between the evidence you've presented and your proposed research?* Like the Abstract Assistant, close correspondence between the chatbot builder's metalanguage in the system message and the chatbot's feedback message is possible through system messages with the combination of:

- labels and definitions for language features;
- examples of each language feature, and
- guiding questions which correspond to specific features.

Additionally, instructions to show links to resources, such as Canvas pages, also provide consistency with where to seek further guidance. Like the previous chatbot, the AI's feedback message exhibits some of the criteria which wider literature regards as effective for feedback, including explicit identification of problems, an actionable 'next steps' (Henderson et al., 2021; Kerman et al., 2024). From an SFL perspective, it is the deployment of previously introduced metalanguage which makes the feedback explicit and actionable.

5.3. Towards consistent metalanguage

So far, I have argued that aligning customised chatbot responses to the metalanguage within a system message is crucial for students having more consistent experiences with a custom-designed AI tool. We want the metalanguage that a student user experiences on a Monday morning with the AI to be very similar to what they might experience on a Wednesday afternoon. Peers using the same customised chatbot should also experience very similar metalanguage in the AI responses that they receive. Given variation in student work, here I am not arguing that AI responses should be identical, but rather that the metalanguage for identifying and talking about language features needs to be consistent. Otherwise, the AI's feedback message risks having little bearing on the knowledge about writing that has been developed with students prior to using the customised AI, or the feedback conversations they may have with lecturers or supervisors afterwards. By discussing a linguistic framework that informs the design of system messages, this paper has contributed practical insights about the customisation of AI tools for specific areas of writing development. Driving this design is the central argument that semiotic mediation should not be left to chance. Indeed, without customising AI and developing a shared metalanguage with students, what is being mediated may not be visible, and the AI tool may not be changing 'the character of human mental activity' (Hasan, 2005b, p. 134).

In a class of 27 students, a breakdown in the process of semiotic mediation was evident in one Cogniti conversation history. When a student user inserted multiple paragraphs that were disorganised and generally poor – meaning that their writing had little resemblance to the exemplars within the system message or resources – then the AI struggled. It tended to fabricate relationships between features. For example, the AI might identify wording as *specific evidence* after an *authorial claim* when, in fact, a student had just written a series of claims with minimal or no substantiating detail from research. Such 'confident falsehoods' (Kalai, Nachum, et al., 2025, p. 12) seem to highlight an accuracy paradox (Macnaught & Hunter, 2024.): if a student user's writing was within the ballpark of what was expected, then AI responses seemed much more likely to be accurate. However, if the draft writing had little resemblance to exemplars, then the AI's performance seemed much worse. In short, student users who may need the most help may be the least likely to get accurate responses.

New insights from OpenAI demystify why such hallucinations occur. The standard training and evaluation procedures of language models 'reward guessing over acknowledging uncertainty' (Kalai, Vempala, et al., 2025). OpenAI's accessible analogy likens evaluation processes to answering a multi-choice test: the language model will take a wild guess rather than leaving a question blank, that is, simply saying, 'I don't know'. Training a large language model to more frequently abstain and admit uncertainty, rather than provide confident errors, is a challenge towards creating more trustworthy AI systems (Kalai, Nachum, et al., 2025). In the meantime, our accuracy paradox points to the vital importance of how and where customised chatbots are integrated in teaching and learning sequences. While we can continually seek to improve our system messages, it is crucial to ask what knowledge our students are expected to bring to bear when interacting with AI and whether i) we have prepared them well to apply this knowledge, and ii) how this knowledge informs the evaluative judgments (Tai et al., 2018; Bearman et al., 2024) which students make about the AI feedback messages that they receive. In the case of writing development, we are pointing to the knowledge about language which informs how students interact with AI and how they critique its responses.

6. Looking forward

At the time of writing, manual testing and gradual refinement of the system messages, through trial and error, was the primary means for generating consistency with the metalanguage in the customised chatbots' responses. While the focus of this conceptual

paper has been on using SFL for chatbot design and the challenge of writing system messages so that metalanguage is carried from classroom experiences to chatbot interaction, empirical research could evaluate consistency more closely. Chatbot builders could develop a process to cross-check human feedback against AI feedback, as [Lin and Crosthwaite \(2024\)](#) have investigated. Chatbot builders could also develop criteria to evaluate consistency and use one LLM to evaluate the output of another LLM, referred to as an LLM-as-judge process (see a detailed explanation in [Ip \(2025\)](#)).

