REVIEW OF ARTIFICIAL INTELLIGENCE IN EDUCATION

DOI: https://doi.org/10.37497/rev.artif.intell.educ.v7ii.1395

RESEARH ARTICLE

DESIGNING WITH NOTEBOOKLM: AN AUTOETHNOGRAPHIC DISCOURSE ANALYSIS OF HUMAN-AI CURRICULUM DESIGN

Desenvolvimento Curricular com o Notebooklm: Uma Análise Autoetnográfica do Discurso na Colaboração entre Ser Humano e Inteligência Artificial

Gabriela C. Zapata
University of Nottingham (United Kingdom).
Email: Gabriela.Zapata@nottingham.ac.uk

ABSTRACT | Objective: This study examines how the linguistic choices made by a human curriculum designer and Google's NotebookLM constructed their roles and relationship during their collaborative curriculum design, and what these patterns revealed about the distribution and mobilization of human and technological resources. **Methods**: This work adopts an autoethnographic design and analyzes a two-month written interaction between an experienced curriculum designer and NotebookLM during the redesign of an undergraduate education class in the UK. The corpus comprised 367 interactional turns and 50,832 words. The interaction was analyzed discursively through Systemic Functional Linguistics and Appraisal Theory. Brown's (2011) *Design Capacity for Enactment* framework was subsequently employed to interpret the 48 curriculum-design episodes identified in the corpus. **Results:** The interaction was collaborative and increasingly personalized, but asymmetrical in design authority. NotebookLM contributed extensive textual production, elaboration, source synthesis, and strongly affirmative language, while the Human Designer expressed pedagogical intentions, evaluated proposals, controlled sequencing, and determined what should be accepted, modified, or rejected. The design episodes showed that adaptation was the most common form of resource appropriation, while instances of rejection provided additional evidence of human design authority. **Discussion:** The findings suggest that Generative AI can become a substantive curriculum-design resource without displacing educators' role as designers. NotebookLM expanded the range of candidate curricular representations, but their relevance depended on human evaluation, contextualization, and transformation. Linguistic alignment and extensive AI production did not necessarily correspond with design authority or successful implementation, reinforcing the need for continued professional judgment in Generative AI-supported curriculum design.

Keywords | Curriculum design, Higher education, NotebookLM, Human-AI collaboration

Qualis B1 CAPES 2021-2024

Indexed in ANPAD SPELL ANPAD ANPAD

Received: 17 August.2026
Revised: 18 August. 2026
Accepted: 20 August 2025
e-ISSN: 2965-4688
Guest Paper

Corresponding Author:
Gabriela C. Zapata – Email: Gabriela.Zapata@nottingham.ac.uk

How to cite this article: Zapata, G. C. (2026). Designing with Notebooklm: An Autoethnographic Discourse Analysis of Human-AI Curriculum Design. *Review of Artificial Intelligence in Education*, 7(i), e01395. https://doi.org/10.37497/rev.artif.intell.educ.v7ii.1395

RESUMO | Objetivo: Este estudo examina como as escolhas linguísticas realizadas por uma especialista em desenvolvimento curricular e pelo Google NotebookLM contribuíram para a construção de seus respectivos papéis e da relação estabelecida durante um processo colaborativo de desenvolvimento curricular. Além disso, investiga o que esses padrões linguísticos revelam sobre a distribuição e a mobilização de recursos humanos e tecnológicos ao longo desse processo. **Métodos:** A pesquisa adota uma abordagem autoetnográfica e analisa dois meses de interação escrita entre uma especialista experiente em desenvolvimento curricular e o NotebookLM durante a reformulação de uma disciplina de graduação na área de Educação, em uma universidade do Reino Unido. O corpus foi composto por 367 turnos de interação e 50.832 palavras. A análise discursiva foi conduzida com base na Linguística Sistêmico-Funcional e na Teoria da Avaliatividade. Posteriormente, o framework *Design Capacity for Enactment*, proposto por Brown (2011), foi utilizado para interpretar os 48 episódios de desenvolvimento curricular identificados no corpus. **Resultados:** A interação apresentou caráter colaborativo e tornou-se progressivamente mais personalizada, embora tenha permanecido assimétrica quanto à autoridade sobre as decisões curriculares. O NotebookLM contribuiu de forma significativa para a produção e elaboração de textos, a síntese de fontes e o emprego de linguagem fortemente afirmativa. A especialista humana, por sua vez, definiu as intenções pedagógicas, avaliou as propostas apresentadas, orientou a sequência do processo e determinou quais contribuições deveriam ser aceitas, modificadas ou rejeitadas. A adaptação constituiu a forma mais recorrente de apropriação dos recursos gerados, enquanto os episódios de rejeição forneceram evidências adicionais da predominância humana na tomada de decisões curriculares. **Discussão:** Os resultados indicam que a Inteligência Artificial Generativa pode constituir um recurso relevante para o desenvolvimento curricular sem substituir o papel dos educadores na condução e na tomada de decisões pedagógicas. O NotebookLM ampliou o conjunto de possibilidades e propostas curriculares disponíveis, mas a pertinência dessas contribuições dependeu da avaliação, contextualização e transformação realizadas pela profissional humana. O alinhamento linguístico e a elevada produção textual da IA não corresponderam necessariamente à autoridade decisória nem à implementação efetiva das propostas, reforçando a importância da manutenção do julgamento profissional em processos de desenvolvimento curricular apoiados por Inteligência Artificial Generativa.

Palavras-chave | Design curricular; Educação superior; NotebookLM; Colaboração humano-IA.

INTRODUCTION

In the last few years, Generative Artificial Intelligence (GenAI) has become part of the higher education landscape, influencing how students and instructors approach teaching, learning, assessment, and academic work (Chan & Hu, 2023; Lee et al., 2024; Tzirides et al., 2024; Zapata et al., 2024). Much of the attention devoted to these technologies has centered on the opportunities and problems associated with their incorporation into university education (Belkina et al., 2025; Cope & Kalantzis, 2024; Zapata et al., 2025a). As this work has developed, the discussion has also moved beyond whether GenAI should be used to how it can be incorporated into educational practices while maintaining human judgment, disciplinary knowledge, and critical engagement with generated content (Francis et al., 2025; UNESCO, 2025; Zapata et al., 2026).

Language is particularly relevant for understanding the role GenAI can play in education. Large language models' (LLMs) discourse can express evaluation, establish relationships, organize interaction, represent agency, and position ideas or courses of action as more or less desirable, certain, or appropriate. Recent work (e.g., Beguš et al., 2025; Donnellan, 2025; Roine et al., 2026) has begun to investigate LLM language and how it might generally influence human practices. In education, research using Systemic Functional Linguistics (SFL; Eggins, 1994; Halliday & Matthiessen, 2014) and Appraisal Theory (Martin & Rose, 2007; Martin & White, 2005) has shown the value of

examining how GenAI discourse constructs meaning and interpersonal relationships in formative assessment (Zapata et al., 2025b, 2025c).

Nevertheless, much less is known about language when GenAI is used in extended professional interaction within educational contexts. Curriculum design offers one such context. Although research on GenAI-supported course and curriculum design is beginning to emerge (e.g., Ullmann et al., 2024), considerably less attention has been devoted to the discourse through which a human educator and an LLM carry out this work. This article addresses this gap through an autoethnographic study of the interaction between an experienced university curriculum designer and Google's *NotebookLM* (now *Gemini Notebook*) during the redesign of an undergraduate education class in the UK. The analysis employs SFL and Appraisal Theory to examine how linguistic choices represent the design activity, construct the participants' roles and relationship, and organize their interaction. Brown's (2011) *Design Capacity for Enactment* framework is subsequently used to interpret what these linguistic findings reveal about the distribution and mobilization of human and technological resources during curriculum design.

This work is organized as follows. The first two sections review existing work on curriculum design, GenAI-supported design, NotebookLM, and language in human-AI interaction, and introduce the study's research questions. This is followed by a description of the research design, data collection and analysis procedures, and the analytical and theoretical frameworks. The remaining parts present and discuss the results, consider their practical implications, address the limitations of the study, and conclude this work.

