
GOLDMIND: A TEACHER-CENTERED KNOWLEDGE MANAGEMENT SYSTEM FOR HIGHER EDUCATION — LESSONS FROM ITERATIVE DESIGN

Gloria Fernández-Nieto

Faculty of Information Technology
Monash University
Melbourne, VIC, Australia
gloriamilena.fernandeznieto@monash.edu

Lele Sha

Faculty of Education
University of Macau
Macau, China
lelesha@um.edu.mo

Yuheng Li

Faculty of Information Technology
Monash University
Melbourne, VIC, Australia
yuheng.li@monash.edu

Yi-Shan Tsai

Faculty of Information Technology
Monash University
Melbourne, VIC, Australia
yi-shan.tsai@monash.edu

Guanliang Chen

Faculty of Information Technology
Monash University
Melbourne, VIC, Australia
guanliang.chen@monash.edu

Yinwei Wei

School of Software
Shandong University
Jinan, Shandong, China
weiyinwei@sdu.edu.cn

Weiqing Wang

Faculty of Information Technology
Monash University
Melbourne, VIC, Australia
teresa.wang@monash.edu

Jinchun Wen

Faculty of Information Technology
Monash University
Melbourne, VIC, Australia
jimwendev@gmail.com

Shaveen Singh

Faculty of Information Technology
Monash University
Melbourne, VIC, Australia
shaveen.singh@monash.edu

Ivan Silva

Faculty of Information Technology
Monash University
Melbourne, VIC, Australia
ivan.silvaferaud@monash.edu

Yuanfang Li

Faculty of Information Technology
Monash University
Melbourne, VIC, Australia
yuanfang.li@monash.edu

Dragan Gašević

Faculty of Information Technology
Monash University
Melbourne, VIC, Australia
dragan.gasevic@monash.edu

Zachari Swiecki

Faculty of Information Technology
Monash University
Melbourne, VIC, Australia
zach.swiecki@monash.edu

August 26, 2025

ABSTRACT

Designing Knowledge Management Systems (KMSs) for higher education requires addressing complex human–technology interactions, especially where staff turnover and changing roles create ongoing challenges for reusing knowledge. While advances in process mining and Generative AI enable new ways of designing features to support knowledge management, existing KMSs often overlook the realities of educators’ workflows, leading to low adoption and limited impact. This paper presents findings from a two-year human-centred design study with 108 higher education teachers, focused on the iterative co-design and evaluation of GoldMind—a KMS supporting in-the-flow knowledge management during digital teaching tasks. Through three design–evaluation cycles, we examined how teachers interacted with the system and how their feedback informed

successive refinements. Insights are synthesised across three themes: (1) Technology Lessons from user interaction data, (2) Design Considerations shaped by co-design and usability testing, and (3) Human Factors, including cognitive load and knowledge behaviours, analysed using Epistemic Network Analysis.

Keywords knowledge management systems, iterative study, human-centered design, epistemic network analysis, higher education

1 Introduction

Knowledge Management Systems (KMSs) increasingly use advanced technologies to support educators in digital environments in the context of higher education. For example, process mining can help inform process insights at scale automatically in teachers’ practices [1, 2]. Generative AI can be used to understand user context (e.g., vision models can extract information from images) and make complex information more accessible, for instance, by summarising lengthy documents [3]. These capabilities are particularly relevant in education, where knowledge in educational practices is highly contextual, distributed across platforms, and challenging to capture and transfer without disrupting daily tasks [4, 5].

In higher education, the ability to effectively capture and reuse institutional knowledge is vital for universities. This is particularly the case in contexts with high staff turnover, the prevalence of sessional academics, and evolving teaching responsibilities [6, 7]. A well-designed KMS promises to support the continuity of teaching practices, foster organisational learning, and enable the transfer of expertise from experienced educators to newcomers. However, implementing such systems in educational settings poses several challenges, particularly when technology, design, and human factors are not adequately integrated.

Technologically, many existing KMSs in education have focused on isolated knowledge management processes — such as storing or accessing content — without addressing knowledge management holistically across its four core processes: access, organisation, sharing, and application [8, 9]. This fragmented design limits the systems’ capacity to support educators in dynamic, real-world teaching environments. For example, a recent literature review by Fan and Beh [10] found limited empirical research on knowledge-sharing tools in higher education, with minimal input from educational stakeholders, suggesting a disconnection between the tools provided and the actual technological needs of educators. Similarly, Bilquise and Shaalan [11] found that although knowledge sharing and creation are the most researched knowledge management processes in universities, the technological support for collaborative practices and organisational culture, such as teachers knowledge workflows, remains underdeveloped. For example, Yun et al. [12], in a non-university context, highlights the need to implement mechanisms that enable knowledge flow and support the connection between existing knowledge and users’ needs. Yet, such tools fell short in addressing complex knowledge flows and routine teaching needs. These shortcomings illustrate persistent **technological challenges**, including the lack of intelligent, context-aware features (e.g., adaptive interfaces, personalised recommendations, or workflow integration) that support effective knowledge management for educators.

To overcome limitations in existing KMSs, a growing body of research advocates for **human-centred approaches** to KMS **design** — i.e., approaches that position educators not merely as end users but as co-designers of systems that reflect their needs, behaviours, and values [13, 14]. Human-Centred Design (HCD) emphasises meaningful participation and iterative refinement based on user input. In the context of education, where teachers work under cognitively demanding and time-constrained conditions, systems that are not carefully designed with the user in mind risk adding cognitive burden or disrupting established practices [15]. These challenges underscore the importance of integrating user feedback throughout the design process to ensure long-term adoption and usability.

Alongside technological and design considerations, **human factors** — such as motivation, trust, and cultural attitudes toward knowledge sharing — play a critical role in the success of KMSs [16]. As emphasised by Sarka et al. [14], future research must more thoroughly address the dynamic interplay between human factors and technology. Their work identifies several key dimensions that shape engagement with knowledge management systems in educational settings: behavioural — e.g., how technology influences teaching and knowledge management practices, cultural — e.g., openness to sharing knowledge, and experience related — e.g., differences in digital literacy across expert groups. These factors are especially relevant in higher education, where teachers operate within complex, cognitively demanding, and time-constrained environments [16, 11]. However, KMSs are often developed without a clear understanding of teachers’ real-world workflows [16], resulting in systems that do not align with daily teaching practices or the cognitive demands educators face [17, 18]. Introducing systems that increase cognitive load or fail to align with existing workflows can discourage usage and reduce long-term impact [18]. To be effective, KMSs must adapt to users’ real needs, routines, and working approaches — not the other way around.

Despite these known concerns, the technological, design, and human factors involved in implementing KMSs in higher education settings — particularly through iterative, human-centred approaches and within the context of real-world teaching tasks — remain underexplored. Exploring these factors can offer insights into how KMSs can be effectively integrated into educators’ workflows, how users engage with such technological support, and how these systems influence knowledge management behaviours over time. In response to this gap, our study presents the design and evaluation of *GoldMind*, a KMS developed through an iterative, human-centred process in collaboration with 108 university teachers over two years. The system supports the capture, storage, sharing, and reuse of expert knowledge in authentic teaching contexts. Through a series of design and evaluation cycles, we examined how teachers interacted with the system and how it supported their knowledge management practices.

Here, we synthesise lessons learnt across three system iterations, highlighting key insights across three interrelated themes: (1) *Technology*: We evaluated user interaction metrics (e.g., task time, failure rates) to assess system *effectiveness* and guide improvements. (2) *Design*: Through co-design sessions and usability testing, we iteratively refined system features and interface elements based on educator feedback. (3) *Human Factors*: We analysed users’ cognitive load, perceptions, and qualitative feedback to evaluate how the system aligned with educators’ work practices. To complement this analysis, we also explored knowledge management behaviours using epistemic network analysis (ENA) to model user interaction logs, uncovering patterns in knowledge management sub-processes such as storing, sharing, and applying knowledge.

By articulating these findings, this paper contributes: (i) practical guidance for designing teacher-centred KMSs based on one of the largest iterative design and evaluation studies with authentic teachers, and (ii) introduces GoldMind, a KMS that supports in-the-flow knowledge capture and dissemination in higher education.

2 Related Work

In this section, we first provide an overview of knowledge management in education. To contextualise the discussion, we draw on insights from recent systematic literature reviews and related works. These includes reviews on knowledge management in higher education [19, 20, 11, 21], future directions in knowledge management research [22], and emerging knowledge management topics in information technologies [23, 12]. We synthesise findings from these works to identify key technological, design, and human factors challenges that must be addressed to ensure that knowledge management practices in higher education, supported by KMSs, remain relevant and impactful.

2.1 Knowledge Management for Education

Knowledge management plays a pivotal role in capturing and disseminating knowledge within organisations, especially in knowledge-intensive fields like education [6]. In universities, where staff turnover is often high — due to retirement, institutional transfers, staff teaching reassignments, or the appointment of sessional academics and teaching assistants — preserving institutional knowledge is essential [7, 24, 25]. Doing so likely mitigates knowledge loss and ensures continuity in teaching and learning practices [26, 27].

Effective knowledge management enables the **access, capture, storage, sharing, dissemination, and application** of knowledge—processes essential for informed decision-making, problem-solving, and innovation [28, 8, 29, 30, 21]. Capture, storage, and access form the foundation for knowledge management activities; sharing allows individuals and teams to replicate effective practices; dissemination and application ensures that knowledge is widely spread and used to enhance performance and organisational growth [31, 32, 33].

Bose and Sugumaran [34] and Wong and Aspinwall [35] outline knowledge management processes that encompass the flow of knowledge within an organisation — from capture to dissemination and application — providing a useful framework for defining knowledge management-related activities. Experts engage in these processes both tacitly and explicitly in their daily workflows [36, 20]. In educational contexts, knowledge management processes are particularly critical for addressing staff turnover, supporting organisational learning, and enabling effective decision-making Asiedu et al. [20], Rehman et al. [16].

Educators engage in knowledge management activities [34, 16, 11] such as navigating repositories (*access*), uploading and organising materials (*storage*), annotating and collaborating on documents (*sharing*), and applying institutional knowledge (e.g., policies or best practices) to complete teaching tasks (*application*). Understanding these processes offers a structured method to capture and disseminate expertise, especially from experts to novices [16, 2].

Drawing on established knowledge management frameworks [37, 28, 8, 38, 6], and motivated by the documented gap in effectively capturing, storing—and especially—disseminating knowledge to mitigate loss and avoid unnecessary disruptions [39], our study focuses on the design and implementation of a KMS that supports both knowledge

management capture and dissemination processes. To address the challenges posed by staff turnover in universities, we designed features in a KMS that allow teachers to capture knowledge (e.g., document relevant information and workflows) while performing their regular tasks. The system also enables the dissemination of both newly captured and existing knowledge (e.g., pushing recommendations), in ways that enable its reuse by others, thus promoting its application and helping complete the knowledge flow within higher education institutions.

2.2 Technological challenges

Given the potential of knowledge management processes to enhance educational practice, researchers have explored how technology can support educators' knowledge management activities. Kaplan and Haenlein [40] demonstrated how collaborative tools such as wikis — with messaging, content integration, and announcement features — can facilitate knowledge sharing. Similarly, email and other communication tools have long been used to support collaboration and information exchange among educators [41]. Carroll et al. [28] developed a tool combining video conferencing, chat, email, and notebooks to help rural educators capture and organise their resources, activity plans, and student outputs for future reuse. More recently, Arpaci et al. [42] evaluated MOOCs as integrative KMSs and found them useful for storing and disseminating knowledge, though with varying degrees of effectiveness depending on the user context.

Generative AI (GenAI) is also creating new opportunities for knowledge management, although most existing research in this area has not focused specifically on educational contexts. For example, Brynjolfsson et al. [43] demonstrated the use of a GenAI-based conversational assistant in the workplace to enhance the productivity of customer support agents, particularly less experienced workers, by improving task performance, and enhancing customer interactions. Similarly, in an organisational context, Alavi et al. [44] illustrated how GenAI can influence knowledge management processes. Their study highlights GenAI's potential in knowledge storage and retrieval, enabling rapid access to large knowledge bases and reshaping how employees access to knowledge and interact with KMS.

Despite these advances, the implementation of knowledge management technologies in higher education remains limited due to persistent challenges that complicate their integration into educational settings. Higher education institutions are inherently knowledge-intensive and multidisciplinary, which increases the complexity of knowledge and makes it difficult for existing technologies to support efficient capture, retrieval and reuse of knowledge [18, 20]. Organisational and technological decisions are often made without adequately considering educators' real needs [17, 5]. Although GenAI models can be tailored with domain-specific datasets to adapt to specific domains, they often fail to incorporate human voices in their training processes [45]. This may cause the lack of trust in AI-generated recommendations [46, 44]. In addition, Galgotia and Lakshmi [18] emphasised the absence of mature, intelligent (e.g., AI-powered) KMSs in higher education settings that align with real educators' needs. Furthermore, Galgotia and Lakshmi [18] emphasised the lack of mature, intelligent (e.g., AI-powered) KMSs in higher education that truly meet the needs of educators. Galgotia and Lakshmi [15] argued that KMSs are rarely embedded in educators' everyday practices, resulting in fragmented knowledge flows and avoidable barriers. Building upon the work by Brynjolfsson et al. [43], Yun et al. [12], which explores the impact of generative technologies on productivity in non-educational domains (e.g., project management), such integration has the potential to enhance educator productivity, particularly for novice teachers. KMSs are essential to manage large, dynamic knowledge bases and provide timely, context-sensitive support (e.g., recommendations [43]). Specifically for GenAI, there is a documented need to balance human-derived tacit knowledge with AI-generated explicit knowledge [44]. Additionally, the risks of using GenAI to disseminate sensitive information remain underexplored [44]. In this paper, we directly address these challenges and discuss the lessons learned during the design and implementation of a KMS for a higher education institution.

2.3 Design Challenges

Key design challenges of KMSs include fragmented support for knowledge processes [42], limited integration of human-centred design, methodological evaluation gaps [22, 47, 14, 48], and a lack of context-specific design tailored to educational settings [49], all of which hinder KMSs adoption [22].

One promising direction to address these challenges is the integration of a holistic knowledge management process approach within established technology adoption frameworks. For example, Arpaci et al. [42] extended the Technology Acceptance Model (TAM) to incorporate knowledge management processes — such as access, storage, sharing, and application — to better evaluate the role of MOOCs in supporting knowledge management. While this work demonstrated that knowledge storage and access were positively perceived, the findings also underscore the need for deeper exploration of user experiences and perceptions to understand the impact of implementation.

Human-Centred Design (HCD) offers an important pathway toward developing more effective KMSs. HCD involves the iterative participation of users, designers, and researchers in co-developing tools that align with real-world needs [50, 13, 51]. As stated by Manesh et al. [22], human-centered approaches to designing KMSs approaches help mitigate

resistance to change and help to unleash true potential of people, by understanding their goals, and particular needs. For example, Nguyen Ngoc et al. [52] reviewed the opportunities of using HCD in KMSs, particularly as a means of incorporating human voices to support variety and customisation in system design. Similarly, Touré et al. [53] conducted focus groups to redesign a poorly adopted knowledge management tool. In a non educational context, HCD has also been explored in the context of knowledge management for manufacturing activities [54]. In another example, Pascal et al. [47] created a knowledge management portal for partner identification in R&D, using a science-based design approach (i.e., inspired by research evidence) combined with human-centered design. However, their approach relied more on literature and application reviews than on participatory design. In the work by Almunawar and Ordñez de Pablos [55, Chapter 4], online surveys were used to investigate university teachers’ perceptions of knowledge management technology use in Asia. While the study captured teachers’ views, it provided only descriptive statistics and lacked qualitative insights into users’ knowledge management preferences, behaviours, and needs.