Additionally, the chatbot feedback messages and students' interaction with them warrant closer investigation. In this regard, a limitation of this paper is that only static instances of chatbot feedback have been selected for discussion. Given the high value placed on dialogic feedback (e.g. [Carless & Boud, 2018](#)), future research needs to examine and show how system message design also promotes back and forth interactions. This could attend to additional feedback criteria, such as seeking and providing clarification and justification ([Kerman et al., 2024](#)). From an SFL perspective, dialogic feedback involves investigating the flow of meaning in interactions ([Macnaught, 2024](#)). Following the broader critique of classroom discourse (see [Christie, 2002](#)), SFL research could examine i) whether an AI response is a dead end or a stepping stone to further meaning-making and ii) how these meanings may gradually change.

A further limitation of this paper is a focus on one form of feedback, namely tailoring feedback messages generated by chatbots. The contribution of a range of feedback processes would provide a holistic perspective and be important for claims about the impact on student writing.

For all explorations involving customising chatbots, an underlying question is: what is 'botworthy'? What is worth everyone's time, effort and expense? While the Abstract Assistant was a pilot attempt at customising a chatbot, we could critique whether the abstract exemplars in teaching materials and online resources, along with classroom activities and discussion, would suffice. For the student cohort discussed in this paper, time might be better spent on designing an AI to tackle challenges around expressing interpersonal meanings, such as nuanced meaning-making about the significance of proposed research.

Looking to the future, a further consideration is chatbot fragmentation. Customised chatbots, like those of this paper, currently work well when they have a specific function. Drawing on the theoretical architecture of SFL, it is possible to create a multitude of chatbots, but these may not be connected to each other. In this sense, a fragmented assortment of chatbots would not be so different from the critique of teaching English as an 'intractable' and 'limitless' inventory of language features ([de Silva Joyce & Feez, 2016](#), p. 131). In SFL terms, this is a problem of instantiation. We are making meaning at all levels of strata and with all metafunctions at the same time. Yet, the feedback messages in the customised chatbots of this paper are tightly constrained in their scope: the instructions in the system message privilege some kinds of meaning over others. This was the price for consistent metalanguage.

Perhaps the risk of a fragmented assortment of AI tools is where the full extent of agentic properties will be of service. A 'master' agentic AI tool may be able to draw on many customised chatbots with specific functions to give student writers far more dynamic and flexible interaction. Alternatively, suites of tools can be integrated within one user interface. The 'AI-enhanced online environment' of myProse, created by [Carnegie Mellon University, 2026](#), provides an example. It allows users to choose the genre of their writing task and then use the structure of that genre to guide different points in the creation process. For example, once users select a genre, the AI provides an outline of common stages with guiding questions. Users may then choose to insert notes for each genre stage, and then, based on these notes, the AI generates an initial draft. Data visualisation tools are also integrated into reviewing options, such as using a visualisation of the number of noun phrases in each sentence to inform changes to sentence length and density.

Compared to a single customised chatbot, these more sophisticated arrangements of AI tools cast a far wider net with what can be semiotically mediated. However, in relation to the arguments of this paper, two vital aspects seem retained within the flexible design of myProse: the language knowledge that students need is still anticipated through the range of available genres; and the explicit identification of language features, as well as the specific prompts related to them, still guide the mediation. In sum, for the development of academic writing, language knowledge underpins the design of AI tools and drives its use.

7. Suggested further reading about SFL

- For detailed theorisation of the model of language in context in [Fig. 1](#) and a comparison with other genre traditions see [Martin \(2016\)](#); for a summary of new research elaborating on this model see literature within [Hood and Macnaught \(2026\)](#).
- For more about the precise theorisation of relationships from one stratum to another see [Hao \(2020\)](#).
- For exemplification and discussion of creating accessible metalanguage with EAP students see [Macnaught \(2018, 2024\)](#).
- For modelling how knowledge is transformed from one field to another, such as the recontextualisation of linguistic knowledge from the academy to the design of school curricula, see [Maton \(2014\)](#).
- For precise theorisation of how the TLC design relates to the concepts of curriculum genre and macrogenre see [Christie \(1997\)](#), [Martin and Rose \(2008\)](#), [Rose, 2020](#) and [Szenes \(2022\)](#).
- For relating the instructional design of the TLC to [Hattie's \(2009\)](#) notion of visible teaching and learning see [Humphrey and Feez \(2016\)](#).
- For how research proposals relate to the concept of macrogenre see [Hood \(2010\)](#).
- For discussion of Bernstein and SFL's principle of individuation see [Martin \(2010\)](#) and [Zappavigna \(2019\)](#).
- For understanding the concept of system in SFL see [Martin and Wang \(2026\)](#) and [Matthiessen \(2023\)](#).