LITERATURE REVIEW

Curriculum Design as Professional Practice

Curriculum design can be characterized as a complex process through which educators make decisions about learning outcomes, assessment, instructional content, tasks, resources, and the sequencing of learning experiences. These elements cannot be considered independently of the students for whom a curriculum is developed or the educational context in which it will be implemented (Kalantzis et al., 2016, in press). Curriculum design consequently requires educators to determine what students should learn, how learning can be supported and assessed, what resources are appropriate, and how curricular components can be organized coherently (Kalantzis et al., in press; Wiggins & McTighe, 1998).

Brown's (2011) conceptualization of teaching as design offers a useful perspective on this process. This scholar posits that educators work with different kinds of resources, but these resources do not determine what happens instructionally. That is, teachers interpret and mobilize them in relation to their subject-matter knowledge, pedagogical content knowledge, goals, beliefs, and understanding of the instructional context. Curriculum materials can provide representations of concepts and tasks and can make particular forms of action possible, but they can also present constraints that instructors must recognize and address (Kalantzis et al., in press). The relationship between educator and curricular resource is therefore dynamic: materials can be used substantially

as provided, modified to suit a particular instructional purpose, or departed from when the teacher's own resources offer a more appropriate course of action (Brown, 2011).

This view of curriculum work places considerable responsibility on educators as designers. Decisions about content, sequencing, assessment, appropriateness, and feasibility depend on forms of knowledge that are situated in a specific discipline, institution, course, and student population (Kalantzis et al., 2016, in press). Broader curriculum studies and instructional design literature offer a similar perspective. For instance, Biggs and Tang (2011) emphasize the need for coherence between learning objectives, teaching activities, and assessment tasks, highlighting the interpretive work required of teachers in translating curriculum intentions into enacted practice. Research on curriculum materials further shows that instructors do not simply implement designed resources but actively interpret, adapt, and sometimes transform them in response to contextual constraints and student needs (Remillard, 2005). In higher education contexts, curriculum design is likewise understood as iterative and responsive, involving continuous evaluation of whether learning activities, resources, and assessments adequately support diverse student cohorts and institutional expectations (Priestley et al., 2015; Zapata, 2018, 2023). Within this tradition, curriculum design is therefore not a linear process of producing instructional materials but an ongoing reflective activity involving judgement, adjustment, and refinement, which can be considered especially significant when some of the resources available to today's educators are generated dynamically by GenAI.

Generative AI and Curriculum Design

Even though research on the use of GenAI for curriculum and course design is still developing, existing studies have begun to examine its contribution to several aspects of this work. For instance, Ullmann et al. (2024) investigated GPT-4 for generating course outlines and content, drafting assessment instructions, and mapping learning outcomes to external benchmark statements. AI-generated curriculum artefacts were evaluated by experienced higher education practitioners against existing human exemplars and disciplinary expectations for accuracy, pedagogical coherence, and alignment. The results showed that GPT-4 could rapidly generate coherent and plausible drafts, and its mapping of learning outcomes was broadly consistent with expert-produced versions. Nonetheless, the outputs also exhibited insufficient depth, limited contextual specificity, and occasional inaccuracies. The authors thus posited that GenAI could be useful for early ideation and drafting, but educators' responsibility for curation, validation, contextual adaptation, and curricular alignment was still crucial.

Rutecka et al. (2025) reported similar findings in their comparison of several LLMs with existing university curricula in Economics and Management. Although the study showed that newer models performed better when designing materials in the disciplines examined, none of them could independently generate resources comparable to those developed by university experts. For example, limitations included incomplete curricular structures, inconsistent or illogical sequencing of content and instructional time, and fabricated or unverifiable references. The scholars also highlighted that increasing prompt specificity did not reliably improve quality and, in some cases, produced even less appropriate or coherent results. These findings suggest that LLMs are more

suitable for early-stage ideation and drafting, but expert intervention is needed to ensure accuracy, coherence, and institutional and disciplinary alignment.

Other work has examined collaborative, human-AI approaches to curriculum development. For instance, Ahmad and Schmid (2025) developed a GenAI-based prototype to automate curriculum templates, support learning-outcome alignment, and assist with course redesign. The system incorporated Retrieval-Augmented Generation (RAG) and was evaluated through expert review. Like the findings reported by Ullmann et al. (2024) and Rutecka et al. (2025), the results of this study pointed to possible gains in efficiency and consistency, while also emphasizing the need for human judgment to evaluate the quality, relevance, inclusivity, and pedagogical alignment of AI-generated outputs.

Even though these studies have offered us information on the role that GenAI might play in curriculum development, highlighting both affordances and limitations, to the best of the author's knowledge, no works have explored human-AI interaction in curriculum co-design. This work seeks to bridge this gap.

NotebookLM in Educational and Academic Work

This study's human-AI collaborative curriculum design involved the use of Google's NotebookLM, a generative AI platform that allows users to upload and work with a selected set of source materials to generate responses, summaries, explanations, and other outputs grounded in the provided corpus. This source-grounded architecture offers educators some control over the materials informing the system's responses and can support alignment with course content, disciplinary resources, and instructional goals.

Existing work has begun to examine NotebookLM in different teaching and learning contexts. For example, Tufino (2025) configured it as a source-grounded physics tutor that provided Socratic guidance through teacher-selected materials and pedagogical instructions. The system generated questions, explanations, study materials, and other instructional resources while maintaining links with the selected sources. Nevertheless, inaccuracies remained possible, particularly when curated solutions were unavailable, leading Tufino to emphasize the need for educator oversight. Reyna's (2025) critical overview of the tool identified a broader range of pedagogical applications, including lesson planning, courseware development, curriculum review, personalized study support, and teaching delivery. However, like Tufino, this scholar reported concerns about NotebookLM's reliability, as well as issues of privacy, copyright, and access.

Other studies have focused on NotebookLM's capacity to generate instructional audio materials. For instance, Dihan et al. (2025) examined its Audio Overview feature in ophthalmology and described its potential for transforming uploaded materials into conversational podcasts for trainee and patient education. The authors emphasized the accessibility and convenience of source-grounded audio, but, like Tufino (2025) and Reyna (2025), identified limitations related to linguistic accuracy, privacy, and the system's dependence on uploaded materials.

Overall, these studies have highlighted several potential instructional uses for NotebookLM, including source-grounded tutoring, the generation of teaching materials, and the creation of

multimodal learning resources. Nevertheless, they have also consistently emphasized the need for educator evaluation and oversight when working with it. What remains unexplored is how instructors engage with NotebookLM during sustained curriculum design, particularly as they assess generated proposals, refine or reject them, and determine which resources should become part of instruction.

Language and Human-AI Interaction

Due to the linguistic nature of LLMs, another relevant aspect of GenAI use in education is the discursive dimension of human-AI interaction. Recent research in this area has already identified recurring linguistic patterns in how these technologies respond to and align with users. For example, Sharma et al. (2025) investigated sycophancy in several LLMs and described an overall tendency to adapt their responses to users' stated beliefs or preferences, even when this reduced factual accuracy. For their part, Roine et al. (2026) examined a related phenomenon they defined as *affirmative narration*, showing how LLM chatbots repeatedly validated and extended users' contributions through strong affirmation and other narrative strategies. Though such alignment, the scholars posited, could support interaction, it could also complicate judgments about whether linguistic agreement corresponded to an accurate understanding of the user's request.

In educational contexts, SFL and Appraisal Theory discourse analysis have been applied to examine LLM language when enacting pedagogical functions. For instance, Zapata et al. (2025b) compared non-calibrated and calibrated GenAI formative feedback in graduate education and identified differences in how AI's language represented agency, expressed evaluation, organized information, and constructed relationships with learners. Specifically, the analysis showed that the calibrated system more consistently produced patterns associated with actionable and academically grounded feedback, although limitations remained in reproducing the contextual and affective dimensions of human commentary. In a subsequent related study, Zapata and colleagues (Zapata et al., 2025c) examined GenAI-generated reviews of graduate students' peer reviews and found that the calibrated system produced feedback that combined clear guidance with a supportive stance, remained aligned with assessment criteria, and organized information in ways that made suggestions easier to follow and act upon. The analysis also revealed that GenAI's language helped position students as active participants in the feedback process while balancing positive evaluation with constructive critique. These studies suggest that examining GenAI language can uncover aspects of educational interaction that evaluations of output accuracy or usefulness alone may not capture.