Our study extends this body of work by involving both individual teachers and teaching teams — i.e., staff employed by higher education institutions in active teaching and coordination of university courses — in the co-design and evaluation of the GoldMind KMS. The system was developed to integrate seamlessly into their workflows and address the challenges identified by Sarka et al. [14], including gaining a holistic understanding of educators’ needs and the mechanisms required to support them. We conducted co-design sessions—a core human-centred design (HCD) approach that fosters equal participation between users and design experts in the development of digital tools [56]. This method supports the creation of culturally and contextually relevant systems by incorporating the lived experiences and perspectives of intended users [49].

Guided by prior research on knowledge management processes and technologies [14, 48, 15], our study gathered qualitative evidence from educators about their current knowledge management practices, tools, and needs. In addition, our iterative design approach allowed us to observe how teachers interacted with GoldMind, helping us identify further opportunities for improvement. This process ensured that human factors — such as working approaches, values, and social dynamics — were embedded in the system’s design.

2.4 Human Factors Challenges

According to Sarka et al. [14], research on knowledge management technologies needs to account for the complex interplay between human actors and technological systems. Their work highlights behavioural (e.g., how technology changes practice), cultural (e.g., openness to sharing), and generational (e.g., differing digital literacy or expertise) factors that influence knowledge management behaviours in education.

A deep understanding of user communities’ needs and workflows is critical for designing usable KMSs and for developing policies that foster their adoption among educators. Successful knowledge management also requires recognising the authentic authors of knowledge — namely, communities of practice [28]. This becomes especially important when incorporating emerging technologies such as GenAI, where intellectual property and ethical considerations are paramount [44]. In line with this, understanding how teaching practices should inform the design of KMSs – to support both individual and collective needs, such as reducing cognitive load – and how these systems, in turn, shape teachers’ practices, remains an open and underexplored area in the knowledge management literature [14, 48].

As highlighted in the literature review by Manesh et al. [22], future avenues for KMS research should include a deeper exploration of quantitative approaches – such as the use of process modelling languages – to better understand how knowledge management functions and to support its application within organisations. Beyond tool design, however, challenges also remain in understanding and supporting individual **knowledge management behaviours**. Gaining a deeper understanding of how educators share, store, access, and apply information can offer valuable insights into whether they lack awareness or structured approaches to their knowledge management behaviours, particularly as revealed through their interactions with KMS tools. This gap underscores the need for KMSs and processes that more closely align with users’ real-world behaviours, practices, and motivations.

2.5 Contribution to HCI and Research Question

The current study contributes to the body of research in human–computer interaction (HCI) by addressing the underexplored intersection between KMSs and HCD in higher education. While KMSs have long been positioned as tools for improving organisational learning and knowledge transfer [57, 58], their design has often been shaped by technological capabilities rather than educators’ lived experiences. Recent innovations — such as the integration of Process Mining and GenAI [59, 60] — present new opportunities to capture and reuse institutional knowledge more effectively [22, 18]. Yet, current implementations in educational settings frequently focus on isolated knowledge management processes such as storage or access [40, 48], often overlooking the complexity of real-world teaching workflows and organisational knowledge flows [15, 17].

Human-centred approaches offer a promising pathway for addressing these limitations. Prior studies have shown that engaging users in iterative design can increase system relevance, usability, and adoption [13, 53, 48, 52]. However, most research has remained limited to technology evaluations or design proposals rather than examining how educators actually engage with KMS over time. By involving 108 teachers in the co-design and iterative evaluation of the GoldMind system, our work contributes empirical insights into how educators’ knowledge management behaviours — such as access, storage, sharing, and application — unfold in authentic teaching contexts [36, 9, 15, 18].

Motivated by the persistent technological, design, and human factors challenges outlined in the literature [14, 42, 28], we pose the following research question (RQ):

RQ *What technological, design, and human factors challenges emerge through an iterative, human-centred implementation of a KMS in a university setting?*

This question reflects the need to advance HCI research on KMSs by surfacing design lessons and behavioural insights that can guide the development of more context-aware, participatory, and sustainable KMSs in education.

3 Iterative Study

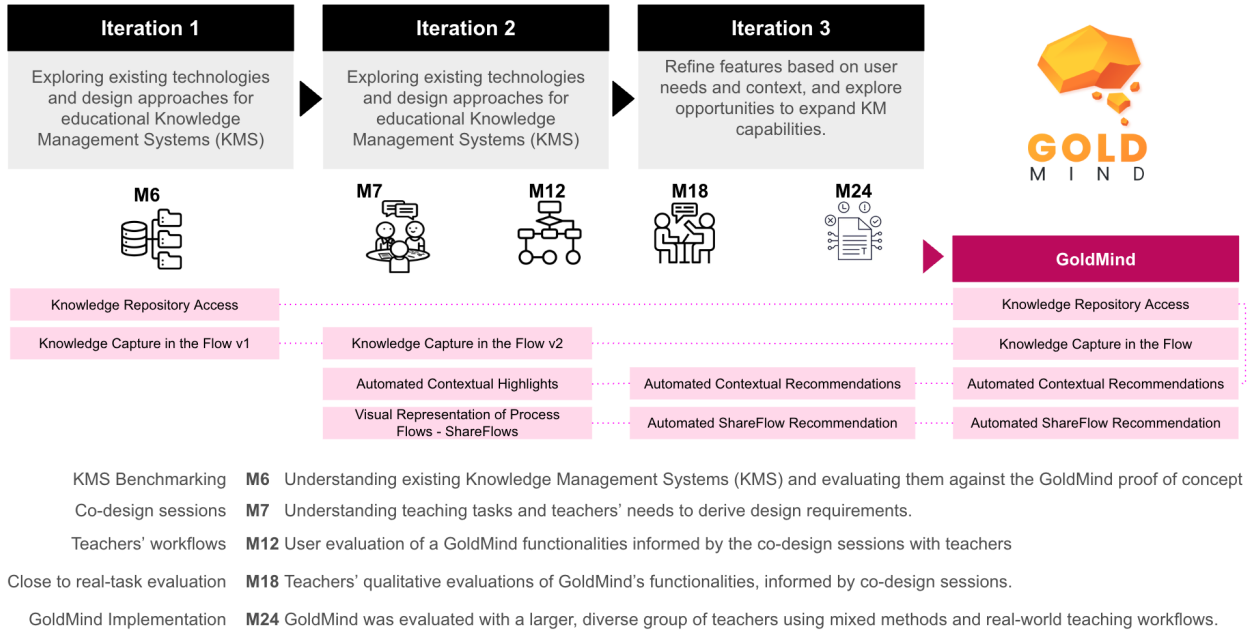


Figure 1: Human-centred iterative design and implementation of the GoldMind Knowledge Management System. Each iteration involved targeted activities, beginning with benchmarking existing KMSs (M6), followed by co-design sessions to understand teachers’ workflows and tasks (M7), user evaluations (M12), and qualitative assessments in near-authentic contexts (M18). The final implementation and large-scale evaluation occurred at M24. Across these phases, features evolved incrementally, including Knowledge Repository, Knowledge Capture in the Flow (v1 to v2), Contextual Highlights, ShareFlow visual representations, and automated recommendation features.

3.1 Context

We present a study that follows a human-centred approach [61]. Figure 1 shows the three iterations reported in this study, which involved interactions with educators from a research-focused university and an interdisciplinary research team (comprising experts in HCI, learning analytics, storytelling, education, and AI). These collaborations led to the co-creation of GoldMind, a KMS designed to support teachers in seamlessly capturing and disseminating knowledge. GoldMind is accessible via a browser extension¹. While the co-design and development of specific GoldMind features are beyond the scope of this paper, further details including technical aspects can be found in other sources [4, 5]. This paper focuses on the iterative evaluations and involvement conducted with teachers.

¹<https://colam.kmass.cloud.edu.au/>

Given the complexity of the teaching context and the need for a KMS to support it, we adopted an iterative, HCD approach involving authentic teaching teams and real teaching tasks. Our aim was to gather teachers’ feedback and experiences in ecologically valid settings that reflect their actual teaching practice. Each iteration was intended to incrementally improve the design based on evaluation results and teacher feedback [62]. We engaged educators in various instructional roles — including course coordinators, course lecturers, and tutors — to ensure a diverse range of perspectives across the iterations.

We used realistic task scenarios to maintain ecological validity in an experimental control setting [63]. This design allowed participants to engage with authentic teaching situations without the risk of affecting live systems (e.g., unintended changes to student marks in the Learning Management System). These tasks were embedded in teaching contexts where educators’ expertise was essential for delivering high-quality instruction. The specific context teaching tasks is introduced in the following section.

3.1.1 Teaching tasks

Teaching Tasks	Course administration	Learning material development	Learning engagement	Teaching Quality Improvement
Teachers' workflows (low-complexity)	<ul style="list-style-type: none"> Consultation timetabling Teaching venue booking 	<ul style="list-style-type: none"> Create assignment activity and Gradebook setup. Create lecture activity 	<ul style="list-style-type: none"> Student group allocation Adaptive quiz updates 	<ul style="list-style-type: none"> Complete peer review survey Identify training program
Teachers' workflows (high-complexity)	<ul style="list-style-type: none"> Moodle site setup Staff supervision Teaching session allocation 	<ul style="list-style-type: none"> Update unit guide Remote teaching session 	<ul style="list-style-type: none"> Exam marking consistency Online poll activity creation Discussion forum setup 	<ul style="list-style-type: none"> Reporting on student survey Update education practice resource for teaching teams

Figure 2: Representative teaching tasks performed by higher education teachers in their teaching workflow. **Red** indicates those tasks that are performed more than once per semester. We underscore the tasks selected for evaluation in this study. Task complexity is measured by the time and the number of resources required to complete the task.

Teachers, individually or in groups, regularly perform complex tasks to support their courses, such as *creating lecture activities*, *ensuring exam marking consistency*, or *updating teaching resources for teams*. These tasks often involve **workflows**, including navigating multiple information systems and accessing various resources. For instance, marking requires high levels of consistency, particularly in team settings, to ensure fairness and uphold student outcomes. Without proper guidance, novice teachers tasked with these responsibilities risk introducing inconsistencies. Similarly, managing teaching materials or integrating activities to foster student engagement further add to the complexity of teaching tasks [5].

The already demanding nature of teaching is further complicated when handovers result in the loss of expert knowledge. Over time, senior teachers develop deep expertise in their teaching, such as identifying effective strategies to engage students. However, teacher turnover — due to factors such as promotions or new job opportunities — often leaves insufficient time for documenting the regular workflows of experts [4]. This creates difficulties for novice teachers, who must navigate procedures and locate resources independently, a challenge exacerbated in universities where information is scattered across different systems.

Based on the co-design sessions described by Fernandez-Nieto et al. [5] and evaluations of teachers’ workflows reported by Sha et al. [2], we identified 19 representative teaching workflows across four categories: Course Administration, Learning Material Development, Learning Engagement, and Teaching Quality Improvement (see Figure 2). In these co-design sessions, teaching teams — including Chief Examiners, Lecturers, and Tutors² — collaboratively mapped their workflows using web-based collaborative tools (e.g., Miro). Through affinity diagramming, we systematically analysed and categorised these workflows, providing insights into the nuanced practices of expert teachers and the complexity of their tasks. This understanding informed the design of *Visual Process Representations*, which aim to simplify and communicate these workflows effectively. The design process and its outcomes are detailed in the following sections.

3.2 Study iterations and participants

The design and implementation of GoldMind followed three iterations. Figure 1 provides specific details on each iteration. The first iteration (Month 6 – **M6**), conducted in the second semester (late July to December) of 2022, focused on (i) understanding the knowledge management systems available in a higher education institution and (ii)

²Chief Examiners are responsible for course coordination, Lecturers manage course delivery and coordination, and Tutors support course management, marking, and tutorial delivery

Table 1: Summary of study iterations, participant roles, knowledge management tasks, and task types. Tasks were selected with a focus on knowledge capture and dissemination, aligning with the nature of our design. Task types are abbreviated as KC (Knowledge Capture) and KD (Knowledge Dissemination)

Iteration	Month	Number	Roles	Number of tasks	[Type] Teaching Tasks
1	M6	N=15	Tutor = 9, CE=3, Lecturer=3	3 KC, 3 KD	[Learning engagement] Create a feedforward poll (KC). [Learning engagement] Adaptive Moodle quiz update (KC). [Learning material development] Update and publish Unit Preview (KC). [Learning material development] Add learning objectives and materials to a Moodle page. (KD). [Learning material development] Create a new assignment in Moodle with a Turnitin draft submission (KD). [Learning material development] Enrol students into Gitlab groups (KD).
2	M7	N=13	Co-design Tutor = 4, CE=4, Lecturer=5	NA	NA
	M12	N=40	Tutor = 26, CE=5, Lecturer=9	6 KC, 6 KD	All M6 tasks [Teaching quality improvement] Marking correction documentation for consistency (KC). [Teaching quality improvement] Reporting and documenting a workshop issue via a video (KC). [Course Administration] Creating Miro boards and making them accessible for future use (KC). [Learning material development] Set up a scheduled Zoom lecture session (KD). [Learning material development] Set up a Moodle forum for a specific unit (KD). [Teaching quality improvement] Validate teacher induction (KD).
3	M18	N=10	Tutor = 7, CE=1, Lecturer=2	3 KC/KD	[Learning material development] Editing gradebook item. [Learning engagement] Create a feedforward poll. [Learning material development] Set up a scheduled Zoom lecture session.
	M24	N=30	Tutor = 19, CE=8, Lecturer=3	3 KC/KD	[Learning material development] Editing gradebook item. [Learning engagement] Create a feedforward poll. [Learning material development] Set up a scheduled Zoom lecture session.

benchmarking existing KMSs against an initial prototype. The second iteration (Month 12 == **M12**), conducted in 2023, involved (i) a participatory design process with authentic teaching teams to understand teachers' needs and identify design requirements for educational KMS (March-April 2023) and (ii) prototyping within the research team and evaluating GoldMind features with educators as they completed teaching tasks (August 2023). Finally, the third iteration, conducted in the last semester of 2023 - **M18** (late July to December) and the first semester **M24** (late February to mid June) of 2024, i) refined GoldMind's features and ii) gathered additional qualitative feedback from educators using the KMS to support their teaching tasks. Participants involved in the co-design sessions took part in iterations one and two.

Table 1 presents the number of participants per iteration (108 in total), participants' teaching roles, and the tasks that were tested in each of the iterations with the features implemented. All participants were linked to a university but were part of different Departments, including Medicine, Nursing and Health Sciences, and Information Technology. In each iteration, we evaluated GoldMind using the teaching tasks described in Section 3.1.1. In line with knowledge management foundations [33, 8], effective knowledge management allows the capture and dissemination of knowledge to be used when needed [38]. Thus, the iterations considered *Knowledge Capture* (KC) tasks and *Knowledge Dissemination* (KD) tasks.

For each iteration, this study examines three main groups of factors, outlined below:

- The **technological factors** refer to the challenges, constraints, and opportunities in developing and implementing KMS for teachers. These includes aspects such as knowledge capture, dissemination (retrieval), system automation, and usability enhancements.
- The **design factors** encompass usability, interface design, and interaction mechanisms required to improve teachers engagement with the KMS. These emphasise iterative design refinements, participatory approaches, and context-aware interfaces.
- The **human factors** consider the individual, collaborative, and behavioral dimensions of adopting KMS in higher education institutions. These includes how teachers engage with the system (e.g., perception, adoption, time on task and quality, and interaction), their cognitive load (e.g., mental effort involved in using the KMS), collaboration and communication (e.g., how the system supports knowledge sharing among educators), and knowledge management behaviors (access, storage, sharing, and application).

3.3 GoldMind Features

This section outlines the main features developed in each iteration to support knowledge management in teaching tasks. These features were informed by knowledge management frameworks and focus on two core processes: *knowledge capture* and *knowledge dissemination*. The goal is to ensure that knowledge is effectively captured, stored, and shared, while also guiding how institutions implement knowledge management practices.

3.3.1 Features – Iteration 1

This iteration aimed to evaluate the existing KMSs at a higher education institution and compare them to the first prototype of GoldMind, which included the following features.