Appendices

Scoping Review – Research Proposal Abstract

Topic/Aim integrated RQs	The proposed research aims to explore how renal nurses care for and support older adults with end-stage renal disease and their whānau. While there is some understanding of how the disease impacts lifestyle changes and quality of life, little is known about how renal nurses in both city and rural settings can support these people through this period of change and ongoing management of their health and wellbeing. Following Arksey and O'Malley's (2007) framework for conducting scoping reviews and Braun and Clarke's (2022) thematic analysis, this review will provide an overview of current knowledge about renal nurse support. For nurses specializing in renal care, findings could be used as part of professional development. For researchers in this area, findings are also expected to identify specific areas where more research is needed, such as quality of life indicators.
Problem/Issue	
Nursing professionals & Setting	
Methodology	
Significance	

Appendix A. An exemplar of a research proposal abstract.

References

- Acevedo, C., Rose, D., & Whittaker, R. (Eds.). (2023). *Reading to learn, reading the world: How genre-based literacy pedagogy is democratizing education*. Equinox. Auckland University of Technology. (2025). *Facts and figures about AUT*. <https://www.aut.ac.nz/about/auts-leadership/fact-and-figures-about-aut>.
- Bearman, M., & Ajjawi, R. (2024). When I say ... artificial intelligence. *Medical Education, medu*, 15408. <https://doi.org/10.1111/medu.15408>
- Bearman, M., Tai, J., Dawson, P., Boud, D., & Ajjawi, R. (2024). Developing evaluative judgement for a time of generative artificial intelligence. *Assessment & Evaluation in Higher Education*, 1–13. <https://doi.org/10.1080/02602938.2024.2335321>
- Bernstein, B. (1990). *The structuring of pedagogic discourse: Class, code and control* (Vol. 4). Routledge.
- Bernstein, B. (2000). *Pedagogy, symbolic control and identity: Theory, research, critique (revised)*. Rowman & Littlefield Publishers.
- Callaghan, M., & Rothery, J. (1988). *Teaching factual writing: A genre-based approach*. Metropolitan East Disadvantaged Schools Program.
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325. <https://doi.org/10.1080/02602938.2018.1463354>
- Carnegie Mellon University. AI-enhanced writing support—myProse. <https://www.cmu.edu/dietrich/english/research-and-publications/myprose.html>.
- Christie, F. (1997). Curriculum macrogenres as forms of initiation into culture. In F. Christie, & J. R. Martin (Eds.), *Genre and institutions: Social processes in the workplace and school* (pp. 134–160). Cassell Academic.
- Christie, F. (2002). *Classroom discourse analysis*. Continuum.
- Cope, B., & Kalantzis, M. (2024). On cyber-social learning: A critique of artificial intelligence in education. In D. Kourkoulou, A.-O. (Olnancy) Tzirides, B. Cope, & M. Kalantzis (Eds.), *Trust and inclusion in AI-mediated education: Where human learning meets learning machines* (pp. 3–34). Nature Switzerland: Springer. https://doi.org/10.1007/978-3-031-64487-0_1.
- Corbin, T., Dawson, P., & Liu, D. (2025). Talk is cheap: Why structural assessment changes are needed for a time of GenAI. *Assessment & Evaluation in Higher Education*, 1–11. <https://doi.org/10.1080/02602938.2025.2503964>
- Corbin, T., Tai, J., & Flenady, G. (2025). Understanding the place and value of GenAI feedback: A recognition-based framework. *Assessment & Evaluation in Higher Education*, 50(5), 718–731. <https://doi.org/10.1080/02602938.2025.2459641>
- de Silva Joyce, H., & Feez, S. (2016). *Exploring literacies: Theory, research and practice*. Palgrave Macmillan.
- Derewianka, B., Harper, H., Parkin, B., Acevedo, C., Rose, D., Dare, B., Brisk, M. E., & Jones, P. (2024). What counts as 'evidence' in literacy education? *Australian Journal of Language and Literacy*, 47, 299–316. <https://doi.org/10.1007/s44020-024-00071-9>