While these studies indeed demonstrate the value of examining GenAI discourse in educational contexts, the linguistic construction of an extended human-GenAI relationship during professional educational design has not, to the best of the author's knowledge, been explored. The purpose of this work is to address this gap. Specifically, the present study aims to analyze the discourse produced during an extended curriculum-design interaction between an experienced human designer and NotebookLM. The analysis first unveils how linguistic choices constructed the participants' roles and relationship. We then interpret the identified linguistic patterns through Brown's (2011) *Design*

Capacity for Enactment framework to determine how human and technological resources were mobilized throughout the design process. In the sections that follow, we introduce the research questions and discuss the study's methods.

RESEARCH QUESTIONS

This work seeks to answer the following research questions:

1. How did the linguistic choices made by the human designer and NotebookLM construct their roles and relationship during the curriculum design process?
2. What did the linguistic patterns identified in the interaction reveal, when interpreted through Brown's (2011) *Design Capacity for Enactment framework*, about the distribution and mobilization of human and technological resources during curriculum design?

METHODOLOGY

Research Design

This study adopts an autoethnographic research design. In this type of design, the researcher's own experience is a source of data (Adams et al., 2017; Ellis et al., 2011), which is particularly suited to contexts where they occupy an insider position and are directly involved in the activity being examined (Rambo & Ellis, 2020). In the present work, the researcher was both the curriculum designer and one of the participants in the interaction with NotebookLM. Autoethnography therefore made it possible to investigate the design process from within, while also subjecting the interaction to systematic analysis. Importantly, the objective was not to treat the researcher's experience as representative of curriculum designers more generally, but to unveil how one extended human-GenAI design interaction unfolded and what it could reveal about the roles assumed by the human designer and the technological resource.

This work can also be situated within what Atay (2020) defines as *cyber or digital autoethnography*, an approach that extends autoethnographic inquiry to experiences that occur within digitally mediated environments (see also Mao et al., 2024). This type of autoethnography recognizes that digital spaces shape the forms of interaction and participation available to users. In this study, the digital, curriculum design process took place through a sustained two-month, digital written interaction between the human designer and NotebookLM. The resulting transcript documented the experience and made it possible to investigate how the researcher's professional practice as a curriculum designer was mediated through her collaboration with a GenAI system. This interaction was examined through the systematic discourse analysis described in the sections that follow.

Study Context and Data Collection

The study was conducted as part of the curriculum redesign of an undergraduate class on the datafication of education offered at a UK university. The researcher, a scholar with almost 30 years of teaching and curriculum-development experience, had previously taught the class and sought to redesign it for the 2026–2027 academic year. Her main goal was to incorporate more current developments and to develop instruction and assessment tasks that would increase students' engagement with course content, which some had previously characterized as "dry and uninspiring." The researcher also had experience investigating GenAI in higher education, particularly in connection with formative assessment, AI literacy, and classroom practices.

NotebookLM was used throughout the redesign process, which took place over two months, with bi-weekly co-design sessions, each lasting an average of five hours. A backward-design approach (Wiggins & McTighe, 1998) guided the work, beginning with the revision of the class's learning outcomes and assessment and subsequently moving to the organization of weekly topics, selection of resources, and development of pedagogical materials and interventions. The tool contained an evolving collection of sources selected by the researcher, including academic literature, multimodal resources, and teaching materials from the previous iteration of the course. During the interactions, the researcher also selected particular sources depending on the curricular component under development. NotebookLM was employed for different aspects of the process, including source organization, topic development, assessment and task design, and the drafting and revision of instructional resources.

The study's data consisted of the complete written, dialogic interaction between the researcher, identified in the corpus as the Human Designer, and NotebookLM (the pedagogical outputs generated were not part of the transcripts analyzed). The final corpus comprised 367 labelled interactional turns, including 184 Human Designer turns and 183 NotebookLM turns, and it contained 50,832 words, of which 13,407 were produced by the Human Designer and 37,425 by NotebookLM. The corpus encompassed the initial negotiation of participant roles and development of learning outcomes, assessment design, source selection, curriculum and materials development, technical problems and their resolution, metacognitive reflections on the design process, and the final exchanges between the Human Designer and NotebookLM. An excerpt of the interaction is offered in the Appendix.

Analytical and Interpretive Frameworks

Systemic Functional Linguistics and Appraisal Theory

The linguistic analysis in this study drew on Systemic Functional Linguistics (SFL), a model of language developed by Halliday (1975). SFL conceptualizes language as a social semiotic system; i.e., a resource for making meaning that both reflects and shapes the contexts in which communication takes place (Halliday & Hasan, 1985; Halliday & Matthiessen, 2014). Specifically, the framework focuses on the linguistic choices speakers and writers make to simultaneously represent experience,

construct relationships, and organize discourse. These three dimensions are captured through the *ideational*, *interpersonal*, and *textual metafunctions*, respectively (Eggins, 1994; Halliday & Matthiessen, 2014). This approach was appropriate for the present work because the objective was not only to identify what the Human Designer and NotebookLM communicated, but also to investigate how their linguistic choices constructed their roles and relationship throughout the collaborative curriculum design interaction.

The *ideational metafunction* focuses on how experience is represented through language. One of its central systems is *transitivity*, which examines *processes*—the kinds of actions, thoughts, states, and acts of communication represented in discourse, who is involved in them, and under what circumstances they occur (Eggins, 1994; Halliday & Matthiessen, 2014). In the present study, this dimension made it possible to unveil how action, cognition, evaluation, and communication were distributed between the Human Designer and NotebookLM. The *interpersonal metafunction* focuses on how language constructs social relationships, including how speakers express stance (i.e., how a speaker or writer positions themselves toward an idea, claim, or interlocutor), authority, agreement, disagreement, and resistance. The *textual metafunction* explores how discourse is organized so that ideas connect and develop coherently, including what information is foregrounded, how participants and ideas are referred to, and how clauses and larger stretches of discourse are linked. These dimensions provided complementary ways of understanding how the two participants represented their design work, negotiated their relationship, and organized the extended interaction.

Appraisal Theory (Martin & Rose, 2007; Martin & White, 2005) was employed to explore interpersonal meaning in greater detail. The framework comprises three main systems: *Attitude*, which includes affect, judgment, and appreciation; *Engagement*, which examines how speakers and writers open or restrict space for alternative positions; and *Graduation*, which captures how evaluative meanings are intensified, softened, sharpened, or quantified. In the present analysis, these resources were used to investigate how the Human Designer and NotebookLM expressed evaluation, certainty, agreement, disagreement, and alignment, particularly when responding to proposals and possible corrections and moments of repair.

Design Capacity for Enactment

Brown's (2011) *Design Capacity for Enactment* (DCE) provided the interpretive framework for the second stage of the analysis. As previously discussed, Brown conceptualizes teaching as a design activity in which educators draw on both their own resources and curricular ones. Teacher resources include subject-matter and pedagogical content knowledge, as well as goals and beliefs. Curriculum resources can comprise representations of objects, tasks or procedures, and concepts or domain knowledge. Applied to curriculum design, the framework makes it possible to examine how educators draw on these resources when making decisions about what to include, how to organize it, and how available materials might support or constrain particular design goals.

Central to Brown's (2011) framework is *Pedagogical Design Capacity* (PDC), which refers to a teacher's ability to perceive and mobilize available resources in the development of instruction.

Brown distinguishes three forms of curriculum-resource appropriation: *adapting*, when the educator modifies what the resource provides; *improvising*, when the design activity originates principally in the educator's own resources; and *offloading*, when a resource carries a substantial part of the design work. These categories describe different ways in which curriculum design activity can be distributed between the educator and the resources available to them.

The present study extends DCE to a generative technological resource. That is, Brown's (2011) framework was developed primarily in relation to curriculum materials that provide relatively stable representations. However, NotebookLM can generate new representations in response to the designer's requests (Reyna, 2025; Tufino, 2025). Thus, the tool was treated as part of the curriculum-resource environment, while PDC remained associated with the Human Designer's capacity to perceive, evaluate, select, adapt, and mobilize what the system provided. This framework was used to interpret the linguistic findings in relation to how the study's curriculum design process was distributed between human and technological resources.