Knowledge Repository: This feature supports *knowledge dissemination* by enabling the storage and organisation of knowledge. It indexes existing resources — such as documents and videos from various repositories — for easy retrieval. It uses a Differential Search Index (DSI) model to represent content as dense vector embeddings, enabling efficient document retrieval based on query similarity [64, 65]. This aims to help educators find relevant content within a platform that integrates information from multiple information systems (e.g., Learning Management System and internal university resources for educators, such as TeachHQ) that were previously siloed (Fig. 3—A).

Knowledge Capture in the Flow: This feature was designed to help teachers *capture knowledge* at varying levels of detail. It aims to enable them to annotate, highlight, and store knowledge while working on regular teaching tasks. To support this, we extended the capabilities of the open web annotation tool *hypothes.is*³. In addition to capturing content from the web, our implementation allows teachers to upload personal and team knowledge from external repositories such as *Google Drive*, integrating it into the knowledge background. We refer to this functionality as *knowledge clipping* (Fig. 3—B).

3.3.2 Features – Iteration 2

Features of Iteration 2 were informed by *co-design sessions* conducted with authentic teaching teams who teach and coordinate subjects in real-world university settings [5]. Teachers identified the need for capturing granular data, which aligned with features from Iteration 1, and the need for accessing knowledge while working on teaching tasks.

Knowledge Capture in the Flow: To complement the *knowledge capture* capabilities, we implemented *Video Annotation* by extending *hypothes.is*, which was originally designed for annotating web pages. Given that videos are commonly used by teachers when planning and preparing their classes, the ability to annotate such resources was considered useful for them. This feature allows teachers to capture knowledge in the form of annotations at multiple levels of granularity and in different modes (e.g., text and video annotations) (Fig. 3—C).

Visual Representation of Process Flows - ShareFlows: To support expert teachers in *capturing knowledge* — identified in the co-design sessions as a key pain point — we designed, implemented, and evaluated a mechanism to record their task-related trace data. This feature enables experienced educators (**experts**) to document task steps seamlessly, which can later be proactively recommended to **novice** teachers via push notifications. In this iteration, the capture process was embedded within teachers’ regular workflows to minimise disruption. The resulting visualisations was designed to support *knowledge dissemination* and were named *ShareFlows* [4]. The design of ShareFlows applies *data storytelling*

³<https://web.hypothes.is>

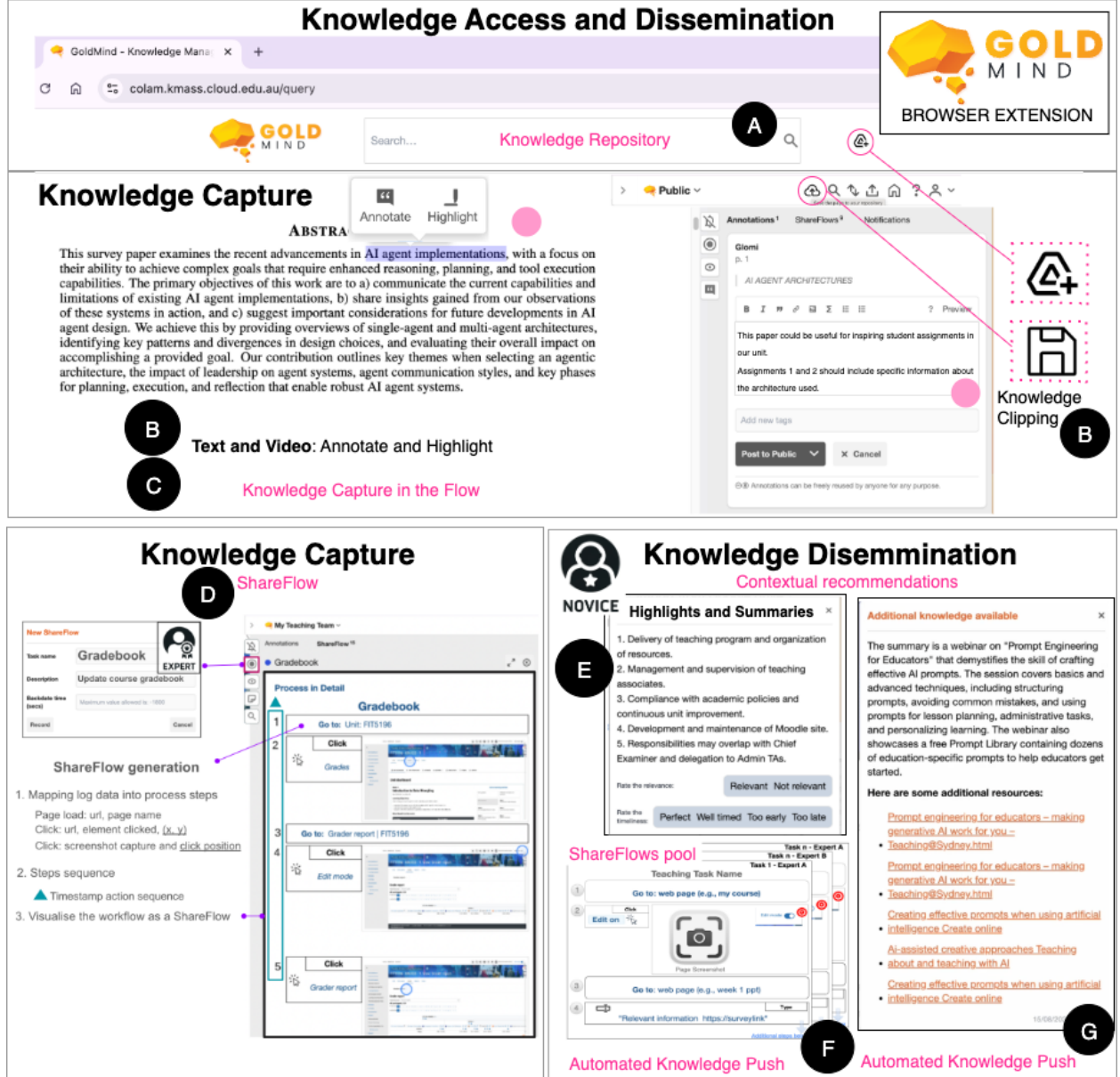


Figure 3: GoldMind Features. The diagram illustrates key components that support knowledge capture and dissemination within the GoldMind platform. Features include: (A) a centralised Knowledge Repository; (B) in-the-flow Knowledge Capture, including knowledge clipping, annotations, and uploading personal knowledge via a browser extension; (C) video annotation capture; (D) Visual Representation of Process Flows through ShareFlows; (E) contextual recommendations powered by highlights and summaries; (F) automated ShareFlow push recommendations; and (G) automated contextual recommendation mechanisms that surface relevant knowledge at the point of need.

principles [66, 67, 68, 69, 70, 71] to present complex expert trace data in a clear, accessible format for novices. The ShareFlow design included: (a) a descriptive title, (b) reduced visual clutter, (c) visual cues (e.g., click highlights) to emphasise key actions, and (d) a scrollytelling layout, a visual storytelling technique, that allows users to navigate steps by scroll up or down a page, enhancing engagement and readability [72] (Fig. 3–D). The functionality to recommend ShareFlows to novice teachers during task completion was implemented in Iteration 3.

Contextual Highlights: To support *knowledge dissemination* we implemented contextual recommendations to address teachers’ need for access to more relevant and situational knowledge. Using a generative language model (LLaMA v3.1), we developed a feature that summarises key information from documents accessed by teachers, listing relevant

insights drawn from existing knowledge sources. When teachers explored a knowledge repository, the system provided recommendations in the form of *highlights*, identifying crucial information within the document (e.g., web page) as pop-up windows (Fig. 3–**E**).

3.3.3 Features – Iteration 3

For Iteration 3, we focused on refining mechanisms for disseminating knowledge proactively. The refinements aimed to facilitate knowledge to be access and use automatically, reducing the effort required from teachers to search for information.

Automated ShareFlow Recommendations. This feature ensures knowledge capture and dissemination occur seamlessly within teachers’ workflows, minimising manual effort and disruptions. GoldMind captures novice teachers’ trace data about actions they take (e.g., page views, mouse moves, clicks) in real-time to provide timely recommendations by proactively pushing relevant ShareFlows from expert teachers (see Section 3.3.2). To extract key steps from process trace data generated through teachers’ interactions with GoldMind, this feature employs the heuristic miner algorithm [73], a widely used process mining algorithm known for its robustness against noisy traces, which are common in human process data [74]. In essence, GoldMind records detailed task traces from experts during their daily workflows, encodes them as visual step-by-step processes, and proactively recommends them to novices performing similar tasks (Fig. 3–**F**).

Automated Contextual Recommendation. Leveraging the Llama LLM (version 3.1), we identify the content of the web pages where teachers are working and recommend relevant resources from the knowledge repository. These recommendations are tailored to the task and content being reviewed by the teacher. The LLM generates a synthetic query based on the page content, which is then used to retrieve related resources from the existing Knowledge Repository (see Section 3.3.1). These recommendations are deliver as pop-up windows containing a list of resources useful for a particular task (Fig. 3–**G**).

4 Methods

4.1 Iterations with authentic teaching teams and tasks

Designing and implementing a KMS requires insights derived from realistic studies and task-based experiments with human participants to understand what works well and what needs refinement [75]. For complex systems — such as those involving contextual recommendations in information retrieval — it is critical to evaluate how effectively the system supports task completion and goal attainment [63]. Both holistic and component-level evaluations are essential for building functional KMS and generating actionable insights for iterative improvement [76].

Using authentic users in iterative studies enhances ecological validity, as it better reflects real-world use compared to evaluations with simulated end-users [77]. This approach also uncovers practical challenges that educators face when engaging with KMS tools in everyday teaching contexts.

To further balance realism with experimental control, we used real teaching tasks and adopted a simulated task-based approach [63]. Simulated tasks allowed participants to engage with the GoldMind system in realistic teaching scenarios without the risks associated with live environments (e.g., accidental changes to student marks in the Learning Management System). The teaching tasks used in this study were elicited through co-design sessions with educators [5] and are detailed in Section 3.1.1, with specific tasks per iteration introduced in Section 3.2.

Given GoldMind’s complexity, its evaluation involved benchmark test collections and overall metric measurements within teaching tasks. This section outlines the metrics used in each iteration to inform continuous system improvement. Moreover, combining iterative studies with authentic, realistic teaching tasks provides valuable insights that inform successive system redesigns (e.g., [78, 79, 80]). Across three iterations, we examined the key challenges of designing and deploying a KMS in higher education contexts. These challenges were analysed through three interrelated lenses: (1) technological factors, (2) design considerations, and (3) human factors.

4.1.1 Task design

Inspired by the notion of evaluating authentic tasks, we selected evaluation tasks for the GoldMind tool based on the following criteria: (1) tasks that are independently performed using the university’s teaching resources (e.g., intranet) and learning management system (e.g., Moodle), (2) tasks that are complex and time-consuming, and (3) tasks that are frequently performed and therefore highly relevant to teachers’ daily workflows.

Based on these criteria, we selected the evaluation tasks summarised in Table 3.2, aiming to reflect diverse teaching contexts. Two researchers (co-authors) conducted prior observational work to document educators’ typical workflows and resources. We also invited three expert teachers to perform the tasks while recording their interactions. Insights from these sessions informed the design of realistic and meaningful evaluation tasks.

4.2 Sources of evidence

Table 2 summarises the sources of evidence collected from teachers through system evaluations, interviews, and surveys. These sources informed the lessons learned and guided the results presented in the following sections. In each iteration, we explored the three key factors, technological, design, and human. However, it was only in the third iteration that we examined knowledge management behaviours using teachers’ trace data.

Iteration	Source of evidence
1 (M6)	Tool evaluation metrics (between-subject design) and SUS and NASA surveys (N=15)
2 (M7 and M12)	Interview (co-design) (N=13)
	Tool evaluation metrics (within-subject design) and SUS and NASA survey (N=40)
3 (M18 and M24)	Tool evaluation metrics (within-subject design) and individual interviews (N=10)
	Tool evaluation metrics (within-subject design), SUS survey, and individual interviews (N=30)

Table 2: Summary of interactions, sources of evidence, and explored themes and behaviours.

In the first iteration of the study (M6), teachers were asked to complete six teaching tasks (see Table 3.2). They were divided into two conditions: Baseline and GoldMind. The baseline condition involved the existing information systems used at the higher education institution. After completing the tasks, participants in each condition were asked to evaluate the usability of the tool(s) using the System Usability Scale (SUS) [81], and to assess cognitive load by using the NASA Task Load Index (NASA-TLX) [82]. This iteration aimed to identify initial technical challenges in existing systems, evaluate a prototype of GoldMind with early features, and gather insights into human factors to inform future design improvements.

In the second iteration (M12), we conducted a co-design session with authentic teaching teams to identify critical pain points and validate design requirements for the system. An updated prototype of GoldMind was evaluated using the same usability and cognitive load measures from the first iteration.

In the final iteration, a refined version of GoldMind was designed and implemented for evaluation. For comparative purposes, metrics were also collected for a Baseline condition, which was based on the Claude LLM. The Baseline system used a Retrieval-Augmented Generation (RAG) technique to ingest domain-specific data (e.g., university policies or course material) and support teachers in completing teaching tasks via a chatbot interface, which was akin to that of ChatGPT or other contemporary LLM-based chatbots. This Baseline was referred to as ChatUI for evaluation purposes.

The goal of this iteration was to assess the effectiveness of GoldMind in supporting teaching tasks compared to an LLM-based approach. To evaluate usability, we asked the participants in both conditions to complete the adapted version of the SUS survey that was developed to evaluate KMSs. The adaption expanded the original SUS 10 questions into 25 questions covering four key knowledge management practices of knowledge access, storage, sharing, and application, as reported in the knowledge management literature [28, 8, 29, 5]. Additionally, semi-structured interviews were conducted at two time points (M18 and M24) to explore teachers’ perceptions based on the following questions: i) *How helpful was the information provided by the Baseline/GoldMind for completing the tasks?* ii) *What features of the Baseline/GoldMind tool did you find most useful?* iii) *Did you encounter any challenges using the Baseline/GoldMind tool? If so, please describe them.*

4.2.1 Procedure

The tool evaluation was conducted in an experimentation room where participants worked independently on teaching tasks while being observed by a researcher from the adjacent room through a one-way mirror. Each participant used a laptop with an additional screen: the laptop displayed task instructions, while the external screen was used to perform and record tasks via the Chrome browser. Participants were first briefed on the study and given five minutes to explore the tool. The assigned tasks per iteration are detailed in Table 1.

4.3 Analysis

Given the multiple iterations and sources of evidence used to evaluate GoldMind, the following sections detail the specific analyses and data collected for each iteration, and how they align with the technological **T**, design **D**, and human factors **H**.

4.3.1 Iterations 1 and 2

Tool evaluation metrics. We adopted a user interaction approach based on trace data to understand how users engage with the system, identify areas for improvement, and inform design decisions [83, 84]. In Iteration 1, metrics were primarily derived from manual video annotations, while in Iteration 2, data capture was automated. The Baseline for comparison was the university’s existing knowledge management systems (e.g., intranet repositories, Google Drive). Metrics focused on two task types: (i) dissemination—measuring the time spent searching for task-related information to assess the system’s effectiveness in supporting information dissemination, and (ii) knowledge capture—measuring time spent capturing information (e.g., note-taking for unit planning). The evaluation metrics included:

- (i) Retrieval speed: Time taken by the system (Baseline or GoldMind) to return a response to a query **T**.
- (i) Relevant responses ratio: Teachers rated the usefulness of responses on a 0–4 scale. This metric also included the number of actions (e.g., clicks and scrolls) performed to find relevant information **H**.
- (i) Time to action: Time taken by users to initiate the next action after receiving a response (e.g., creating an assignment in Moodle) **H**.
- (i, ii) Task duration: Total time spent on the task **H**.
- (ii) Knowledge capture efficiency: The average cost of capturing knowledge, calculated using time on task (for capture tasks), time spent capturing knowledge per condition, and task completion quality **T**.
- (i) Just enough responses: Number of queries needed to retrieve useful information for completing the task **T**.
- Task completion ratio: proportion of tasks completed by participants per condition **H**.
- (i, ii) Task completion quality: Task performance rated on a scale from 0 to 1, based on a predefined rubric **H**.