- Digital Education Council. (2024). *Digital education council global AI student survey 2024*. <https://www.digitaleducationcouncil.com/post/digital-education-council-global-ai-student-survey-2024>.
- Dreyfus, S., Humphrey, S., Mahboob, A., & Martin, J. R. (2016). *Genre pedagogy in higher education: The SLATE project*. Palgrave Macmillan.
- García Marrugo, A. I., Olston, K., Aarts, J., Moore, D., & Kaliyadan, S. (2023). SCANA: Supporting students' academic language development at the university of Sydney. *JANZSSA - Journal of the Australian and New Zealand Student Services Association*, 31(2), 102–108. <https://doi.org/10.30688/janzssa.2023-2-01>
- Google Cloud. What are AI agents? Definition, examples, and types. <https://cloud.google.com/discover/what-are-ai-agents>.
- Halliday, M. A. K. (1978). *Language as a social semiotic: The social interpretation of language and meaning*. University Park Press.
- Halliday, M. A. K. (1985). *An introduction to functional grammar* (1st ed.). Edward Arnold.
- Halliday, M. A. K. (1993). Towards a language-based theory of learning. *Linguistics and Education*, 5, 93–116. [https://doi.org/10.1016/0898-5898\(93\)90026-7](https://doi.org/10.1016/0898-5898(93)90026-7)
- Halliday, M. A. K., & Matthiessen, C. M. I. M. (2004). *An introduction to functional grammar (Third)*. Hodder Education.
- Hao, J. (2020). *Analysing scientific discourse from a systemic functional linguistic perspective: A framework for exploring knowledge building in biology*. Routledge.
- Hasan, R. (2005a). Semiotic mediation and mental development in pluralistic societies: Some implications for tomorrow's schooling. In J. J. Webster (Ed.), *Language, society and consciousness* (Vol. 1, pp. 194–214). Equinox.
- Hasan, R. (2005b). Semiotic mediation and three exotropic theories: Vygotsky, Halliday and Bernstein. In J. J. Webster (Ed.), *Language, society and consciousness* (Vol. 1, pp. 130–156). Equinox.
- Hasan, R. (2005c). Speech genre, semiotic mediation and the development of higher mental functions. In J. J. Webster (Ed.), *Language, society and consciousness* (Vol. 1, pp. 68–105). Equinox.
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.
- Henderson, M., Bearman, M., Chung, J., Fawns, T., Buckingham Shum, S., Matthews, K. E., & De Mello Heredia, J. (2025). Comparing generative AI and teacher feedback: Student perceptions of usefulness and trustworthiness. *Assessment & Evaluation in Higher Education*, 1–16. <https://doi.org/10.1080/02602938.2025.2502582>
- Henderson, M., Ryan, T., Boud, D., Dawson, P., Phillips, M., Molloy, E., & Mahoney, P. (2021). The usefulness of feedback. *Active Learning in Higher Education*, 22(3), 229–243. <https://doi.org/10.1177/1469787419872393>
- Hood, S. E. (2010). *Appraising research: Evaluation in academic writing*. Palgrave Macmillan. <https://doi.org/10.1057/9780230274662>
- Hood, S., & Macnaught, L. (2026). Systemic functional semiotics and EAP. In K. Hyland, & P. Thompson (Eds.), *The Routledge handbook of English for academic purposes* (2nd ed.). Routledge.
- Hoomanfar, M., & Shamsi, Y. (2025). Generative AI in dissertation writing: L2 doctoral students' self-reported use, AI-giarism, and perceived training needs. *Journal of English for Academic Purposes*, 78, Article 101570. <https://doi.org/10.1016/j.jeap.2025.101570>
- Humphrey, S. (2017). *Academic literacies in the middle years: A framework for enhancing teacher knowledge and student achievement*. Routledge. <https://doi.org/10.4324/9781315625584>
- Humphrey, S., & Feez, S. (2016). Direct instruction fit for purpose: Applying a metalinguistic toolkit to enhance creative writing in the early secondary years. *Australian Journal of Language and Literacy*, 39(3), 207–219. <https://doi.org/10.1007/BF03651974>