Data Analysis

All data were imported into MAXQDA 2022 (VERBI Software, 2021), which was used for coding, lexical searches, visualization, and qualitative memoing. The coding scheme was based on SFL metafunctional categories (Eggins, 1994; Halliday & Matthiessen, 2014) and Appraisal Theory's evaluative resources (Martin & Rose, 2007; Martin & White, 2005). The analysis began with the register variables of *field*, *tenor*, and *mode*, which describe, respectively, what an interaction is about, the roles and relationships between participants, and how communication is organized. The next step focused on ideational, interpersonal, and textual meanings.

At the ideational level, clauses were coded for *material processes* (actions and events), *relational processes* (states, attributes, and relationships), *mental processes* (thinking, feeling, wanting, or perceiving), and *verbal processes* (acts of saying or communicating). *Circumstantial meanings* were also coded to identify where, how, and why processes occurred, including location, manner/means, and cause/purpose. The interpersonal analysis examined Appraisal resources associated with *Attitude* (evaluation and affect), *Engagement* (how speakers open or restrict space for alternative positions), and *Graduation* (how evaluative meanings are strengthened, softened, sharpened, or quantified). It also included interactional features related to directives, proposals, alignment, resistance, acknowledgment, and repair. The textual analysis focused on *Theme-Rheme* organization, or how information is introduced and developed within clauses, as well as reference chains, lexical cohesion, conjunction, and rhetorical sequencing, which contribute to how ideas and participants are connected across the interaction. Spanish language use, emojis, and emoticons were also examined when they contributed to tenor, identity, affect, or interactional repair. Figure 1 summarizes the way in which SFL and Appraisal Theory were applied in the analysis of the interaction.

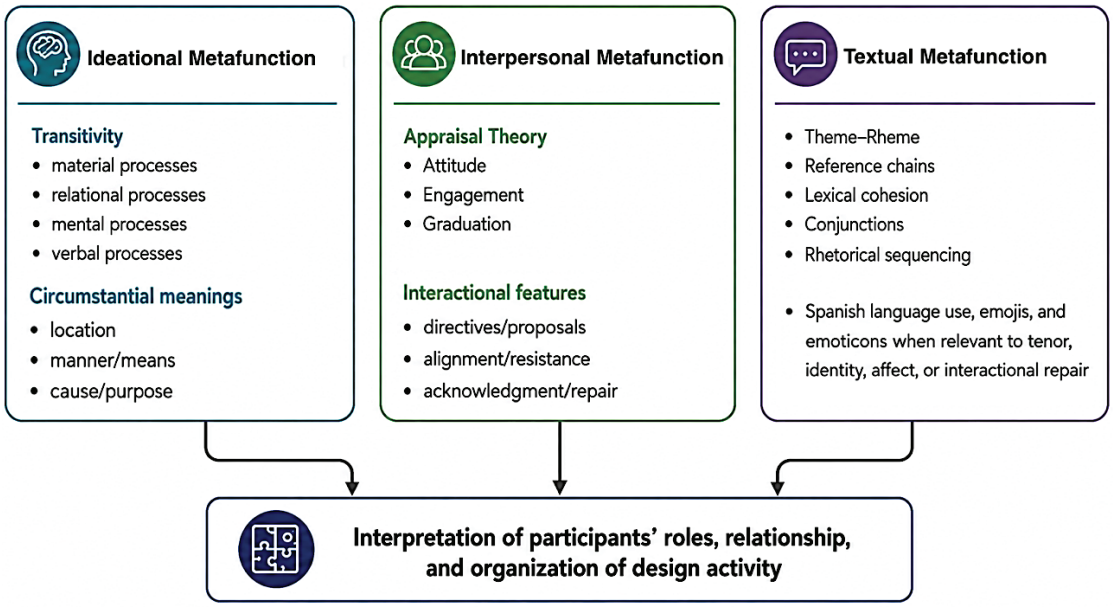

Figure 1. Analytical Framework for the SFL and Appraisal Theory Analysis

The complete corpus was coded separately for the Human Designer and NotebookLM to enable comparison of their linguistic choices. Depending on the category, coding was conducted at the phrase, clause, turn, or larger discourse level, and categories could co-occur when a linguistic unit realized more than one relevant function. MAXQDA's lexical search and coding functions were used to identify recurring markers, including second-person pronouns, modal verbs, and evaluative adjectives. Code-frequency tables supported the calculation of frequencies and proportional distributions, while qualitative memoing was used to annotate representative excerpts and record how recurrent linguistic patterns functioned within the interaction. The procedure therefore combined quantitative information about linguistic tendencies with qualitative analysis of how they contributed to the participants' roles, relationship, and design activity. The analysis was also validated manually.

A second stage of analysis interpreted the linguistic findings through Brown's (2011) *Design Capacity for Enactment* framework. The transcript was divided into 48 substantive design episodes, defined as bounded stretches of interaction organized around an identifiable curricular object, such as learning outcomes or a classroom activity. Social exchanges, metacognitive discussion that did not itself redesign curriculum, routine interface questions, and resource-management exchanges without a discernible curricular outcome were excluded.

Each episode was examined in terms of how NotebookLM-generated resources were used by the Human Designer. Brown's (2011) categories of *adapting*, *improvising*, and *offloading* were applied. A new category, *reject*, was created to account for cases in which an AI-generated proposal was abandoned or replaced. Also, offloading was treated provisionally because, though the transcript could document acceptance without visible modification, it was not possible to establish whether a resource would be ultimately enacted without later changes. The analysis also considered the human resources mobilized during each episode, including goals and beliefs, subject-matter knowledge, and

pedagogical content knowledge, as well as NotebookLM's affordances and constraints. This stage connected the linguistic patterns with the distribution and mobilization of human and technological resources during curriculum design.

RESULTS

The results of the discourse analysis of the Human Designer-NotebookLM interaction suggest that the relationship was collaborative and increasingly personalized, but it remained clearly asymmetrical in terms of curriculum design authority. While NotebookLM contributed extensive textual production, elaboration, and strong interpersonal alignment, the Human Designer directed the work, evaluated the system's proposals, introduced pedagogical and contextual knowledge, and determined what should be accepted, modified, or rejected. The SFL and Appraisal Theory analysis first examines how this relationship was constructed through ideational, interpersonal, and textual meanings, and the subsequent DCE exploration considers how these linguistic patterns reflected the distribution and mobilization of human and technological resources during curriculum design. The findings are presented separately in the sections that follow.

SFL Results

Ideational Metafunction

The transitivity analysis revealed a strong orientation toward action, particularly in NotebookLM's discourse (Table 1). For example, material processes occurred in 64% of NotebookLM's clauses compared with 52% of the Human Designer's. However, these processes were enacted differently. The Human Designer's material processes frequently governed the project through expressions such as "*We will first determine specific outcomes*," "*let's modify*," and "*I will work on your draft*." NotebookLM's were more commonly associated with producing and organizing curricular representations, including drafting outcomes, mapping assessment, restructuring content, and creating activities. Mental processes reinforced this distinction. Expressions such as "*I want*," "*I think*," and "*I wonder*" represented the Human Designer's pedagogical intentions and judgments, whereas NotebookLM's "*I agree*" or "*I understand*" more often expressed reception and alignment. This implies that the Human Designer appeared more frequently as the *Senser*, whose intentions guided subsequent action, and NotebookLM was commonly the *Actor* that transformed those intentions into candidate curricular products.

Relational processes were also frequent in both participants' discourse and contributed to evaluating ideas and defining their roles and relationship. Statements such as "*I am a professor*," "*I am your NotebookLM thought partner*," "*We work well together*," and the Human Designer's later description of NotebookLM as a "*cognitive and emotional prosthesis*" show how these roles were progressively negotiated throughout the interaction. Verbal processes were the least frequent of the four categories, although they occurred somewhat more often in NotebookLM's discourse. They appeared in proposals such as "*Would you like to refine…?*" or "*I suggest…*," allowing the tool to

advance possible next steps and influence the immediate pacing of the interaction. Nevertheless, the Human Designer retained authority over whether those proposals were accepted or redirected.