For time to action, task duration, knowledge capture efficiency, and just enough responses, we tested for statistical differences between conditions using the mixed-effects model in Formula 1, with a fixed effect for condition (c) and a random effect for participant (p).

$$Y_{\text{time, duration, capture, responses}} = \beta_0 + \beta_1 \times c + \gamma_p + \epsilon \quad (1)$$

We used a mixed-effects regression model to assess statistical differences in retrieval speed, task completion time, and quality metrics (dissemination and capture) across the two experimental conditions. The model included fixed effects for condition (c) and task (t), and a random effect for participant (p); see Formula 2.

$$Y_{\text{quality}} = \beta_0 + \beta_1 c + \beta_1 t + \gamma_p + \epsilon \quad (2)$$

For task completion and the relevant responses ratio, we used a binomial logistic regression model to assess statistical differences between the two conditions (GoldMind vs. Baseline). We analysed the relationship between the condition (independent variable, denoted as c) and the dependent variables (completion and relevance); see Formula 3.

$$\log \left(\frac{p}{1-p} \right) = \beta_0 + \beta_1 c \quad (3)$$

We adopted a usability questionnaire to gain quantitative insights into teachers’ perceptions of the tools. We used the **SUS** survey[85] **D**, which elicited relevance scores based on participants’ perceptions of how useful the system was in supporting the task context (e.g., [86]). Responses were collected on a 5-point Likert scale (1 = strongly disagree, 5

= strongly agree). The numerical values of these responses were used to calculate a final usability score according to the standard SUS scoring procedure [81].

NASA-TLX was used to measure cognitive load across six dimensions [H]: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration [87]. Responses were collected using a 5-point Likert scale (1 = Very Low, 5 = Very High). The numerical values were used to calculate the mean score for each dimension and were visualised using box plots.

4.3.2 Iteration 3

For this iteration, we followed a within-subjects design and simplified the evaluation metrics. The main focus was to assess the system’s effectiveness in making pre-existing knowledge (e.g., captured by experts) accessible to new teachers via automated push dissemination (see features in Section 3.3.3). A chatbot tool based on the Claude LLM served as the Baseline. We also gathered qualitative insights through post-task interviews with teachers.

Tool evaluation metrics. The set of metrics considered for this iteration are as follows:

- Reduction in time to access knowledge: We measured the time-on-task for each participant and evaluated how long they spent locating relevant information to complete the assigned tasks [T].
- Failure rate: During the evaluation, teachers were given a limited amount of time to complete the tasks. We then recorded how many tasks each teacher successfully completed within the allotted time [H].
- Task completion quality: We tracked user actions—such as clicks, scrolling, and navigation steps—performed while searching for the correct response. These actions were used to assess the efficiency and focus of participants’ task completion [H].

For task completion time and reduction in knowledge management time, we measured statistical differences between conditions using the mixed-effects regression model defined in Formula 1. For the failure rate, we used the binomial logistic regression model described in Iteration 1.

For participants’ task quality, we used fixed effects for condition, start condition (i.e., using Baseline tool first or GoldMind tool first, denoted as s) and task (denoted as t), considering the interactions between condition (i.e., baseline vs. GoldMind) and start condition, as well as interactions between condition and task, and a random effect for participants, see Formula 4.

$$Y_{\text{quality}} = \beta_0 + \beta_1 c + \beta_2 s + \beta_3 (c \times s) + \beta_4 t + \beta_5 (t \times c) + \gamma_p + \epsilon \quad (4)$$

Knowledge management SUS survey [D]. We adopted a usability questionnaire to gain quantitative insights into participants’ perceptions of the tool. The questionnaire was adapted from the most widely used standardized questionnaire (i.e., SUS) for the assessment of perceived usability [81]. The adaption expanded the original 10 questions into 25 questions covering four key knowledge management practices of knowledge access, storage, sharing, and application, as reported in the knowledge management literature [28, 8, 29, 4].

Teachers’ Interviews. To analyse the three four-ended questions described in Section 4.2, one of the authors, with a background in HCI, coded the transcripts of the responses and used audio recording data to support interpretation. The coding process was guided by predetermined categories outlined in the study procedure (Section 2): (i) educators’ perceptions of the usefulness of GoldMind and the Baseline [D] and (ii) educators preferred features of GoldMind and the Baseline [T], and (iii) educators’ perceptions of the challenges they faced using GoldMind and the Baseline [H]. Since the categories were distinct, none of the quotes overlapped, and the coder categorised each teacher’s quotes into each of the categories above (i-iii). The resulting coded statements were then reviewed by the authors, who engaged in several discussions to select instances that best illustrated the usefulness, preferred features, and challenges of GoldMind and Baseline.

As part of the human factor exploration [H], we conducted **Epistemic Network Analysis (ENA)** [88] to analyse teachers’ knowledge management behaviours while using the systems (Baseline and GoldMind). In brief, ENA creates network models for each unit of analysis. The nodes of these networks are codes applied to the data—i.e., labels that describe important themes or actions that the units take; the edges represent the co-occurrence of these codes within pre-define windows of data for that unit. Thicker, more saturated edges indicate higher frequency co-occurrences

between codes. In this way, ENA can model the patterns of interconnected actions that individuals take. These networks can be analysed using two coordinated representations: (i) a network graph (as described above), and (ii) a summary statistic derived from a dimensional reduction on set of networks in the data. Networks and their summary statistics (called ENA scores) are visualised in the same low-dimensional space where the dimensions are linear combinations of the original co-occurrence variable that explain the maximal amount of variance between the units of analysis. The dimensions can be interpreted using the network diagrams, which are arranged such that edges on the extremes of the dimension correspond the co-occurrences that distinguish units of analysis the most. Both ENA scores and ENA networks can be aggregated to compare subpopulations within the data and networks for different subpopulations can be subtracted to show the co-occurrences that were stronger in one subpopulation compared to the other. Standard statistical techniques can be applied to the ENA scores to test for systematic differences between subpopulations. For more details on ENA, see [89]. This analysis was only performed for Iteration 3, as GoldMind and the Baseline were mature enough at that stage to collect teachers’ trace data.

The ENA model was configured as follows: *units* corresponded to participants by task and condition; *codes* represent knowledge management sub-processes (see Table 3); *conversations* include the system and task name (Task 1 or Task 2); *window* size was set to 40 lines, manually derived by sampling 50 actions in the trace data and using the median window length; and *metadata* included task outcome (Pass or Fail) and condition order. We constructed the ENA model using the means rotation (MR) approach, which identifies the network embedding space whose first dimension maximises between-group variance. Here, we maximised the difference between the Baseline and GoldMind participants. To visualise the differences between conditions, we generated a subtracted network between the mean network of the GoldMind participants and the mean network of the Baseline participants. This graph showed the connections that occurred more frequently in one condition compared to the other. We complemented this network comparison with statistical tests using a mixed-effects regression model with position on the means-rotated dimension (*MRI*) as the outcome variable, condition (*c*), task (*t*), task order (*s*), and task outcome (*o*) as fixed effects, and participant (*p*) as a random effect, see Formula 5.

$$Y_{MRI} = \beta_0 + \beta_1 c + \beta_1 t + \beta_1 s + \beta_1 o + \gamma_p + \epsilon \quad (5)$$

To code the data, we identified sequences in the teachers’ trace data and automatically⁴ mapped them to knowledge management sub-processes — i.e., knowledge access, storage, sharing, and application — in alignment with established knowledge management frameworks [28, 8, 29, 5]. The teachers’ trace data used to identify the codes included actions (e.g., clicks, type, navigation, or uploading of new pages) and other metadata, such as URLs, to validate the resources accessed by teachers. Table 3 summarises the code names, knowledge management sub-processes and definitions, illustrations of mapped sequences from trace data, and examples of what these may represent in a task context.

Code name	KM Process - Definition	Sequences	Example
AP.1 Search and Explore	Access process: The user searches for a page, explores it, and moves to a new resource.	['Navigation', 'Scroll', 'Click', 'Navigation']	Searching for information across multiple web pages.
AP.2 Switch Tabs	Access process: The user navigates across tabs in the browser.	['Navigation', 'Navigation', 'Navigation']	Searching for the correct tab where information is located.
AP.3 Read/Engage with Content	Access process: The user interacts with a single resource for a period of time.	['Click', 'Click']	Engaging with a page, likely reading relevant information.
AP.4 Open and Interact with Content	Access process: The user opens a new resource and starts interacting with it.	['Navigation', 'Click']	Looking for information on a newly opened page.
SP.1 Type and Interact	Store process: The user interacts with a resource and types information for a period of time.	['Click', 'Click', 'Click', 'Click', 'Type']	Creating a forum in Moodle.

⁴using process variation analysis and implemented with the pm4py Python library

SP.2 Select and Type	Store process: The user selects an option from a form on a page.	['Select', 'Type']	Selecting an option from a form.
SP.3 Type-driven Content	Store process: The user types in a single resource for a period of time.	['Type', 'Type']	Adding a description or explanation for an assessment.
SP.4 Open and Type	Store process: The user accesses a resource and types information.	['Navigation', 'Click', 'Type']	Searching for a type of activity in the LMS.
ShP.1 Share Resource	Share process: The user creates a resource and makes it available to others.	['Submit']	Creating a forum for future teacher or student use.
ShareFlow with Interaction	Apply process: The user explores an automated recommendation within a task context and interacts with it.	['push recommendation', 'click pop up window']	Interacting with an automated recommendation to complete a task.
ShareFlow without Interaction	Apply process: The user explores an automated recommendation within a task context, but interaction is unknown.	—	Exploring an automated recommendation, but interaction is unknown.
Knowledge with User Interaction	Apply Process: Automatic Recommendation of Organisational Knowledge (GoldMind only)	['push recommendation', 'click pop up window']	Teacher click and engage with the recommendation
Knowledge without User Interaction	Apply Process: Automatic Recommendation of Organisational Knowledge (GoldMind only)	—	Exploring an automated recommendation, but interaction is unknown.
Querying	Knowledge Access: Query interaction available in GoldMind and Baseline (LLM)	Trace data when teachers interact with the query engine (e.g., Chatbot)	Teachers asking for information to support their tasks execution

Table 3: Coding Scheme for Epistemic Network Analysis (ENA) mapping Knowledge Management sub-processes

5 Lessons Learnt

This section presents the results from the three iterations of the GoldMind system development an evaluation. The results per iteration are organised around three main dimensions: **technological**, **design**, and **human factors**. At the end of each iteration, we highlight key lessons learned and their contributions to the field of Human-Computer Interaction (HCI). Knowledge management behaviours, which fall under human factors, are reported in a separate subsection.

5.1 Iteration 1 (Month 6): Initial Evaluation and Benchmarking

Table 4 summarises the evaluation metrics from the first iteration, comparing GoldMind with the Baseline. The results are organised and explained according to the three evaluation factors. Metric abbreviations (e.g., **m1**) are provided in the summary table for reference.

Technological Factors: The results indicated that GoldMind outperformed the Baseline in several key aspects related to knowledge capture and dissemination. In terms of dissemination, the retrieval speed **m1**, referring to the query response time, improved significantly ($p = 0.000$, $t = -4.11$) with GoldMind, which was 73% faster. Regarding knowledge capture efficiency ($p = 0.04$, $t = -2.05$) **m5**, the effort required to record knowledge was lower in GoldMind compared to Baseline. Specifically, participants using GoldMind required 41% less effort to capture knowledge in the

Table 4: Iteration 1: Comparison of GoldMind and Baseline performance using descriptive and statistical analyses, including 95% confidence intervals (CI) on the difference between GoldMind and Baseline and measures of improvement. For the NASA metrics, the abbreviations stand for: Mental demand (Md), Physical demand (Pd), Temporal demand (Td), Performance (P), Effort (E), and Frustration (F).

Metric	Fac.	GoldMind	Baseline	Statistical Analysis	CI (difference)	Improvement
m1 Retrieval speed	T	39 secs, SD = 108.52	146 secs, SD = 112.69	$\beta = -98.99, p = 0.000, t = -4.11$	-146.21 – -51.78	73% faster
m2 Time to action	H	32.53 secs, SD = 41.19	70.64 secs, SD = 61.46	$\beta = 38.63, p = 0.027, t = -2.22$	48.974 – 124.96	35 secs faster
m3 Relevant response ratio	H	2.77	1.86	$\beta = 0.098, p = 0.75, t = 0.51$	-0.56 – 0.95	48% more relevant responses
m4 Task duration	H	717 secs, SD = 464.8	1172 secs, SD = 479	$\beta = -485.76, p = 0.000, t = -3.80$	-64.05 – 285.71	1.68% faster
m5 Knowledge capture efficiency	T	23.32 secs, SD = 39.79	39.89 secs, SD = 46.37	$\beta = -22.68, p = 0.040, t = -2.05$	-44.283 – -1.077	41% lower effort
m6 Just Enough Response	T	0.137, SD=0.56	0.20, DS=0.44	$\beta = -38.63, p = 0.027, t = -2.21$	-72.76 – -4.50	-44% fewer relevant responses
m7 Task Completion ratio	H	0.92, SD=0.89	0.85, SD=91	$\beta = 0.22, p = 1, t = -0.36$	0.50 – 0.80	6% increase
m8 Task completion quality [0–1]	H	0.96, SD = 0.10	0.97, SD = 0.10	$\beta = 0.22, p = 0.37, t = -0.09$	-0.080 – 0.055	-1.9% worse
m9 NASA	H	Md: 2.60, SD = 0.97 Pd: 2.26, SD = 1.10 Td: 2.55, SD = 1.15 P: 4.42, SD = 0.88 E: 2.44, SD = 0.97 F: 2.23, SD = 1.12	Md: 2.63, SD = 1.01 Pd: 2.25, SD = 1.13 Td: 2.58, SD = 1.20 P: 4.30, SD = 0.98 E: 2.44, SD = 1.05 F: 2.27, SD = 1.18	Md: $\beta = -0.51, p = 0.14, t = -1.47$ Pd: $\beta = -0.20, p = 0.62, t = -0.48$ Td: $\beta = -0.72, p = 0.08, t = -1.74$ P: $\beta = 0.46, p = 0.17, t = 1.36$ E: $\beta = -0.46, p = 0.23, t = -1.20$ F: $\beta = -0.78, p=0.038, t = -2.07$	Md: -1.19 – 0.16 Pd: -1.02 – 0.61 Td: -1.52 – 0.08 P: -0.20 – 1.12 E: -1.20 – 0.28 F: -1.52 – 0.04	-1.1% less Md 0.4% more Pd -1% less Td 0.2% more P same E -1 % less F
m10 SUS	D	59.3, SD = 22.67	37.8, SD = 20.30	$\beta = 22.84, p = 0.048, t = 2.20$	0.25 – 45.44	16% higher

system (e.g., capturing and saving video tutorials). However, GoldMind underperformed in the Just Enough Response ($p = 0.027, t = -2.21$) m6 metric (-44%), indicating that the responses were not always sufficiently relevant to support task completion.

Design Factors Early usability testing revealed barriers to adoption, emphasising the need for greater stakeholder involvement in system refinement. System Usability Scale (SUS) ($p = 0.048, t = 2.20$) m10 scores showed a significant improvement with GoldMind (59/100) compared to the Baseline (38/100). See Figure 4. This is in line with the task duration metric that was shorter in GoldMind than the Baseline.