- Humphrey, S., & Macnaught, L. (2016a). Functional language instruction and the writing growth of English language learners in the middle years. *Tesol Quarterly*. <https://doi.org/10.1002/tesq.247>
- Humphrey, S., & Macnaught, L. (2016b). Developing teachers' professional knowledge of language for discipline literacy instruction. In H. de Silva Joyce (Ed.), *Language at work in social contexts: Analysing language use in work, educational, medical and museum contexts* (pp. 68–87). Cambridge Scholars Publishing.
- Humphrey, S., Martin, J. R., Dreyfus, S., & Mahboob, A. (2010). The 3x3: Setting up a linguistic toolkit for teaching academic writing. In A. Mahboob, & N. Knight (Eds.), *Applicable linguistics: Texts, contexts, and meanings* (pp. 185–199). Continuum.
- Ip, J. (2025). Confident AI. LLM-as-a-judge simply explained: The complete guide to run LLM evals at scale. <https://www.confident-ai.com/blog/why-llm-as-a-judge-is-the-best-llm-evaluation-method>.
- Jiang, F., & Hyland, K. (2025). Metadiscursive nouns in academic argument: ChatGPT vs student practices. *Journal of English for Academic Purposes*, 75, Article 101514. <https://doi.org/10.1016/j.jeap.2025.101514>
- Kalai, A. T., Nachum, O., Vempala, S. S., & Zhang, E. (2025). *Why language models hallucinate* (arXiv:2509.04664). arXiv. <https://doi.org/10.48550/arXiv.2509.04664>
- Kalai, A., Vempala, S., Nachum, O., Zhang, E., Robinson, D., Jain, S., Mitchell, E., Beutel, A., & Heidecke, J. (2025). *Why language models hallucinate*. OpenAI. <https://openai.com/index/why-language-models-hallucinate/>.
- Kaufer, D., & Ishizaki, S. (2025). *Future of writing and writing instruction: In the disciplines and professions [White paper]*. Carnegie Mellon University. <https://www.cmu.edu/dietrich/english/research-and-publications/myprose.html>.
- Kerman, N. T., Noroozi, O., Banihashem, S. K., Karami, M., & Biemans, H. J. A. (2024). Online peer feedback patterns of success and failure in argumentative essay writing. *Interactive Learning Environments*, 32(2), 614–626. <https://doi.org/10.1080/10494820.2022.2093914>
- Liang, W., Zhang, Y., Cao, H., Wang, B., Ding, D., Yang, X., Vodrahalli, K., He, S., Smith, D., Yin, Y., McFarland, D., & Zou, J. (2023). Can large language models provide useful feedback on research papers? A large-scale empirical analysis (arXiv:2310.01783). arXiv. <https://doi.org/10.48550/arXiv.2310.01783>
- Lin, S., & Crosthwaite, P. (2024). The grass is not always greener: Teacher vs. GPT-assisted written corrective feedback. *System*, 127, Article 103529. <https://doi.org/10.1016/j.system.2024.103529>
- Macnaught, L. (2018). Multimodal metalanguage. In H. de Silva Joyce, & S. Feez (Eds.), *Multimodality across classrooms: Learning about and through different modalities* (pp. 144–160). Cambridge Scholars Publishing.
- Macnaught, L. (2024). *Writing with students: New perspectives on collaborative writing in EAP contexts*. Bloomsbury Academic.
- Macnaught, L., Bassett, M., van der Ham, V., Milne, J., & Jenkin, C. (2024). Sustainable embedded academic literacy development: The gradual handover of literacy teaching. *Teaching in Higher Education*, 29(4), 1004–1022. <https://doi.org/10.1080/13562517.2022.2048369>
- Macnaught, L., & Hunter, K. (2024). *Reimagining research writing teaching and learning through AI assistants*. Cogniti mini-symposium. <https://www.youtube.com/watch?v=aqPSyZK5GM>.
- Macnaught, L., Winnington, R., & Macdiarmid, R. (2026). Preparing graduate entry nursing students to successfully complete research projects: A scoping review. *ESP Today*, 14(1), 89–114. <https://doi.org/10.18485/esptoday.2026.14.1.5>
- Martin, J. R. (2008). Tenderness: Realisation and instantiation in a Botswana town. In N. Norgaard (Ed.), *Systemic functional linguistics in use: Odense working papers in language and communication* (Vol. 29, pp. 30–58).
- Martin, J. R. (2010). Semantic variation - Modelling realisation, instantiation and individuation in social semiosis. In M. Bednarek, & J. R. Martin (Eds.), *New discourse on language* (pp. 1–34). Continuum.
- Martin, J. R. (2013). Embedded literacy: Knowledge as meaning. *Linguistics and Education*, 24, 23–37. <https://doi.org/10.1016/j.linged.2012.11.006>
- Martin, J. R. (2016). One of three traditions: Genre, functional linguistics, and the “Sydney School.” In N. Artemeva, & A. Freeman (Eds.), *Genre around the globe: Beyond the three traditions* (pp. 31–77). Trafford.
- Martin, J. R., & Matthiessen, C. (1991). Systemic typology and topology. In F. Christie (Ed.), *Literacy in social processes: Papers from the inaugural Australian systemic linguistics conference* (pp. 345–383).