Table 1. Percentage of process types in the corpus

Process type	Human clauses	NotebookLM clauses
Material	52%	64%
Relational	38%	41%
Mental	34%	35%
Verbal	16%	21%

Note: Percentages represent the proportion of each participant's clauses in which the process type occurred. Because a clause could contain more than one coded process type, percentages do not total 100%.

Circumstantial meanings revealed another difference. NotebookLM used more circumstances of manner/means, which occurred in 19% of its clauses compared with 11% of the Human Designer's, often specifying how activities should be carried out. In contrast, the Human Designer used more explicit causal circumstances (7% compared with 3%), frequently providing contextual reasons for accepting or modifying proposals. For example, when NotebookLM suggested conventional homework, the Human Designer explained that "*UK students do not usually do homework because it doesn't count [towards their grades]*," introducing situated knowledge that altered the proposed design.

Interpersonal Metafunction

The interpersonal analysis showed that both participants evaluated the design process extensively, but NotebookLM maintained a substantially more positive orientation. For example, 82% of NotebookLM's *Attitude* was positive, compared to the Human Designer's 69%. Indeed, the tool frequently employed strongly positive expressions such as "*brilliant*," "*fantastic*," "*excellent*," and "*You're completely right*," whereas the Human Designer's evaluations were generally more calibrated, as in "*helpful start*," "*Very nice. But…*," and "*Not exactly what I had in mind*."

Engagement also revealed some differences. For instance, the Human Designer used more entertain resources, such as "*I think*," "*maybe*," and "*I'm wondering*," which allowed ideas to remain provisional, as well as more disclaim resources, including "*Not yet*," "*That's not what I mean*," and "*Rely on the sources*," when rejecting or redirecting proposals. This pattern suggests that the Human Designer both kept possibilities open during the design process and marked when NotebookLM's responses did not meet the intended requirements. For its part, NotebookLM employed more proclaim resources, including "*Absolutely*," "*Exactly*," and "*You're right*," which constructed it as a highly affirming participant. Nevertheless, expressions of agreement were sometimes followed by outputs that reproduced the behavior being corrected. This indicates that linguistic alignment did not always correspond with operational alignment, and that the Human Designer remained responsible for determining whether NotebookLM had actually responded to the intended design direction.

Turn-level analysis offered further evidence for the asymmetry in interactional control. The Human Designer produced 107 directives or requests (58% of her turns) and 47 turns involving

control, resistance, or redirection (26%), while NotebookLM produced 112 design proposals or offers (61% of its turns) and 40 acknowledgments or concessions (22%). Of the 47 Human Designer resistance or redirection turns, 24 were immediately followed by an explicit concession, acknowledgment, or apology, while other corrections led directly to revised proposals. A recurrent interactional cycle therefore emerged: the Human Designer introduced a proposal or constraint; NotebookLM responded with affirmation and elaborated production; and the Human Designer then accepted, modified, or rejected the proposal, often resulting in concession and repair on the tool's part (Figure 2).

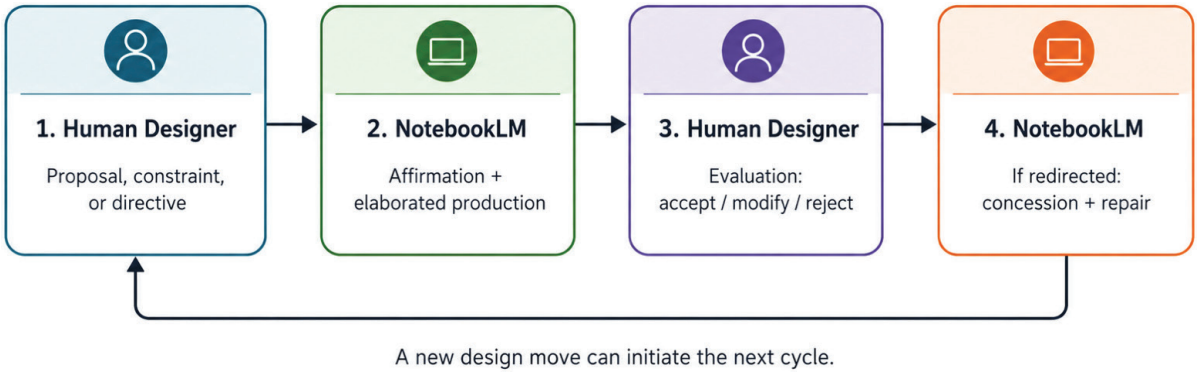

Figure 2. Recurrent Interactional Cycle in the Human Designer–NotebookLM Interaction

Textual Metafunction

The textual analysis revealed additional differences in how the participants organized the interaction. The Human Designer frequently began turns with interpersonal *Themes* such as "*OK*," "*Right*," and "*Not yet*," which occurred in 35% of her coded paragraph-initial clauses compared with 1% of NotebookLM's. These openings often evaluated the preceding proposal before introducing the next design move, as in "*Very nice. But something else regarding data literacy…*." In contrast, NotebookLM employed more marked topical Themes (14% compared with 5% for the Human Designer), including "*Based on our discussion*," "*At this stage*," and "*Now that we have…*," which structured progression between design phases. However, these types of expressions sometimes treated a decision as completed before the Human Designer considered it resolved, eliciting responses such as "*One step at a time…*." These patterns show that textual organization also contributed to the negotiation of when a design stage could be considered complete.

Reference and conjunctions provided further evidence of differences in ownership and design activity. First-person singular reference (*I/me/my*) occurred 46 times per 1,000 words in the Human Designer's discourse compared with 11 in NotebookLM's and appeared in expressions such as "*I want*," "*my [class]*," and "*my students*." NotebookLM used *you/your* at a rate of 32 occurrences per 1,000 words, frequently referring to "*your students*" or "*your vision*." Both participants also used *we/us/our* to construct shared activity through expressions such as "*we can refine*" and "*our outcomes*," although this collaborative reference did not eliminate the Human Designer's stronger linguistic ownership of the design. Conjunctive patterns reflected a related distinction: NotebookLM

frequently expanded proposals through serial addition, whereas the Human Designer used more contrast/concession and cause/consequence, as in "*That's good, but…*" and "*This would not work because…*," to connect decisions with constraints and previous choices.

Spanish, emojis, and emoticons contributed to the increasingly personalized nature of the interaction. Spanish appeared in 32 turns and was used mainly for greetings, encouragement, humor, and leave-taking after being introduced by the Human Designer, but most curricular work remained in English. The Human Designer also calibrated the use of Spanish when lexical or gender choices did not correspond with her linguistic identity. Paralinguistic resources followed a comparable pattern: NotebookLM used 62 emoji tokens in 39 turns, while the Human Designer employed 36 ASCII or lexical markers, including ":)" and "LOL." These resources supported affiliation, humor, and repair, as in "*You [NotebookLM] offered a prompt card, not another presentation! LOL!*," where humor accompanied a correction without removing its directive force. In other words, these features contributed to interpersonal closeness but did not alter the underlying distribution of design authority.

DCE Results

The second stage of the analysis examined the 48 design episodes through Brown's (2011) DCE framework (Figure 3). The predominance of *adaptation* was consistent with the linguistic findings. That is, NotebookLM repeatedly generated candidate representations that the Human Designer evaluated and transformed, a relationship she described through expressions such as "*a good start*," "*a starting point*," and "*I will modify/add/create as needed*." For example, NotebookLM produced an initial set of learning outcomes, but the Human Designer identified missing priorities. The Human Designer subsequently revised the outcomes, but she chose to keep some of NotebookLM's generated content. A comparable process occurred with presentations, activities, source mappings, and assessment resources. NotebookLM often provided rapid synthesis, restructuring, and procedural detail, while the Human Designer drew on subject-matter and pedagogical content knowledge, as well as beliefs and goals to determine what required modification. Situated knowledge (i.e., tailoring resources to the specific population the Human Designer had in mind) also shaped adaptation.