Human Factors: The time to action m2 metric indicated that GoldMind users were faster (32 secs) in deciding which next action to execute ($p = 0.027, t = -2.22$). This suggests that when responses were relevant to the task context, they also provided clear guidance on the next steps to be executed. Task duration ($p = 0.00, t = -3.80$) m4 was shorter for GoldMind users compared to those using the Baseline. The Relevant Response Ratio m3 showed that

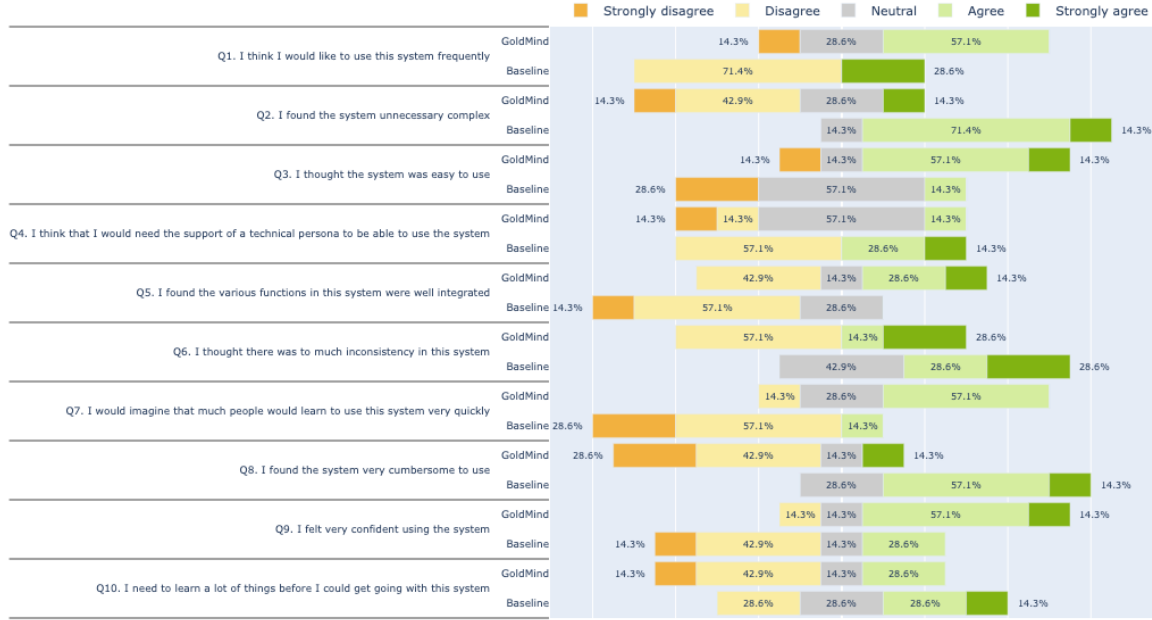


Figure 4: System Usability survey response rating by participants in Iteration 1

participants found GoldMind’s responses to be more relevant (manually rated on a 0–4 scale). The Task Completion Rate (m7) was higher for GoldMind users, showing an improvement over the Baseline. Similarly, Task Completion Quality (m8) was also higher for GoldMind users. Regarding the NASA Task Load Index (NASA-TLX) (m9), the results showed that GoldMind outperformed the Baseline in five out of six usability dimensions, with the exception of Q2 – Physical Demand. Users reported feeling less rushed, experiencing lower mental demand, and feeling more confident in completing their tasks when using GoldMind. Figure 5 presents the detailed NASA-TLX survey responses. Among all dimensions, only the frustration item showed a statistically significant difference ($p = 0.038$, $t = 2.20$).

What we learnt from this iteration:

Lesson 1. *In-context Knowledge Capture and Unified Access Improve Efficiency and Task Support.* GoldMind’s integrated features — such as the Knowledge Repository (A) and Knowledge Capture in the Flow (B) — effectively addressed key limitations of the baseline system, including disconnected and decontextualised knowledge. These features enabled teachers to capture and retrieve information directly within their workflow, leading to more efficient retrieval (m1), improved task execution (m4, m7), and higher task quality (m8).

Lesson 2. *Context-aware Retrieval Requires Further Refinement.* Although GoldMind improved performance across several metrics, it underperformed in delivering consistently relevant responses within specific task contexts (m6). This suggests a need to refine the contextual retrieval mechanisms or explore alternative strategies for aligning recommendations with users’ immediate task needs.

Lesson 3. *Alignment with Teachers’ Workflows and Expectations Needs Improvement.* GoldMind received lower usability scores (SUS) (10) and was rated lower on human factors (m9), particularly in terms of users’ perceived performance. These results indicate a need to enhance the system’s user experience and better align its functionality with diverse teaching workflows and expectations.

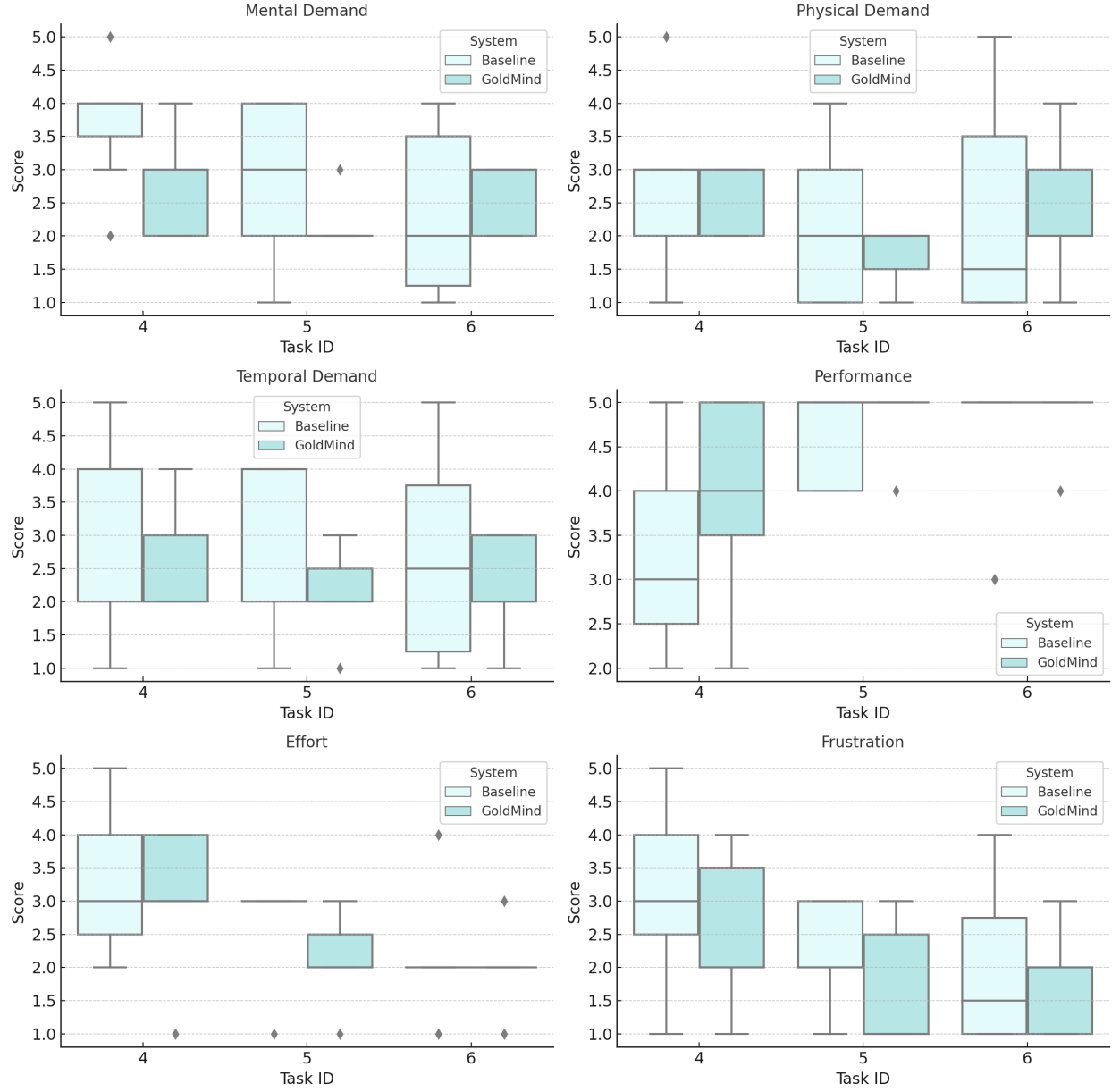


Figure 5: The results of Iteration 1 (M6) – the descriptive statistics of the responses to the NASA task load index survey

5.2 Iteration 2: Co-Design and Iterative Prototyping

The second iteration consisted of a co-design session and the evaluation of the systems informed by the co-design. The lessons learned on each of these stages are detailed below.

5.2.1 Co-design (Month 7)

Building on the findings from the first tool evaluation in this iteration, six co-design sessions were conducted with six teaching teams (13 teachers) to identify their needs and further refine the system’s functionalities. In these sessions, the teams described their teaching workflows, evaluated the existing tools used in their practices, and, using inspirational cards, articulated the technological support they require to enhance their Knowledge Management practices. Detailed descriptions of the activities of the co-design sessions are provided in the work by Fernandez-Nieto et al. [5].

Design Factors: The co-design sessions led to the development of four core **design requirements** (DRs) for Knowledge Management Systems:

- DR1** Facilitate access to relevant knowledge. Educators need a single point to access and retrieve existing knowledge efficiently, as current repositories are scattered across multiple platforms.
- DR2** Granular Knowledge Storage: Educators benefit from multimodal document management that enables structured storage and retrieval of knowledge at varying levels of detail. Features like content summarisation, annotations, and metadata can enhance usability and facilitate knowledge application.
- DR3** Proactive Recommendations and contextualised suggestions. Educators require personalised, context-aware recommendations integrated into their workflows to support knowledge application.
- DR4** Facilitate knowledge Sharing and Expert identification: Educators seek tools to locate and connect with community knowledge and expertise, fostering collaboration and reducing knowledge silos. Social software can enhance knowledge exchange, supporting communities of practice in higher education.

Technological factors: These design requirements (DRs) informed the design of the GoldMind system, helping identify which functionalities from Iteration 1 aligned with teachers’ needs and guiding the design and implementation of new features in Iterations 2 and 3. The GoldMind system integrates four key functionalities (F1–F4).

- F1** Aggregating distributed knowledge: To address DR1, this functionality indexes knowledge from various repositories (e.g., university documents and videos) for easier retrieval. In Iteration 1, we introduced the Knowledge Repository (Section 3.3.1, **A**) and, based on its alignment with co-design session needs, retained it in later iterations.
- F2** Capturing and linking multimodal knowledge: Addressing DR2, this functionality supports knowledge capture and storage at multiple levels of granularity. In Iteration 1 (Section 3.3.1, **B**), the *Knowledge Capture in the Flow* feature was used to highlight and annotate documents within their workflows. In Iteration 2 (Section 3.3.2, **C**), we extended this to video annotation and introduced *ShareFlows*, a feature that captures expert interactions and translates them into reusable step-by-step workflows (Section 3.3.2, **D**).
- F3** Context-aware recommendations: Addressing DR3 and DR4, this functionality delivers relevant knowledge based on users’ workflows and tasks. In Iteration 2 (Section 3.3.2, **E**), we implemented a contextual recommendation system. In Iteration 3 (Section 3.3.3, **G**), we refined it by incorporating additional context, such as the timing within the semester, to improve relevance and effectiveness.
- F4** Knowledge sharing during workflows: Supporting DR4, this functionality enables knowledge exchange through GoldMind’s browser extension and web-clipping features, similar to *Zotero* and *Evernote*. Introduced in Iteration 1 (Section 3.3.1), it was retained in later iterations based on co-design feedback. In Iteration 3, automatic recommendations were added, triggered by teachers’ workflows and task relevance (Section 3.3.3, **F**, **G**).

Human Factors The co-design sessions provided the guidance for the implemented functionalities. This iteration informed functionalities from the user needs derived from the DR.

5.2.2 System Evaluation (Month 12)

The functionalities described above (F1–4) were implemented for this subsequent evaluation in line with the DR. Table 5 summarises the evaluation metrics from the second iteration, comparing GoldMind with the Baseline. The results are

organised and explained according to the three evaluation factors. Metric abbreviations for this iteration are provided in the summary table for reference.

Table 5: Iteration 2: Comparison of GoldMind and Baseline performance using descriptive and statistical analyses, including 95% confidence intervals (CI) on the difference between GoldMind and Baseline and measures of improvement. For the NASA metrics, the abbreviations stand for: Mental demand (Md), Physical demand (Pd), Temporal demand (Td), Performance (P), Effort (E), and Frustration (F).

Metric	Fac.	GoldMind	Baseline	Statistical Analysis	CI (difference)	Improvement
m1 Retrieval speed	T	1.96 secs, SD = 7.07	4.24 secs, SD = 7.01	$\beta = -2.62, p = 0.000, t = -6.19$	-3.45 – -1.79	53.84% faster
m2 Relevant response ratio	H	0.62	0.38	$\beta = 3.28, p = 0.000, t = 6.31$	0.10 – 0.22	60.85% more relevant
m3 Task duration Knowledge Capture (KC) Knowledge Dissemination (KD)	H	KC: 560.47 secs, SD = 379.33 KD: 597.16 secs, SD = 315.51	KC: 618.05 secs, SD = 367.65 KD: 759.91 secs, SD = 317.20	KC: $\beta = -57.58, p = 0.40, t = -0.84$ KD: $\beta = -148.266, p = 0.009, t = -2.62$	KC: -191.66 – 76.49 KD: -259.13 – -37.40	KC: 9.3% faster KD: 21% faster
m4 Knowledge capture efficiency [low - high]	T	1: 49.65 secs, SD = 60.56 2: 49.39 secs, SD = 60.56 3: 43.42 secs, SD = 62.26	1: 80.15 secs, SD = 62.93 2: 86.16 secs, SD = 63.47 3: 52.47 secs, SD = 63.88	$\beta = -26.74, p = 0.000, t = -3.94$	-40.04 – -13.44	1: 38.05% reduction - low 2: 42.67% - mid 3: 17.24% - high
m5 Just Enough Response	T	0.29, SD = 0.29	0.08, SD = 0.17	$\beta = 1.928, p = 0.000, t = 7.29$	1.41 – 2.44	273% more useful responses
m6 Task Completion ratio	H	0.98, SD = 0.27	0.93, SD = 0.29	$\beta = 0.070, p = 0.053, t = 1.93$	-0.01 – 0.11	5.35% improvement
m7 Task completion quality [0–1]	H	0.91 score, SD = 0.22	0.85 score, SD = 0.21	$\beta = 0.050, p = 0.144, t = 0.52$	-0.001 – 0.140	6.85% slightly better
m8 NASA	H	Md: 2.56, SD = 1.26 Pd: 1.84, SD = 1.14 Td: 2.28, SD = 1.04 P: 3.95, SD = 1.05 E: 2.61, SD = 1.1 F: 2.21, SD = 1.21	Md: 2.58, SD = 1.26 Pd: 1.88, SD = 1.15 Td: 2.30, SD = 1.05 P: 3.92, SD = 1.05 E: 2.62, SD = 1.1 F: 2.23, SD = 1.22	Md: $\beta = -0.59, p = 0.000, t = -3.85$ Pd: $\beta = -0.28, p = 0.009, t = -2.59$ Td: $\beta = -0.35, p = 0.001, t = -2.53$ P: $\beta = 0.54, p = 0.001, t = 3.32$ E: $\beta = -1.15, p = 0.000, t = -8.20$ F: $\beta = -1.06, p = 0.000, t = -6.15$	Md: -0.89 – 0.29 Pd: -0.49 – 0.07 Td: 0.22 – 0.85 P: -0.20 – 1.12 E: -1.42 – 0.87 F: -1.40 – 0.72	-0.07% less Md -0.2% less Pd - 0.08% less Td 0.07% more P - 0.03% less E - 0.03% less F
m9 SUS	D	85.13, SD = 24.95	41.13, SD = 24.98	$\beta = 44.00, p = 0.000, t = 12.32$	37.10 – 50.89	100% more usable

Technological Factors: In this second iteration, GoldMind also demonstrated significant improvements over the Baseline in knowledge retrieval efficiency ($p = 0.000, t = -6.19$) m1. For Knowledge Capture Efficiency m4, aligned with DR2, this iteration evaluated the capture capabilities using different levels of information complexity (low, medium, high). GoldMind outperformed the Baseline across all complexity levels ($p = 0.00, t = -3.94$). Although the improvement for capturing high-complexity information was modest, the gains for low and medium complexities were substantial.