- Martin, J. R., & Rose, D. (2007). *Working with discourse: Meaning beyond the clause* (2nd ed.) Continuum.
- Martin, J. R., & Rose, D. (2008). *Genre relations*. Equinox.
- Martin, J. R., & Wang, P. (2026). A practical guide to system networks: Modelling paradigmatic relations for language description. In *Systemic functional linguistics*. Bloomsbury Academic.
- Martin, J. R., & White, P. (2005). *The language of evaluation: Appraisal in English*. Palgrave Macmillan.
- Maton, K. (2014). *Knowledge and knowers: Towards a realist sociology of education*. Routledge.
- Matthiessen, C. M. I. M. (2020). Trinocular views of register: Approaching register trinocularly. *Language, Context and Text. The Social Semiotics Forum*, 2(1), 3–21. <https://doi.org/10.1075/langct.00019.mat>
- Matthiessen, C. (2023). *System in systemic functional linguistics: A system-based theory of language*. Equinox.
- Matthiessen, C. M. I. M., Wang, B., Ma, Y., & Mwinlaaru, I. N. (2022). *Systemic functional insights on language and linguistics*. Springer. https://doi.org/10.1007/978-981-16-8713-6_6
- Mitchell, T. D., & Pessoa, S. (2021). Using the 3x3 toolkit to support argumentative writing across disciplines. In A. Hirvela, & D. Belcher (Eds.), *Argumentative writing in a second language: Perspectives on research and pedagogy* (pp. 48–63). The University of Michigan Press.
- Moldoveanu, M. C., & Siemens, G. (2025). Interactionism: Re-designing higher learning for the large language agent era (arXiv:2501.00867). arXiv. <https://doi.org/10.48550/arXiv.2501.00867>
- Monbec, L. (2020). Systemic functional linguistics for the EGAP module: Revisiting the common core. *Journal of English for Academic Purposes*, 43, Article 100794. <https://doi.org/10.1016/j.jeap.2019.100794>
- Pessoa, S., Mitchell, T. D., & Gómez-Laich, M. P. (2024). *Analysis and argument in first-year writing and beyond: A functional perspective*. University of Michigan Press.
- Pessoa, S., Mitchell, T. D., & Miller, R. T. (2018). Scaffolding the argument genre in a multilingual university history classroom: Tracking the writing development of novice and experienced writers. *English for Specific Purposes*, 50, 81–96. <https://doi.org/10.1016/j.esp.2017.12.002>
- Rajagopalan, U. M., & Jie, S. (2016). Adopting a SFL-oriented approach for evaluating genre-based academic abstracts of EST undergraduate students in Japan. 2016 international conference on advanced mechatronic systems (ICAMechS) (pp. 510–515). <https://doi.org/10.1109/ICAMechS.2016.7813501>
- Rose, D. (2020). Building a pedagogic metalanguage I: Curriculum genres. In J. R. Martin, K. Maton, & Y. J. Doran (Eds.), *Accessing academic discourse: Systemic functional linguistics and legitimation code theory* (pp. 268–302). Routledge.
- Rose, D., & Martin, J. R. (2012). *Learning to write, reading to learn: Genre, knowledge and pedagogy in the Sydney school*. Equinox.
- Rothery, J. (1996). Making changes: Developing an educational linguistics. In R. Hasan, & G. Williams (Eds.), *Literacy and society* (pp. 86–123). Longman.
- Szenes, E. (2022). Re-visiting the modelling of “big texts” in SFL: The role of embedding in construing “depth” in long business reports. In D. Caldwell, J. Knox, & J. R. Martin (Eds.), *Applicable linguistics and social semiotics: Developing theory from practice* (pp. 355–381). Bloomsbury.
- Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education*, 76(3), 467–481. <https://doi.org/10.1007/s10734-017-0220-3>
- University of Sydney. (n.d.) Cogniti.ai [Computer software]. <https://cogniti.ai/>.
- University of Sydney. (2024). *How does Cogniti work?* <https://cogniti.ai/docs/how-does-cogniti-work/>.
- University of Sydney. How is my data protected?. <https://cogniti.ai/docs/how-is-my-data-protected/>.
- van der Veer, R. (2008). Multiple readings of Vygotsky. In B. van Oers, W. Wardekker, E. Elbers, & R. van der Veer (Eds.), *The transformation of learning: Advances in cultural-historical activity theory* (pp. 20–27). Cambridge University Press.
- Vygotsky, L. S. (1978). *Mind in society*. Harvard University Press.
- Wang, D. (2024). Teacher-versus AI-generated (Poe application) corrective feedback and language learners' writing anxiety, complexity, fluency, and accuracy. *International Review of Research in Open and Distance Learning*, 25(3), 37–56. <https://doi.org/10.19173/irrodl.v25i3.7646>
- Wiesinger, J., Marlow, P., & Vuskovic, V. (2025). *Agents [White paper]*. Google. https://www.kaggle.com/whitepaper-agents?utm_source=SAIL&utm_medium=email&utm_campaign=sail-agents-agents-agents.