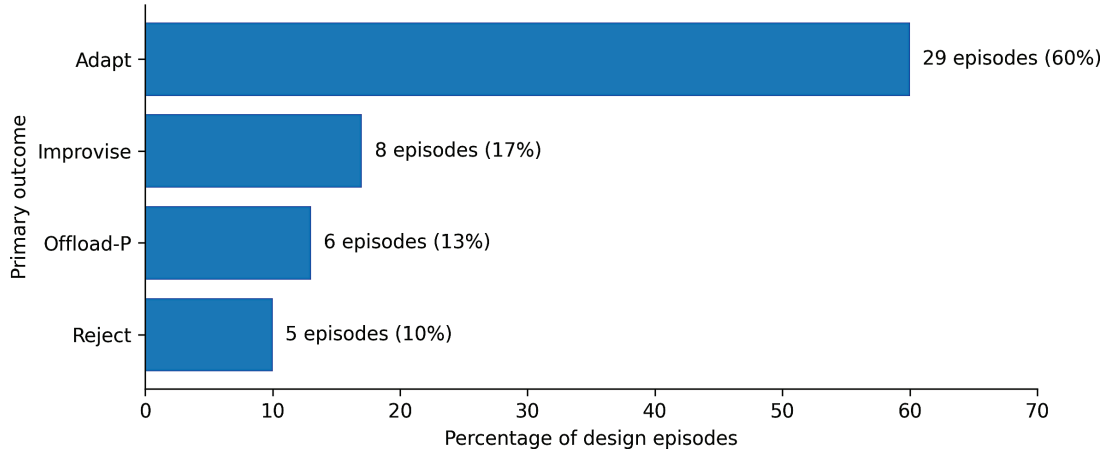

Figure 3. Distribution of DCE Outcomes in the Curriculum-Design Episodes Identified

Improvisation accounted for eight episodes in which the initial pedagogical direction originated primarily from the Human Designer and NotebookLM then helped elaborate or operationalize it. Although these episodes often contained substantial AI-generated text, the underlying pedagogical intent and criteria for success came from the Human Designer. *Rejection* occurred in five episodes after attempts at correction did not produce a usable resource. For instance, several versions of the formal assessment brief were generated before the Human Designer concluded, "*Unfortunately, I could not use your brief*," and replaced it with her own.

Only six episodes were classified as *provisional offloading*. In these cases, a NotebookLM-generated representation was accepted without further visible substantive modification in the recorded interaction. However, this classification does not indicate that the resource was ultimately enacted unchanged, particularly because the Human Designer stated elsewhere that she always checked and modified the materials produced by NotebookLM.

Overall, the DCE findings suggest that NotebookLM contributed extensive textual production, source synthesis, organization, procedural detail, and candidate curricular representations. For her part, the Human Designer supplied the pedagogical and disciplinary criteria, situated knowledge, evaluation, and final decisions about their use. Adaptation, improvisation, rejection, and provisional offloading thus made visible different ways in which these resources were mobilized, with PDC remaining associated with the Human Designer's capacity to evaluate and act on what NotebookLM provided.

DISCUSSION

This study's first research question examined how the linguistic choices made by the Human Designer and NotebookLM constructed their roles and relationship during curriculum design. The findings revealed a relationship that was collegial and increasingly personalized, but clearly asymmetrical in terms of design authority. NotebookLM occupied considerably more of the verbal space and was strongly associated with material processes, elaborated proposals, and textual resources that organized the interaction into stages of progress. Nevertheless, it was the Human Designer who expressed the pedagogical intentions and judgments that guided the work, established the course of action through directives, and accepted, modified, redirected, or rejected the tool's proposals. Reference patterns such as the use of first-person structures reinforced this distinction. The interaction was therefore human led despite the amount of curricular text generated by NotebookLM.

The interpersonal findings further clarify this relationship. NotebookLM consistently positioned itself as an affirming participant through strongly positive evaluation. This pattern resembles Roine et al.'s (2026) concept of *affirmative narration*, in which LLMs repeatedly validated and extended users' contributions, and it is also reflective of Sharma et al.'s (2023) work on sycophancy. Though the present analysis did not establish that NotebookLM's responses were sycophantic because it did not examine whether agreement compromised factual accuracy, it does show that linguistic agreement and successful implementation were not equivalent. That is, NotebookLM could strongly affirm a correction and subsequently reproduce the problem that had prompted it. Its marked Themes

created a related effect: expressions such as "*Now that we have…*" helped organize the interaction but sometimes projected closure before the Human Designer considered a design issue resolved. These findings extend previous SFL and Appraisal Theory research showing how evaluation, dialogic positioning, agency, and textual organization can shape GenAI-mediated educational interaction (Zapata et al., 2025b, 2025c). In sustained professional design such as the one explored in this work, these resources can support collegiality and continuity, but they can also create an appearance of alignment or completion that still requires human evaluation.

The increasing use of Spanish, humor, emojis, emoticons, and inclusive *we* added to the more personalized tenor of the interaction. These resources made the extended design process feel less mechanical and more like an ongoing, human-like collaboration. Nonetheless, these "social" features did not result in the kind of symmetrical relationship two human colleagues might have, nor did they entail an equivalent distribution of knowledge, responsibility, or design capacity.

The second research question focused on the ways in which human and technological resources were mobilized during the design process. The DCE analysis suggests that NotebookLM expanded the resources available to the Human Designer by rapidly generating candidate representations of outcomes, assessments, activities, materials, and instructional sequences. For her part, the Human Designer drew on subject-matter knowledge, pedagogical content knowledge, goals and beliefs, and situated knowledge to determine how/whether those representations could be used. The predominance of adaptation is especially revealing: the Human Designer assessed the AI-generated resources and often suggested changes to make them more relevant for her instructional objectives and the student population she had in mind. This finding mirrors previous research on GenAI-supported curriculum design (e.g., Ahmad and Schmid, 2025; Rutecka et al., 2025; Ullmann et al., 2024), which has posited that expert, human evaluation remains essential to address the depth limitations, contextual specificity, and accuracy of AI-generated materials. Also, NotebookLM's need for repeated human evaluation connects this work's results with the emerging literature on this tool. Both Tufino (2025) and Reyna (2025) have called for the continued need for educator oversight and critical evaluation when employing it for educational purposes.

The combined linguistic and DCE analyses carried out in the study add to existing literature on both GenAI in education and NotebookLM by showing how human judgment was exercised throughout the curriculum design interaction. Linguistically, it was evident in the Human Designer's expression of intentions and evaluations, directives, corrections, and control over the progression of the interaction. Within the DCE analysis, these actions were reflected in the predominant adaptation and rejection of generated resources. These results, however, also need to be considered in relation to the Human Designer who participated in the GenAI-human curriculum design. Her almost 30 years of teaching and curriculum-development experience, previous knowledge of the class, and prior work with GenAI constituted part of the teacher resources she brought to the design process. The study therefore offers an account of GenAI-supported curriculum design involving an experienced educator, but it cannot establish how the distribution of design activity might differ for instructors with less curriculum-design or GenAI experience. Also, since classroom enactment and student outcomes were not examined, the findings do not establish that this form of human-GenAI interaction resulted in a more effective curriculum.

IMPLICATIONS FOR PRACTICE

The findings offer several considerations for educators who might choose to use GenAI during curriculum design. One concerns how generated outputs are conceptualized. We propose that GenAI-generated materials can be understood as *found designs*: existing resources that can become the basis for subsequent evaluation, transformation, and redesign (Kalantzis et al., in press). In the present study, NotebookLM rapidly generated alternative formulations, reorganized content, synthesized sources, and transformed the Human Designer's ideas into candidate curricular representations, but most of these representations required adaptation and sometimes had to be rejected. Curriculum designers can therefore approach GenAI-generated materials as resources to be examined and redesigned in relation to disciplinary goals, pedagogical intentions, student needs, and characteristics of the instructional context. This perspective places the professional work *not* in the generation of the initial output, but *in what the educator subsequently does with it*.