Users also received more relevant recommendations aligned with the task context, as indicated by Just Enough responses ($p = 0.00, t = 7.29$) (m5).

Design Factors: System usability ($p = 0.000, t = 12.32$) (m9) improved significantly, with GoldMind scoring 85.13/100, compared to the Baseline score of 41.13/100 (a 106.99% increase). This score also exceeded GoldMind’s performance in Iteration 1 (59), indicating that the design improvements informed by the co-design sessions were effective.

Human Factors: The proportion of useful responses ($p = 0.000, t = 6.31$) (m2) increased from 38% in the Baseline condition to 62% with GoldMind. Relevant response ration motivated by DR4, were evaluated based on the teacher’s level of expertise. For this analysis, participants were grouped as experts, novices, or journeyman. In the Baseline condition, the ratio of relevant responses was 36.63%, 36.54%, and 41.51% respectively. Responses metric, which improved from 8% in the Baseline to 29% in GoldMind, representing a 273% improvement. In contrast, the GoldMind condition showed significantly higher ratios – 62.16%, 62.42%, and 61.08% – with lower variance. These results indicate that GoldMind not only generated more relevant responses but also adapted more consistently across different expertise levels compared to the Baseline. This was an improvement compared to the previous iteration.

Task duration (m3) showed improvements, with participants completing capture-focused tasks more efficiently and performing dissemination tasks significantly faster under the GoldMind condition. Tasks oriented toward evaluating dissemination (KD) were completed significantly faster with GoldMind compared to the Baseline ($p = 0.009, t = -2.62$). Additionally, participants completed more tasks when using GoldMind (m6). Task quality (m7) also improved, with scores rising from 0.85 in the Baseline to 0.91 with GoldMind, indicating a 6.85% improvement and suggesting that GoldMind facilitated slightly higher-quality task outcomes.

In terms of cognitive load reduction (m8), users reported requiring less mental effort ($p = 0.000, t = -3.85$), less effort to complete knowledge management tasks ($p = 0.000, t = -8.20$), experiencing less frustration ($p = 0.000, t = -6.15$), and perceiving improved performance when using GoldMind ($p = 0.001, t = 3.32$). Figure 6 presents the responses for each question in the NASA survey.

What we learnt from this iteration: Lesson 4: Co-Design Enhances Alignment with Teachers’ Needs. The active involvement of teachers in co-design and evaluation sessions was essential for tailoring GoldMind to the real-world needs of educational teams. This collaborative approach enabled: (i) a deeper understanding of specific requirements, (ii) meaningful refinements to teachers’ interactions with the system, and (iii) the development of new features that enhanced both usability and pedagogical relevance. These improvements are reflected in the significantly higher SUS scores (m9), indicating better user experience and acceptance. Furthermore, a clearer understanding of teachers’ workflows provided valuable insights into how passive data capture (e.g., trace data) could be used to contextualise and inform the recommendation process, demonstrated by a notable improvement in the relevance of responses for task completion (m5).

Lesson 5: Reducing Cognitive Load Enhances User Satisfaction and Task Performance. The integration of AI-powered support features in GoldMind reduced users’ cognitive load by streamlining information retrieval (m1) and simplifying task execution (m3); see Section 3.3, feature E. NASA-TLX results (m8) indicated lower mental demand, reduced effort and frustration, and improved perceptions of task performance. These outcomes highlight GoldMind’s effectiveness in helping educators manage complex teaching tasks with less mental strain and greater confidence.

Lesson 6: Co-Design Informs the Need for Capturing and Sharing Expert Knowledge to Support Retention and Transfer. The co-design sessions provided valuable insights into the direction needed to support teachers’ knowledge management practices. These collaborative discussions guided the research towards addressing these needs. Building on the existing *Knowledge Capture in the Flow* capabilities (Section 3.3, C), the team iterated on alternative ways to meet these requirements. One key outcome was the conceptualisation of recording expert teachers’ workflows and transforming them into ShareFlows, structured, accessible representations designed to support novice users (Section 3.3, D). This design directly addressed the critical need for enhancing knowledge retention and transfer within educational teams, as identified in the co-design sessions. The concept of ShareFlows was further evaluated in subsequent iterations and reported in Sha et al. [4].

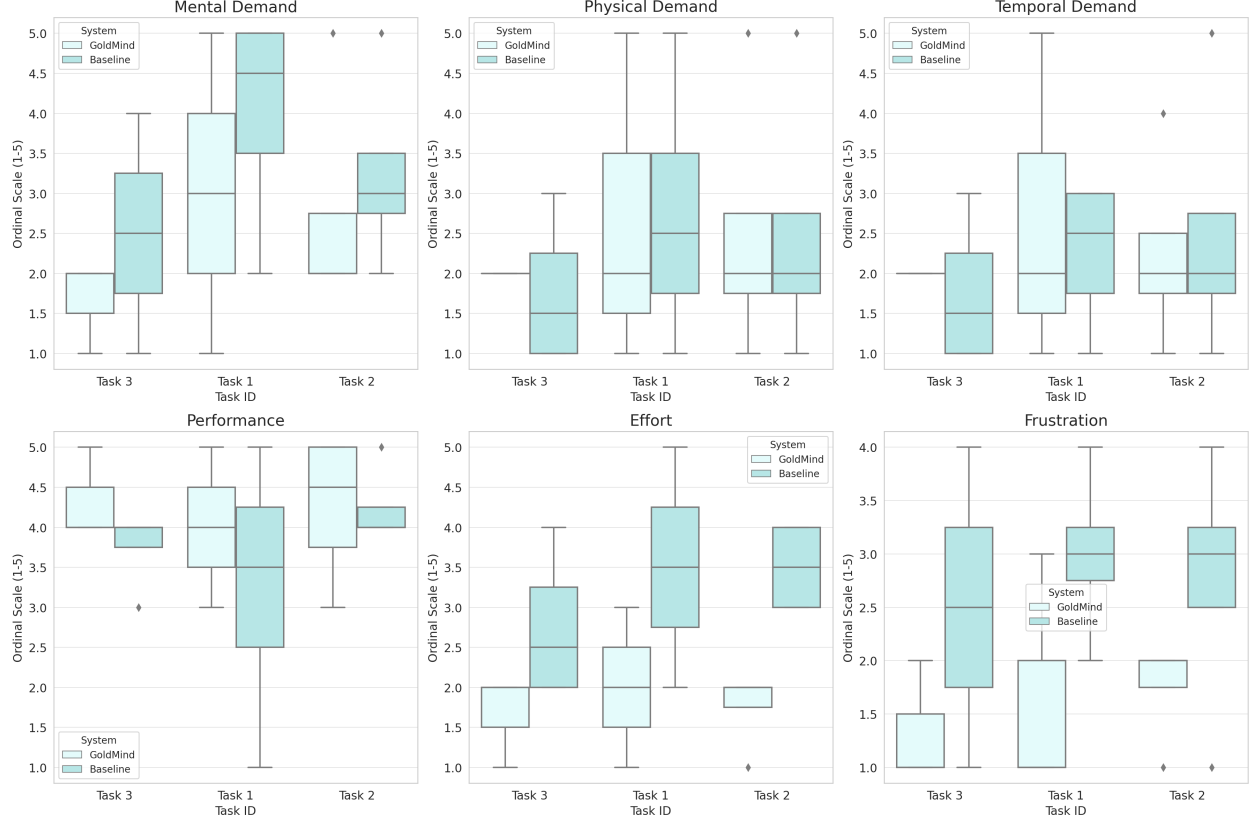


Figure 6: Results of Iteration 2 (M12) - Descriptive statistics of the NASA task load index survey responses

5.3 Iteration 3: Close to Real-World Deployment and Simplified Evaluation

The introduction of user expertise (e.g., expert and novice) in Iteration 2 motivated the use of contextualised recommendations. The success of context-aware recommendations according to the user expertise highlighted the need to simplify evaluations to real-life task performance, leading to the redefinition of evaluation metrics in Phase 3. During the co-design sessions in Iteration 2 teachers provided details on real teaching tasks that they perform regularly. Motivated by these findings, the final phase focused on completing task identified during the co-design sessions to assess how well GoldMind supported knowledge management in authentic teaching scenarios. Metrics were refined to directly measure efficiency and effectiveness in practical use cases. In addition, aiming to capture teachers’ perceptions of the KMS to support their real-world tasks, we gathered information from interviews to understand their perceptions and to inform continual improvement. This section presents lessons learnt from two iterative tool evaluations.

5.3.1 System Evaluation: new study design (Month 18)

This first evaluation was an initial test of a study with new metrics, features, and baseline. This new study design aimed to: evaluate knowledge management features; compared to a state-of-the-art Baseline tool (LLM-based Generative AI chatbot); and to simplify the evaluation process to validate effectiveness of our KMS against GenAI support. Table 7 summarises the evaluation metrics from the third iteration (Month 18), comparing GoldMind with the Baseline. The results are organised and explained according to the three evaluation factors. Metric abbreviations for this iteration are provided in the summary table for reference (gray colour).

Technological Factors: For the reduction in Knowledge Management time ^{m2}, GoldMind users retrieved relevant information 55% faster than those using the Baseline ($p = 0.00$, $t = -3.54$). This improvement is primarily attributed to GoldMind’s automated recommendations, which eliminated the need for teachers to manually query the system for support.

Table 6: Results of Iteration 2 interviews (M18). Teachers’ perceptions on GoldMind and Baseline (LLM)

Question	GoldMind	Baseline
What features of the Baseline (LLM)/GoldMind tool did you find most useful? T	<p><i>"Additional knowledge and the use of Highlights, as it was displayed when I needed it." (M18-P6).</i></p> <p><i>"Notifications were useful, but the frequency should be revised." (M18-P8).</i></p> <p><i>"Recommendations help other people learn between different users." (M18-P1).</i></p>	<p><i>"Q&A, the chance to provide additional explanations to get support." (M18-P2, P3).</i></p> <p><i>"The possibility to ask questions to locate a resource." (M18-P10, P9).</i></p>
How helpful was the information provided by the Baseline (LLM)/GoldMind for completing the tasks? GoldMind: 4 positive, 1 negative Baseline: 2 positive, 3 negative D	<p><i>"New roles (novices) would benefit from GoldMind tips" (M18-P1).</i></p> <p><i>"The push notifications are very useful, but some content seemed duplicated." (M18-P5)</i></p> <p><i>"Additional knowledge was helpful but I used it only once as I would like to complete tasks without much help." (M18-P6, P4).</i></p>	<p><i>"At the beginning I got a few good responses but then I returned to Moodle and decided to complete the task by myself." (M18-P3).</i></p> <p><i>"For a single instruction, it works well (e.g., to find where a button is located), but for complex instructions, it is not quite informative." (M18-P7).</i></p> <p><i>"It was not helpful as it always provided wrong instructions." (M18-P9).</i></p>
Did you encounter any challenges using the Baseline (LLM)/GoldMind tool? If so, please describe them. H	<p>None (M18-P1, P4, P5).</p> <p><i>"The recommendation push is a bit passive. Having multiple pushes can also be confusing. A way to highlight what needs attention would help." (M18-P6).</i></p> <p><i>"It may have some redundant instructions." (M18-P8).</i></p>	<p><i>"From the responses, it seems the chat is not smart enough to understand what you mean; it was asking me for more context." (M18-P2, P10).</i></p> <p><i>"It did not generate the whole instruction or response, or it generated late." (M18-P3, P7, P9).</i></p>

The technological factor is also informed by teachers’ perceptions, as shown in the first row of Table 6. This row provides responses to the question: “What features of the Baseline (LLM)/GoldMind tool did you find most useful?” m5. In general, for GoldMind, teachers appreciated the highlights, context-aware notifications, and shared recommendations, although some suggested that the notification frequency could be improved. For the Baseline, teachers valued the ability to ask follow-up questions and locate specific resources, but they primarily relied on it for straightforward tasks (e.g., ask questions to locate a document).

Design Factor: By integrating directly into teachers’ workflows, GoldMind minimised disruption and enhanced adoption. In this iteration, teachers’ perceptions, shown in the second row of Table 6, reflect their views on *How helpful was the information provided by the Baseline (LLM)/GoldMind for completing the tasks?* m7. For GoldMind, most teachers found GoldMind helpful, particularly for novice users, with positive comments on timely push notifications and contextual support. A few noted content duplication and a preference for working independently, which may indicate additional opportunities to improve the content of the notification. For Baseline, feedback was mixed to negative. While some found it helpful for simple queries, many felt the chatbot failed to support more complex tasks or gave inaccurate information.

Human Factors: Task duration ($p = 0.006$, $d = 1.47$) m1 significantly shorter in GoldMind, likely due to the automated support provided, which eliminated the need for additional time spent querying the system. Task completion quality m3 was also slightly higher for GoldMind. Notably, no participants failed to complete the tasks, which may be attributed to the relatively small sample size in this evaluation.

The human factor is also informed by teachers’ perceptions, as shown in the third row in Table 6. This row provides responses to the question: *Did you encounter any challenges using the Baseline (LLM)/GoldMind tool?* m6. In general, for GoldMind, most teachers reported no major issues, though some mentioned redundant content and the

Table 7: Iteration 3: Comparison between M18 and M24 Studies. GoldMind and Baseline performance using descriptive and statistical analyses, including 95% confidence intervals (CI) on the difference between GoldMind and Baseline and measures of improvement.

Metric	Fac.	GoldMind	Baseline	Statistical Analysis	CI (difference)	Improvement
M18 – New study design						
m1 Task Duration	H	2037.9 secs, SD = 449.0	3261 secs, SD = 1080.4	$\beta = -1223.1, p = 0.006, t = -2.77$	-2088.3 – -357.9	37% faster
m2 Reduction in KM time	T	209.21 secs, SD = 96.15	469.24 secs, SD = 328.8	$\beta = -0.117, p = 0.000, t = -3.54$	-0.182 – 0.053	55% less KM time
m3 Task completion quality [0–1]	H	0.91 score, SD = 0.096	0.87 score, SD = 0.071	$\beta = 0.032, p = 0.475, t = 0.714$	-0.055 – 0.119	26% better
m4 Failure rate reduction	H	No one failed	No one failed	–	–	–
M24 – Main study						
m1 Task Duration	H	708.6 secs, SD = 213.58	843.9 secs, SD = 213.29	$\beta = -62.75, p = 0.054, t = -1.92$	-126.62 – 1.11	16% faster
m2 Reduction in KM time	T	152.40 secs, SD = 139.48	274.55 secs, SD = 139.56	$\beta = -111.735, p = 0.324, t = -0.98$	-333.58 – 110.11	44% less KM time
m3 Task completion quality [0–1]	H	0.8 score, SD = 0.35	0.538 score, SD = 0.35	$\beta = 0.27, p = 0.000, t = 3.6$	0.13 – 0.43	47% better
m4 Failure rate reduction	H	0.13, SD = 0.34	0.33, SD = 0.47	$\beta = -2.33, p = 0.001, t = -3.28$	-3.72 – -0.94	60% fewer failures

need for clearer prioritisation of notifications. For the Baseline, teachers frequently struggled with vague or incomplete responses, delayed replies, and lack of understanding from the chatbot.