- Wiley & Sons Inc. (n.d.). Journal of clinical nursing author guidelines. Wiley Online Library. <https://onlinelibrary.wiley.com/page/journal/13652702/homepage/forauthors.html#3>.
- Xu, Y., Polio, C., & Pfau, A. (2024). Optimizing AI for assessing L2 writing accuracy: An exploration of temperatures and prompts. In C. Chapelle, G. Beckett, & J. Ranalli (Eds.), *Exploring AI in applied linguistics* (pp. 151–174). Iowa State University Digital Press. <https://doi.org/10.31274/isudp.2024.154.10>.
- Zapata, G. C., Cope, B., Kalantzis, M., & Sears Smith, D. (2025). Calibrated generative AI as meta-reviewer: A systemic functional linguistics discourse analysis of reviews of peer reviews (arXiv:2509.15035). *arXiv*. <https://doi.org/10.48550/arXiv.2509.15035>
- Zapata, G., Tzirides, A., & Cope, B. (2025). *A systemic functional linguistics discourse analysis of learner-centered, generative AI feedback in higher education*. OSF. https://doi.org/10.35542/osf.io/nuhq2_v1
- Zappavigna, M. (2019). Language and social media: Enacting identity through ambient affiliation. In G. Thompson, W. Bowcher, L. Fontaine, & J. Y. Liang (Eds.), *The Cambridge handbook of systemic function linguistics* (pp. 714–737). Cambridge University Press.
- Zou, M., Kong, D., & Lee, I. (2025). Doctoral student's strategy use in GAI chatbot-assisted L2 writing: An activity theory perspective. *Journal of English for Academic Purposes*, 76, Article 101521. <https://doi.org/10.1016/j.jeap.2025.101521>

Lucy Macnaught, PhD, is a Senior Lecturer/Learning Advisor at Auckland University of Technology. Her book, *Writing with Students: New perspectives on collaborative writing in EAP contexts*, was shortlisted for the 2025 M.A.K. Halliday Book Prize. It illuminates how metalanguage and the organisation of classroom talk enables students to critique and justify their choices and teachers to guide but not provide wording. Current research interests include GenAI for writing development and embedding academic literacy in coursework and research programs.