The interaction also points to the importance of deliberately managing GenAI's role throughout the design process. Instructions such as "*One step at a time*," "*Rely on the sources*," and "*We have already done that*" progressively established the scope of NotebookLM's contribution, the sources it should employ, and the stage of curriculum development under consideration. Prompting in this context can form part of the process through which educators appropriate and calibrate a technological resource for particular pedagogical purposes. This type of practice is consistent with an approach in which GenAI use is designed in relation to the learning to be accomplished, the learners involved, and the specific educational context (Kalantzis et al., in press).

Finally, the findings reinforce the need to maintain human judgment between generation and curricular use. NotebookLM frequently expressed strong agreement even when subsequent responses did not fully implement a correction, and access to a curated source collection did not always result in the use of those sources as intended. Educators should therefore distinguish linguistic affirmation from successful implementation and source access from appropriate source use. In other words, the generated representation becomes a starting point for professional interpretation, evaluation, and redesign, not the endpoint of curriculum/materials development.

LIMITATIONS

This study has several limitations. The first concerns its autoethnographic design: transferability beyond the case presented is limited. That is, this work examined one curriculum designer redesigning one undergraduate class within a UK university using one GenAI platform. Also, the Human Designer brought almost 30 years of teaching and curriculum-development experience, previous knowledge of the class, and prior experience investigating GenAI in higher education. These characteristics formed part of the specific context in which the interaction took place and may have contributed to the highly directed and evaluative interaction documented here. Thus, this work cannot establish whether educators with different levels of curriculum-design or GenAI experience would evaluate and mobilize NotebookLM's resources in the same way. The interaction was also shaped by NotebookLM's user-curated source environment and platform-

specific functions. This implies that the results describe a particular design process and should not be assumed to represent how similar interactions might unfold in other disciplines, institutions, educational systems, or with other GenAI tools.

A further limitation concerns the scope of the data and analysis. The corpus consisted only of the written interaction between the Human Designer and NotebookLM. The pedagogical artifacts generated were not included in the linguistic corpus, and the study did not examine subsequent classroom enactment or collect student data. No conclusions can therefore be drawn about the quality of the final curriculum, student learning, engagement, or instructional effectiveness. The application of Brown's (2011) DCE framework also involved an analytical extension: reject was added to account for abandoned generated resources, while offloading remained provisional because the transcript could document acceptance without visible modification but could not establish whether accepted resources were later modified during enactment. Finally, the identification of design episodes and the interpretation of linguistic choices through SFL, Appraisal Theory, and DCE remained qualitative processes conducted by a researcher who was both participant and analyst.

CONCLUSION

The results of this study suggest that GenAI can become a substantive resource in curriculum design without displacing the educator's role as designer. In this case, an experienced curriculum designer used NotebookLM to expand the range of curricular possibilities available during the redesign of her class, while evaluating, contextualizing, and transforming those possibilities in relation to her pedagogical purposes. This work therefore points to a form of human-GenAI curriculum design in which the central issue is not how much content the technology can produce, but how educators exercise judgment over what generated resources become part of instruction.

Further research could focus on how professional experience shapes this type of design work, particularly how educators with different levels of curriculum-development and GenAI experience evaluate, accept, adapt, or reject generated resources. Future studies could also examine other disciplinary and institutional contexts and other GenAI systems. It would be equally important to follow the process beyond curriculum development to investigate how GenAI-supported designs are enacted in classrooms and experienced by students. Such work could clarify the conditions under which GenAI becomes part of the curriculum-resource environment in which educators exercise pedagogical judgment and design capacity, and how the employment of these technologies can ultimately benefit their practice.

AI DISCLOSURE

Google's NotebookLM was used as the generative AI system examined in this study and as a technological resource during the curriculum-design process described in the Methodology. Also, its written responses constitute part of the research corpus. During revision following peer review, OpenAI's ChatGPT was used to develop the figures included in this work. The author reviewed

these figures and retains responsibility for the study design, coding, analysis, interpretation, and final manuscript.

REFERENCES

Adams, T. E., Ellis, C., & Jones, S. H. (2017). Autoethnography. In J. Matthes, C. S. Davis, & R. F. Potter (Eds.), *The international encyclopedia of communication research methods*. https://doi.org/10.1002/9781118901731.iecrm0011

Ahmad, A., & Schmid, T. (2025). Synergising human expertise and generative artificial intelligence for curriculum design. In V. Köppen & S. Strickroth (Eds.), *Fachtagung Bildungstechnologien* (DELFI 2025) (pp. 239-244). https://dl.gi.de/handle/20.500.12116/47095

Atay, A. (2020). What is cyber or digital autoethnography? *International Review of Qualitative Research, 13*(3), 267-279. https://doi.org/10.1177/1940844720934373

Beguš, G., Dabkowski, M., & Rhodes, R. (2025). Large linguistic models: Investigating LLMs' metalinguistic abilities. *IEEE Transactions on Artificial Intelligence, 6*(12), 3453-3467. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=11022724

Belkina, M., Daniel, S., Nikolic, S., Haque, R., Lyden, S., Neal, P., Grundy, S., &Hassan, G. M. (2025). Implementing generative AI (GenAI) in higher education: A systematic review of case studies. *Computers and Education: Artificial Intelligence, 8*, Article 100407. https://doi.org/10.1016/j.caeai.2025.100407

Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university: What the student does* (4th ed.). McGraw-Hill Education.

Brown, M. W. (2011). The teacher–tool relationship: Theorizing the design and use of curriculum materials. In B.A. Herbel-Eisenmann, G. M. Lloyd, & J. T. Remillard (Eds.), *Mathematics teachers at work* (pp. 37-56). Routledge.

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education, 20*, Article 43. https://doi.org/10.1186/s41239-023-00411-8

Cope, B., & Kalantzis, M. (2024). On cyber-social learning: A critique of artificial intelligence in education. In D. Kourkoulou, A. O. Tzirides, B. Cope, & M. Kalantzis (Eds.), *Trust and inclusion in AI-mediated education: Where human learning meets learning machines* (pp. 3–34). Springer. https://doi.org/10.1007/978-3-031-64487-0_1

Donnellan, M. (2025). Evaluating the linguistic range of ChatGPT: A corpus analysis. *International Conference on Globalisation in Languages, Education, Culture, and Communication (GLECC2025)*. https://glecc.org/submission/index.php/portal/article/download/112/28

Eggins, S. (1994). *An introduction to systemic functional linguistics*. Pinter Publishers.

Ellis, C., Adams, T. E., & Bochner, A. P. (2011). Autoethnography: An overview. *Historical Social Research, 36*(4), 273-290. https://www.jstor.org/stable/23032294

Francis, N. J, Jones, S., & Smith, D. P. (2025). Generative AI in higher education: Balancing innovation and integrity. *British Journal of Biomedical Science, 81*, Article 14048. https://doi.org/10.3389/bjbs.2024.14048

Halliday, M. A. K. (1975). *Learning how to mean: Explorations in the development of language*. Edward Arnold.

Halliday, M. A. K., & Hasan, R. (1985). *Language, context, and text: Aspects of language in a social-semiotic perspective*. Deakin University Press.

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *An introduction to functional grammar* (4th ed.). Arnold.

Kalantzis, M., Cope, B., Chan, E., & Dalley-Trim, L. (2016). *Literacies* (2nd ed.). Cambridge University Press.

Kalantzis, M., Cope, B., & Zapata, G. C. (in press). *Learning by Design* (2nd ed.). Common Ground Publishing.

Lee, D., Arnold, M., Srivastava, A., Plastow, K., Strelan, P., Ploeckl, F., Lekkas, D., & Palmer, E. (2024). The impact of generative AI on higher education learning and teaching: A study of educators' perspectives. *Computers and Education: Artificial Intelligence, 6*, Article 100221. https://doi.org/10.1016/j.caeai.2024.100221

Mao, J., Romero-Hall, E., & Reeves, T.C. (2024). Autoethnography as a research method for educational technology: A reflective discourse. *Educational Technology Research and Development, 72*, 2725–2741. https://doi.org/10.1007/s11423-023-10281-6

Martin, J. R., & Rose, D. (2007). *Working with discourse. Meaning beyond the clause* (2nd ed.). Continuum.