5.3.2 System Evaluation: main study (M24)

For this iteration, we conducted a more extensive evaluation to validate the study design tested in the first evaluation of this iteration. We recruited 30 teachers to assess GoldMind’s performance and usability metrics in the context of real teaching tasks compared to a new Baseline (GPT model). Table 7 summarises the evaluation metrics from the third iteration (M24), comparing GoldMind with the Baseline. The results are organised and explained according to the three evaluation factors. Metric abbreviations for this iteration are provided in the summary table for reference (purple colour).

Technological Factors: Our tool demonstrated a reduction in knowledge management time m2. Under the Baseline (LLM) condition, participants spent an average of 111.735 seconds more searching for the information needed to complete the task.

The technological factor is also informed by teachers’ perceptions, as shown in the first row in Table 8. This row provides responses to the question *What features of the Baseline(LLM)/GoldMind tool did you find most useful* m5. In general, for GoldMind, ShareFlows were widely appreciated for providing visual, structured task guidance with examples, relevant links, and a “show-and-tell” experience. Users liked the push notifications and centralised search, although some preferred to search independently. For the Baseline, participants positively mentioned its simplicity and the ability to revisit past queries, as they were highly familiar with chatbot interfaces. However, users often had to repeat queries, and the features lacked the depth and visual clarity provided by GoldMind.

Design Factors: Overall, for Knowledge Management usability m7, participants rated both GoldMind and the Baseline (LLM) tools highly for simplicity and ease of use. This suggests that adopting a knowledge tool — whether GoldMind or the Baseline — in the form of a web browser extension to support task workflows is positively received and can

Table 8: Results of Iteration 3 interviews (M24). Teachers’ perceptions on GoldMind and Baseline (LLM) systems

Question/Factor	GoldMind	Baseline
What features of the Baseline (LLM)/GoldMind tool did you find most useful? T	<p>"ShareFlows provide examples I can follow." (M24-P37, P35)</p> <p>"Searching in a centralised repository was useful." (P33)</p> <p>"ShareFlows provided a visual representation of the task." (P31, P26, P25, P24, P15, P12, P10, P6, P5, P2)</p> <p>"ShareFlow provided additional information such as links relevant for the task." (P31)</p> <p>"It's kind of like talking with someone next to you, and then they show me." (P25)</p> <p>"ShareFlow push will help me know where to go to complete the task." (P18)</p> <p>"I think my natural inclination is to search." (P17)</p>	<p>"The chat is quite simple to use." (M24-P37, P36, P18)</p> <p>"The chat has an active response, but I had to ask the same questions a couple of times." (P35)</p> <p>"Searching." (P33)</p> <p>"With ChatUI (LLM), I found it easy to find my previous queries." (P17)</p>
How helpful was the information provided by the Baseline (LLM)/GoldMind for completing the tasks? GoldMind: 26 positive, 3 negative, 1 unknown Baseline: 7 positive, 17 negative, 6 unknown D	<p>"As long as the task has been done by previous experts, I can just follow their steps to complete the task." (M24-P38)</p> <p>"It is helpful, but for some tasks, it is not straightforward because other people's experiences usually don't completely align with mine." (M24-P37, 36)</p> <p>"It is very helpful as I can see every step I need in the flowchart (ShareFlow) for the task." (M24-P36, P22)</p> <p>"I can see the information I need." (M24-P35)</p> <p>"I was able to access the links I needed." (P17)</p> <p>"The timing of the recommendation needs to be considered, and the amount of information should be reviewed as it can be confusing." (P34)</p> <p>"The information here is generated by experts and is trustworthy." (P33)</p> <p>"The information is accurate, but it is not possible to search within ShareFlow." (P32)</p> <p>"GoldMind was better at guiding me to begin the task." (P18)</p> <p>"I know a lot of teaching staff struggle, and having a pop-up like that with instructions on how to do things would be extremely useful for them." (P16)</p>	<p>"As long as I can type my question, I was able to get some information to continue with the task." (M24-P38)</p> <p>"I believe it is helpful, but it fails in providing links to relevant resources for the task. Although easy to use, the information could be inaccurate." (M24-P37, P32, P31, P24)</p> <p>"Not so much, as I need to think about what question I need to ask, then the response is not quite useful for what I needed." (M24-P36)</p> <p>"It is useful, but it is just like any other chat." (M24-P35)</p> <p>"Useful and easy to use." (P34, P28)</p> <p>"The information here is AI-generated and is less trustworthy." (P33, P5)</p> <p>"The information provided by ChatUI (LLM) is more personalized." (P30)</p> <p>"If you do not know how to ask your questions, it is unlikely to get a good response." (P4)</p>
Did you encounter any challenges using the Baseline (LLM)/GoldMind tool? H	<p>"GoldMind has some delays in detecting the question I have." (P30)</p> <p>"GoldMind has so many functionalities that I feel confused with the UI." (P28)</p> <p>"I could not find the history of searches or navigation." (P17, P8, P25)</p> <p>"Pulling up notifications again was an issue." (P16, P6, P11, P14, P16, P21, P23)</p> <p>"When I tried to search for some work others have done, I felt like there is too much information." (P4, P3, P20)</p> <p>"Although it provided guidance, it is not completely tailored to my own workflow." (P5, P13)</p> <p>"The guidance was not doing what I was doing." (P10)</p>	<p>"Responses were only textual and the instruction was literally asking me to click the icon, but I did not know which one." (P26, P25, P5, P10)</p> <p>"The chat is not recalling the history and the context needs to be provided." (P24, P9, P11)</p> <p>"Only provided text, links were omitted." (P22, P20, P21, P23)</p> <p>"You are required to provide context in order to get a good response." (P18, P17, P13)</p> <p>"Although easy to use, the information it gave me was completely wrong." (P16, P6, P14)</p> <p>"Some responses were not correct." (P4)</p> <p>"There were hallucinations." (P5)</p>

effectively deliver timely information. However, GoldMind outperformed ChatUI (LLM) in helping participants find accurate and trustworthy information. As a result, more teachers reported that the ShareFlow tool supported task completion. Participants also expressed greater satisfaction with GoldMind and a stronger interest in using it regularly compared to the Baseline, indicating its strong potential for adoption in practice (see additional details of the Knowledge Management SUS metrics in [2]).

In this iteration teachers’ perceptions, shown in the second row of Table 8, reflects their views on *How helpful was the information provided by the Baseline (LLM)/GoldMind for completing the task?* ^{m8}. In general, feedback for GoldMind was positive, consistent with the Month 18 evaluation. Teachers valued ShareFlows for their step-by-step guidance, visual clarity, and trustworthiness due to expert-generated content. While some noted that the experiences did not always align perfectly with their own tasks, highlighting the need to improve the precision of contextual recommendations that account for users’ expertise, most agreed that GoldMind supported task completion. For Baseline (LLM), Responses were mixed to negative. Although a few participants found it helpful for straightforward queries, many criticised its lack of context awareness, vague or inaccurate answers, and difficulty formulating effective prompts. AI-generated content was perceived as less trustworthy and often unhelpful for complex tasks.

Human Factors: Task completion time ^{m1} was shorter with GoldMind. Task completion quality ($p = 0.000$, $t = 3.6$) ^{m3} (based on rubrics in appendix) was also higher when using the GoldMind tool compared to the Baseline (LLM). The results further highlight that there were fewer task failures ^{m4} with GoldMind ($p = 0.001$, $t = -3.28$).

Teachers’ perceptions also informed the human factors in this iteration. The insights presented in the third row of Table 8 address the question: Did you encounter any challenges using the Baseline (LLM)/GoldMind tool? ^{m6}. For GoldMind, participants experienced UI complexity, difficulty finding notifications, and information overload. These perceptions could be linked to the fact that multiple features were evaluated, and the combination of learning to use a new tool while completing the task added complexity, making GoldMind feel difficult to navigate for some teachers. While the system provided guidance, teachers felt that it sometimes lacked personalisation and responsiveness to their individual workflows. However, it was also noted that GoldMind helped reduce cognitive effort. For the Baseline, key challenges included vague text-only responses, missing contextual memory, and hallucinated or incorrect information. Participants found it frustrating to repeatedly reframe their questions to receive meaningful answers. This was attributed to the tool’s inability to understand the task context on its own, requiring teachers to manually provide that context for each task in order to obtain useful responses.

To contribute to our understanding of human factors for Iteration 3, we examined teachers’ knowledge management behaviours while completing tasks using both GoldMind and the Baseline, through Epistemic Network Analysis (ENA). These visual representations illustrate how teachers engaged in knowledge management processes during Iteration 3 (Month 24), with a focus on knowledge retrieval through automated push recommendations.

The results contrast the knowledge management behaviours between the Baseline LLM-based chatbot querying system and the GoldMind system (see Section 4.3.2 for a full explanation of codes used in the ENA representation).

The first dimension of the ENA space (on the right) primarily distinguished teachers who made connections among *Querying* and *Search and Explore* activities from those who made connections to *ShareFlow Push*. This can be interpreted as a contrast between ‘pull’ behaviours (e.g., actively querying information) and ‘push’ behaviours (e.g., receiving recommendations via ShareFlow), represented on the left. This creates a clear distinction between pull (user-initiated) and push (system-initiated) knowledge management behaviours in the ENA representation.

Figure 7 shows the subtraction between conditions across tasks. It presents teachers’ knowledge management behaviours when using Baseline (LLM) and GoldMind. Thicker lines indicate more frequent connections between behaviours. For Baseline, the most common connection was from *Querying* to *Searching and Exploring*, suggesting teachers primarily used the Baseline chatbot to find information and then explore it. In contrast, GoldMind showed two prominent connections: from *ShareFlow Push with No Interaction* to both *Search and Explore* and *Read/Engage with Content*.

The average ENA scores (light salmon and dark orchid), corroborate the regression analysis in Table 9, where participants in GoldMind condition (‘pull’ behaviours) are significantly farther to the left on MR1⁵ ($t = -1.255$, $\beta = -0.56$, $p = < 2e - 16$, $d = 4.15$), indicating that the differences highlighted by the subtraction graph are statistically significant controlling for covariates.

What we learnt from this iteration:

⁵MR refers to "means rotation", which is a dimension reduction technique that maximise the variance between conditions. In this study, we did not consider the second dimension, as the rotation method ensures the mean value of the conditions is zero in that dimension [89]

A) Knowledge Management behaviours

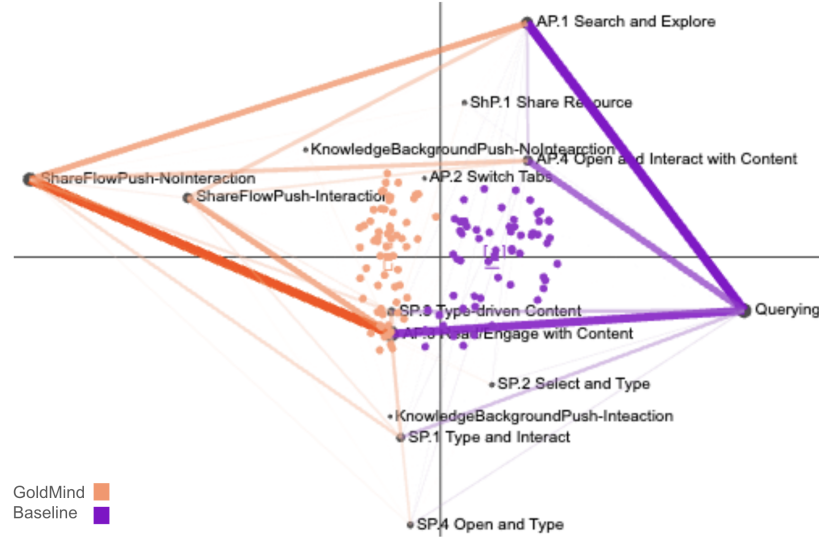


Figure 7: ENA representations of teachers' knowledge management interactions using GoldMind (light salmon), which focused on automated push recommendations, and the Baseline (dark orchid), which relied the LLM-based chatbot system for querying information.

Fixed Effect	β	Std. Error	df	t value	p-value
(Intercept)	0.328	0.046	110.75	7.154	***
SystemType (KMASS)	-0.569	0.025	89.15	-22.769	***
Task Order	-0.031	0.025	86.99	-1.255	0.212
Task Name (Task 2)	0.081	0.025	86.92	3.282	**
Task Evaluation (Pass)	-0.044	0.030	113.42	-1.466	0.145

Table 9: Fixed effects estimates from the linear mixed-effects model. Significance codes: *** $p < .001$, ** $p < .01$, * $p < .05$

Lesson 7: Process Mining and AI-Supported Systems Improve Task Efficiency and Quality. GoldMind significantly improved both task completion time ($m1$, $m1$) and task quality ($m3$, $m3$) compared to the baseline tools. These results suggest that integrating AI-powered knowledge management systems into teachers' workflows can effectively support day-to-day teaching activities. In particular, by automating teacher support via LLMs (e.g., Llama) for contextual recommendations (Section 3.3, **G**) and process mining techniques to understand experts and novices workflows (Section 3.3, **F**), GoldMind enabled teachers to perform tasks more efficiently and with higher quality outcomes.

Lesson 8: Human-Centred Design Enables Capturing and Sharing Expert Knowledge to Support Retention and Transfer. The introduction of new Knowledge Capture in the Flow mechanisms (Section 3.3, **C**) in earlier iterations enabled the recording of expert teachers' workflows and their transformation into ShareFlows (**D**). By incorporating Data Storytelling principles, ShareFlows were designed to facilitate intuitive understanding and reuse of teaching practices, demonstrating the potential of representational tools for sustained knowledge sharing.

Building on these foundations, the introduction of automated ShareFlow recommendations (**F**) illustrated the impact of human-centred design in providing timely and relevant support for teaching tasks (**G**). This feature enabled seamless knowledge sharing within teaching teams and contributed to improved task outcomes ($m3$, $m3$) and reduced failure rates ($m4$). In addition, teachers' perceptions of ShareFlows indicated that these representations were useful in helping them complete their tasks $m8$. These results underscore the value of designing tools that align with educators' workflows and support collaborative knowledge retention and transfer.

Lesson 9: Iterative Refinement and Usability Feedback Drive Adoption. The evaluation underscored the importance of iterative, user-driven co-design and usability testing in refining GoldMind. Qualitative feedback revealed the system’s strengths (e.g., usefulness and preferred features) and usability challenges, informing opportunities for further enhancement. Addressing these refinements is essential to improve user experience, increase adoption, and better support teachers in achieving their instructional goals.

Lesson 10: Automated KM support can encourage alternative behaviours: The technological support provided by GoldMind prompted richer and more diverse knowledge management behaviours. This suggests that context-aware recommendations (e.g., ShareFlows push based on task context) may lead teachers to explore alternative pathways (e.g., interactions with additional resources) when completing teaching tasks.

6 Discussion

In this section, we summarise the lessons learnt from our iterative study; then discuss the implications of these findings for practice, identify various limitations of our study, and suggest some potential directions for future research and development.

6.1 Summary of Lessons Learnt and contributions to HCI research

Table 10 summarises the 12 lessons learnt across each iteration and the ENA analysis. Based on the three main factors explored in this paper, our discussion is organised into technological, design, and human factors.