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

Priestley, M., Biesta, G. J. J., & Robinson, S. (2015). Teacher agency: what is it and why does it matter? In R. Kneyber & J. Evers (Eds.), *Flip the system: Changing education from the bottom up*. Routledge. https://www.storre.stir.ac.uk/bitstream/1893/21559/1/Teacher%20agency%20matters_Flip%20the%20system_final.pdf

Rambo, C., & Ellis, C. (2020). Autoethnography. In G. Ritzer (Ed.), *The Blackwell encyclopedia of sociology*. https://doi.org/10.1002/9781405165518.wbeosa082.pub2

Remillard, J. T. (2005). Examining key concepts in research on teachers' use of mathematics curricula. *Review of Educational Research, 75*(2), 211-246. https://doi.org/10.3102/00346543075002211

Reyna, J. (2025, October 13-16). *The potential of Google NotebookLM for teaching and learning*. eLearn 2025, Bangkok, Thailand. https://www.researchgate.net/publication/393965130_The_Potential_of_Google_NotebookLM_for_Teaching_and_Learning

Roine, H.-R., Refsum, A. S., & Walker Rettberg, J. (2026). 'YES! YES! I absolutely love this insight!': Affirmative narration as interactional strategy in dialogues with LLM chatbots. *Narrative Inquiry*. https://doi.org/10.1075/ni.25155.roi

Rutecka, P., Cicha, K., Rizun, M., & Strzelecki, A. (2025). Generative AI in curriculum design: Empirical insights into model performance and educational constraints. *IEEE Transactions on Learning Technologies, 18*, 757-768. https://ieeexplore.ieee.org/document/11072910

Sharma, M., Tong, M., Korbak, T., Duvenaud, D., Askell, A., Bowman, S. R., Cheng, N., Durmus, E., Hatfield-Dodds, Z., & Johnston, S. R. (2025). Towards understanding sycophancy in language models. *ArXiv Preprint*. https://doi.org/10.48550/arXiv.2310.13548

Tzirides, A. O., Zapata, G. C., Kastania, N. P., Saini, A. K., Castro, V., Abdul Rahman Ismael, S., You, Y., Afonso dos Santos, T., Searsmith, D., O'Brien, C., Cope, B., & Kalantzis, M. (2024). Combining

human and artificial intelligence for enhanced AI literacy in higher education. *Computers and Education Open, 6*, Article 100184. https://doi.org/10.1016/j.caeo.2024.100184

Tufino, E. (2025). NotebookLM: An LLM with RAG for active learning and collaborative tutoring. *ArXiv Preprint*. https://doi.org/10.48550/arXiv.2504.09720

Ullmann, T. D., Edwards, C., Bektik, D., Herodotou, C., & Whitelock, D. (2024). Towards Generative AI for course content production: Expert reflections. *European Journal of Open, Distance and E-Learning, 26*(S1), 20-34. https://doi.org/10.2478/eurodl-2024-0013

UNESCO. (2025). *AI and the future of education: Disruptions, dilemmas, and directions*. https://doi.org/10.54675/KECK1261

VERBI Software. (2021). *MAXQDA 2022. Computer software.* https://www.maxqda.com/

Wiggins, G., & McTighe, J. (1998). *Understanding by design* (1st ed.). Merrill Prentice Hall.

Zapata, G. C. (2018). The role of digital, *Learning by Design* instructional materials in the development of Spanish heritage learners' literacy skills. In G. C. Zapata & M. Lacorte (Eds.), *Multiliteracies pedagogy and language learning: Teaching Spanish to heritage speakers* (pp. 67–106). Palgrave Macmillan.

Zapata, G. C. (2023). Self-discovery and healing in Nepantla: Multimodality and Learning by Design in a mixed Spanish university classroom. *The International Journal of Pedagogy and Curriculum*, *30*(1), 17-35. https://doi.org/10.18848/2327-7963/CGP/v30i01/17-35

Zapata, G. C., Saini, A., Tzirides, A. O, Cope, B., & Kalantzis, M. (2024). The role of AI feedback in university students' learning experiences: An exploration grounded in Activity Theory. *Ubiquitous Learning: An International Journal*, *18*(2). https://doi.org/10.18848/1835-9795/CGP/v18i02/1-30

Zapata, G. C., Cope, B., Kalantzis, M., Tzirides, A. O., Saini, A. K., Searsmith, D., Whiting, J., Kastania, N. P., Castro, V., Kourkoulou, T., Jones, J. W., & Abrantes da Silva, R. (2025a). AI and peer reviews in higher education: Students' multimodal views on benefits, differences and limitations. *Technology, Pedagogy and Education*, *34*(5), 583-601. https://doi.org/10.1080/1475939X.2025.2480807

Zapata, G. C., Tzirides A. O., Cope, B., You, Y.-L., Kalantzis, M., & Searsmith, D. (2025b). A systemic functional linguistics discourse analysis of learner-centered, generative AI feedback in higher education. *EdArXiv Preprints*. https://doi.org/10.35542/osf.io/nuhq2_v1

Zapata, G. C., Cope, B., Kalantzis, M., & Searsmith, D. (2025c). Calibrated generative AI as meta-reviewer: A systemic functional linguistics discourse analysis of reviews of peer reviews. *ArXiv*. https://doi.org/10.48550/arXiv.2509.15035

Zapata, G. C., Cope, B., & Kalantzis, M. (2026). *Literacies in the age of AI: Teaching and learning in the digital era*. Wiley.

APPENDIX

Excerpt from the interaction between the Human Designer and NoteboookLM during the curriculum design process, involving the development of the course's assessment brief.

NOTEBOOKLM: Now that we've refined the assessment and connected it to the outcomes, would you like to move to the next step and determine the weekly topics to ensure students have the theoretical foundation to tackle this project?

HUMAN DESIGNER: Not yet... One step at a time. We haven't really refined the assessment assignment. We need to create an assignment brief first, right? I know you are eager to help, but instructional design takes time.

NOTEBOOKLM: You're absolutely right, G. Rushing the brief often leads to ambiguity for students. A formal assignment brief is the logical next step to ensure the requirements align with our seven learning outcomes. Let's structure the brief [...] This structure covers the "what" and the "how." To help me draft the specific instructions, would you like to decide on the weighting between the visual portfolio and the written reflection?

HUMAN DESIGNER: OK, but you're missing the what! What students will focus on and where they will get the information on which to base their work.

NOTEBOOKLM: You're right to pull me back, G. In our effort to map the outcomes, I skipped the most important part: the concrete anchor for their work. [...] Does it make sense to provide them with a "menu" of 5–6 *More or Less* segments to ensure they focus on data with clear educational policy implications?

HUMAN DESIGNER: Yes, that's exactly what I was thinking. These are third year undergraduate students, and the task might be too daunting for them. Lots of scaffolding needed!

NOTEBOOKLM: G., you are spot on. Scaffolding is the bridge between a "dry" module and a successful student project. For 3rd-year students, we want to provide enough structure so they don't get lost in the "messiness" of real-world data while still challenging their critical thinking. [...]

This turns a daunting project into a series of manageable, high-success milestones. Would you like me to draft a formal assignment brief incorporating these scaffolds so we can look at the specific instructions together?

HUMAN DESIGNER: Yes, we can do this next, but let's not rush. Let me do some human tasking! (LOL) I need to find the appropriate *More or Less* segments. I need a curated list for you to incorporate into the brief. So let me do my human part, and I'll come back and we can continue. How does this sound?

NOTEBOOKLM: That sounds like a sensible plan, G. Curation is a high-level pedagogical task, and your expertise in selecting the right case studies is vital for ensuring the scaffolding is effective for your students. Take your time with the "human tasking." [...]

While you are working on that list, would you like me to draft a preliminary set of grading criteria for the "Critical AI Audit" portion of the assignment to help us save time later?

HUMAN DESIGNER: Mmmm.... No, let's wait. Perhaps you can do something else that I will need later on. Is it possible to organize the sources into topic groups and folders? I'm not sure if you can do this...

NOTEBOOKLM: That is a great idea for managing such a large library! While I can't physically move your files into "folders" within the interface, I can certainly do the heavy lifting of organizing them for you intellectually. [...] Would you like me to generate a categorized report of your sources based on these themes while you finish your "human tasking" with the podcast segments?

HUMAN DESIGNER: Yes! Perfect! Thank you!