Table 10: Summary of Lessons Learnt and HCI Contributions Across Iterations

Iteration	Key Lessons Learnt	High-Level HCI Contributions
1	<p>L1. In-context knowledge capture and unified access improve efficiency and task support. (m1, m4, m7, m8)</p> <p>L2. Context-aware retrieval requires further refinement. (m6)</p> <p>L3. Alignment with teachers’ workflows and expectations needs Improvement (m9, SUS)</p>	<ul style="list-style-type: none"> - Embedded KM tools in workflows enhance efficiency, specially information retrieval for this iteration. - Early identification of user experience (UX) misfits aids design alignment
2	<p>L4. Co-design enhances alignment with teachers’ needs. (m5, m9)</p> <p>L5. Reducing cognitive load enhances user satisfaction and task performance. (m1, m3, m8)</p> <p>L6. Co-design informs the need for capturing and sharing expert knowledge to support retention and transfer. (features C, D)</p>	<ul style="list-style-type: none"> - Co-design improves educational tool usability - Cognitive load-aware interfaces enhance satisfaction - Visual representation tools (<i>ShareFlow</i>) support expert-novice knowledge transfer
3	<p>L7. Process mining and AI-supported systems improve Task Efficiency and Quality. (m1, m3, m4, F, G)</p> <p>L8. Human-centred design enables capturing and sharing expert knowledge to support retention and transfer. (m3, m4, m8, D)</p> <p>L9. Iterative refinement and usability feedback drive adoption.</p>	<ul style="list-style-type: none"> - Process mining and AI-powered, context-aware tools enhance knowledge capture and dissemination - Workflow-capture tools drive scalable knowledge reuse - Iterative feedback loops foster adoption and continuous improvement
ENA analysis	<p>L10. Automated KM support can encourage alternative behaviours.</p>	<ul style="list-style-type: none"> - Pattern variations by condition suggest opportunities for systems to better understand knowledge management behaviours and to detect user struggles, enabling timely and tailored support.

6.1.1 Technological factors

Lessons one and seven (**L1**, **L7**) advance our understanding of how technology can be effectively integrated into educational contexts, particularly through in-context knowledge capture, process mining, and AI-based task support. This aligns with the work of Di Vaio et al. [19] and Asiedu et al. [20], who argue that educational contexts are inherently complex due to their knowledge-intensive and multidisciplinary nature, and that further research is needed to implement effective technological support for educators. Our findings demonstrate that these technological approaches enhance task efficiency and provide meaningful support for educators in their daily workflows.

One major gap in the literature is that technological decisions in education are often made without adequately considering teachers' actual needs [17, 5]. Our study addresses this issue by employing a participatory design approach, which was instrumental in identifying context-specific requirements for KMSs (see **L4**). In particular, challenges related to task turnover and handover processes—highlighted in the literature as a pressing concern [25], were only revealed through close engagement with authentic teaching teams. This underscores the importance of co-design practices in the development of effective educational technologies.

Another persistent issue is the failure of AI systems to meaningfully incorporate human voices, resulting in a lack of mature, intelligent systems that genuinely align with educators' needs [20]. This was underscored in our study during Iteration 1, where **L2** and **L3** highlighted the need for ongoing refinement in system support and understanding of teachers' workflows. This lesson, along with insights from the co-design sessions, helped us recognise that KMSs are rarely embedded in teachers' everyday practices. Our findings suggest that mechanisms such as context-aware retrieval and in-context knowledge capture can help address this gap. Importantly, the effectiveness of context-aware recommendations depends on a deep understanding of domain-specific workflows — something best achieved through participatory and iterative design processes. The co-design activities in this study revealed the importance of supporting knowledge retrieval and capture based on specific contextual needs, as emphasised in **L8**. Our findings align with previous work [43], which shows that such systems can enhance both the speed and quality of task completion.

Additionally, in response to concerns about using generative AI to disseminate sensitive information [44], our study suggests that when data is aligned with specific user needs and systems offer fine-grained access control, these risks can be mitigated. Future research should explore and define privacy design principles to address the potential risks of disseminating sensitive or privileged information [44]. It should also investigate how knowledge capture and sharing can be embedded within KMSs to ensure secure and context-appropriate knowledge exchange among authorised users only.

6.1.2 Design factors

The literature highlights the need for more integrated KMSs in education, as most existing systems focus narrowly on isolated knowledge management processes like access or storage [42]. Addressing this gap, our study evaluated a KMS that supports a broader set of knowledge management processes. **L6** underscored the value of combining knowledge capture and dissemination to close *the knowledge management loop*, while **L8** demonstrated the importance of moving beyond passive storage to actively capturing and sharing expert knowledge. Through co-design with educators (**L2**), we identified the need to embed such systems into real-world workflows to support timely, relevant access to knowledge. Our findings show that educators value tools that align with their daily practices and support both task execution and professional collaboration.

Our findings underscore the value of participatory design (HCD) in shaping KMs in educational settings, consistent with calls for more inclusive and context-aware systems [22]. The participatory design approach (**L4** and **L6**) enabled us to involve teachers and teaching teams throughout the iterative design and evaluation phases (**L9**), ensuring the system reflected their real-world needs and practices. (**L9**) underscore the need for continual feedback collected from usability surveys (SUS) and user interviews (Tables 6 and 8) and further confirmed that HCD methods support the development of functionalities that are more closely aligned with teachers' expectations. These insights reinforce the need to embed end-users in the design process from the outset, not only to improve usability, but to ensure adoption and long-term value of KMSs in educational settings [14].

6.1.3 Human factors

As highlighted by Sarka et al. [14], a deeper understanding of human factors is essential for the effective design and implementation of Knowledge Management Systems (KMSs). Our iterative study addressed this gap by incorporating design considerations that accounted for users' cognitive and behavioural responses. Specifically, **L5** revealed that system complexity can increase cognitive load (Iteration 1), ultimately reducing user satisfaction. To identify such issues, we employed NASA-TLX scores and user interviews, which provided valuable insights into unforeseen complexities that may hinder usability and reduce the effectiveness of support. This evaluation process must remain ongoing, as

emphasised in **L9**; usability and user perceptions are critical human factors that should be continuously monitored to guide system refinements. The final iterations of our study underscored the need to reduce complexity in ShareFlows. This finding aligns with Wang et al. [90], who emphasised the importance of balancing contextual information, repetition, and sequencing. Future co-design of ShareFlows will aim to avoid excessive repetition and information overload, both of which can hinder usability.

An additional human factor of effective knowledge management is the requirement for acknowledging the authentic authors of knowledge [28, 44]. Our implementation of ShareFlows (Iteration 2) addressed this by allowing expert teachers to be recognised for their contributions while making their practices accessible to novices (**L6** and **L8**). This approach supports respectful and contextually relevant knowledge sharing.

In response to calls for more qualitative and human-centred approaches in KMS design [22], and in line with broader efforts to account for the complex interplay between human actors and technological systems [14], our study foregrounded the role of human factors through both methodological and design choices. We employed ENA as a quantitative ethnographic tool to analyse teachers’ knowledge management patterns, embedding these insights throughout the co-design and evaluation phases. This approach ensured that the system aligned with real-world educational practices and user needs, contributing to the development of usable, meaningful, and adoptable technologies. Beyond usability, our ENA-based analysis allowed us to examine how specific system features, particularly automated recommendations, influenced teaching behaviours. Lesson **L10**, suggest that such features may positively shape task engagement and performance, demonstrating the potential of ENA to support future investigations into how KMSs mediate and transform professional practice. By revealing how teachers interact with system features over time, the ENA analysis provided a structured understanding of behavioural patterns grounded in authentic practice. This contributes to HCD by offering an empirical basis for tailoring system functionality to real user workflows and decision-making processes.

6.2 Technological Implications: Designing for Knowledge Capture, Context, and Transfer

The GoldMind system demonstrates how existing human-centered AI and process mining technologies can be effectively applied to support knowledge capture and dissemination in real-world educational workflows. Across iterations, we observed that features such as institutional knowledge organisation (Knowledge repository feature **A**, in Section 3.3.1), in-flow knowledge capture (Features **B** and **C** in Section 3.3.2), and automated recommendations (e.g., Automated ShareFlow Recommendations **F** and Automated Contextual Recommendation **G** in Section 3.3.3) significantly enhanced knowledge management (e.g., retrieval speed, reduction in KM time) as presented in Tables 4, 5, and 7. These features improved retrieval accuracy, enabled knowledge capture in usable formats (e.g., ShareFlows feature **D**), reduced the time to access and record relevant information, and improved task performance, all of which may contribute to increased teacher productivity. These findings align with recent research on the use of GenAI in the workplace [43]. We posit that the features developed across Iterations 1 to 3 hold high promise to be similarly impactful in other domains that rely on distributed digital information and operate within structured, time-sensitive workflows (e.g., health [91], software development teams [60], scientific research contexts [92], legal information systems [3]).

A key contribution of this work is the implementation of in-the-flow knowledge capture functionalities that allow educators to annotate, summarise, and store resources without disrupting their tasks, supporting knowledge retention and transfer. ShareFlows, in particular, proved effective (see Iteration 3 results in Table 7) for capturing teaching processes and enabling knowledge transfer when automatically recommended to teachers working on similar tasks. Knowledge transfer from experts to novices remains a persistent challenge [93], and has motivated research into AI-driven solutions. For example, Rodgers et al. [46] proposed an AI-powered system to support expert decision-making in Smart Grid implementation, while Ottersböck et al. [94] explored a self-learning assistance system to mitigate knowledge loss in organisations caused by high staff turnover. Although such approaches are promising, a persistent challenge is users’ limited trust in AI-generated recommendations [44]. In our case, ShareFlows were seen as more trustworthy because teachers could identify the expert behind the recommendation (e.g., as mentioned by P33 during the interview in Iteration 3).

Furthermore, GoldMind’s knowledge repositories (e.g., integrating sources using indexing algorithms), integration of context-aware, AI-powered recommendations showed a marked improvement in user usability, relevance of information retrieved, and alignment with teachers’ needs across levels of expertise. This demonstrates the feasibility of building adaptive knowledge systems that cater to users’ cognitive load and task complexity in real-time educational environments.

6.3 Design Implications: Iterative, Participatory Approaches Lead to Usable and Trusted Tools

Our iterative HCD approach has been instrumental in teachers’ trust in GoldMind, supporting both its adoption and perceived effectiveness. Participatory design sessions with teachers revealed four critical design requirements: (1) unified access to distributed knowledge, (2) support for multimodal knowledge storage, (3) contextual and proactive recommendations, and (4) tools to support peer knowledge sharing. These requirements shaped GoldMind’s evolving feature set and interface design.

The design of the visual representations of teachers’ workflows, or ShareFlows (feature **D** in Section 3.3.2), as storytelling artefacts enabled educators to engage with knowledge in an intuitive and structured way. This was particularly helpful for novices, who could follow task workflows visually and step by step. Teachers consistently reported that these features made complex institutional knowledge more accessible and actionable (see interviews reported in Table 8), aligning with the foundations of data storytelling, which aim to make complex information easier to understand [95, 91]. Our findings also show that automated recommendations save time in task completion and improve task quality.

Our results suggest that further work is needed in system design, particularly in balancing contextual information within the visual representations of captured knowledge. This challenge is closely tied to designing a user experience (UX) that supports teachers in completing their tasks while making effective use of the system’s features. Given the complexity of teachers’ workflows and the rich sequences captured, this study provides lessons learned for the design and implementation of a KMS features (e.g., ShareFlows), while also highlighting opportunities for future refinements — such as improved pattern highlighting, contextual scaffolding, and strategies to address usability concerns."

6.4 Human factors: Understanding Behaviour, Cognitive Load, and Knowledge Practices

Not only can human-centred design improve system adoption [13, 53], but it can also reshape how users interact with, capture, and apply knowledge. Although GoldMind is still in the process of being scaled across the university, the iterative design and implementation process—as well as the findings—highlight key human factors that influence these developments.

Not only can human-centred design improve system adoption [13, 53], but it can also reshape how users interact with, capture, and apply knowledge. Although GoldMind is still in the process of being scaled across the university, the lessons learned from its iterative design and implementation highlight key human factors that influence this development. The iterative development of GoldMind illustrates how evaluating AI-powered knowledge management tools through successive co-design cycles can support teachers by adapting to their emerging needs. For example, the requirement to access knowledge from multiple repositories informed the integration of artificial intelligence for indexing and retrieval. Likewise, teacher feedback emphasised the need to maintain uninterrupted task flow (as identified during the co-design sessions and reflected in design requirements DR3 and DR4 in Section 5.2.1), which led to the optimisation of contextualised AI recommendations aligned with educators’ workflows. This iterative study underscores the importance of participatory design and real-world evaluation in HCI-driven AI research to develop solutions that are attuned to authentic educational contexts.

The ENA and trace-based analyses used to understand teachers’ knowledge management processes while interacting with KMSs exemplify how these analyses can be instrumental in identifying patterns in knowledge management processes. However, these approaches could be further enriched in future research through methods such as Contextual Think-Aloud [96], which has proven effective in capturing nuanced user experiences in complex digital environments, such as those involving software development teams.

6.5 Limitations

While our study involved 108 educators across two faculties over two years, it was conducted within a single higher education institution. Broader validation across multiple institutions and disciplines is needed to test the generalisability of our findings. Additionally, although we evaluated effectiveness using realistic digital tasks, longitudinal studies are necessary to assess how GoldMind supports ongoing knowledge retention, cultural adoption, and peer collaboration over time.

Although the evaluation was grounded in realistic tasks, it was conducted in a controlled setting. Future work should explore in-the-wild deployments that integrate seamlessly into educators’ daily teaching routines and organisational systems.

Finally, while our ENA and interaction metrics provided valuable insights into patterns of knowledge management processes, future research should investigate the social and collaborative dimensions of knowledge management, such

as peer knowledge co-construction and mentoring, to deepen our understanding of the human ecosystem surrounding KMS use. Additionally, a limitation of our ENA approach is that the coding process (outlined in Section 4.3, Table 3) was not formally validated. Typically, this involves comparing human-coded data with automated coding to assess reliability. Future studies should incorporate this validation step to uncover further insights into teachers' behaviours when using KMSs to complete their knowledge management tasks.

6.6 Conclusion

This paper presented a two-year human-centred study on the iterative design, implementation, and evaluation of GoldMind, a knowledge management system developed to support educators in higher education. Grounded in knowledge management processes and informed by real teaching tasks, our approach underscore technological, design, and human challenges that provide evidence for successful adoption and impact of knowledge management system that can be extended to not only educational settings.

Through three iterations involving 108 teachers in higher education, we identified key lessons learnt – including challenges and opportunities – in capturing, representing, and sharing expert knowledge to support teaching workflows. Our findings demonstrate the value of integrating AI technologies such as process mining and generative AI-based summarisation within a co-designed system, enabling teachers to access timely, relevant, and context-aware information. The use of visual process representations (e.g., ShareFlow), proactive knowledge recommendations, and a focus on usability contributed to more efficient task completion and improved user satisfaction.

Critically, this work illustrates how iterative, participatory, and interactive design processes not only improve the usability and effectiveness of intelligent systems but also deepen our understanding of human knowledge management processes. The analysis of teachers' knowledge management processes revealed the importance of timing, context, and trust in shaping engagement with system recommendations and support mechanisms.

By combining human-interaction and knowledge management perspectives, our study contributes practical insights into designing knowledge management systems that are adaptable, usable, and sensitive to educators' evolving needs. Future work should expand these findings through longitudinal, multi-institutional deployments and explore collaborative and cultural aspects of knowledge sharing in greater depth.

References

- [1] Qiwei He, Francesca Borgonovi, and Marco Paccagnella. Leveraging process data to assess adults' problem-solving skills: Using sequence mining to identify behavioral patterns across digital tasks. *Computers & Education*, 166:104170, 2021.
- [2] L. Sha, G. Fernández-Nieto, Y. Li, Y. Tsai, G. Chen, J. Wen, S. Singh, I. Silva, D. Gasevic, and Z. Swiecki. Shareflows: Seamless knowledge capture and proactive push for efficient teacher workflows in higher education. In *In 30th International Conference on Intelligent User Interfaces (IUI '25)*, IUI '25, page 24–27, 2025.
- [3] Mohammad Hossein Jarrahi, David Askay, Ali Eshraghi, and Preston Smith. Artificial intelligence and knowledge management: A partnership between human and ai. *Business Horizons*, 66(1):87–99, 2023.
- [4] lele sha, Gloria Milena Fernandez-Nieto, Yuheng Li, Yi-Shan Tsai, Guanliang Chen, Jim Wen, Shaveen Singh, Ivan Silva Feraud, Dragan Gasevic, and Zachari Swiecki. Shareflows: Seamless knowledge capture and proactive push for efficient teacher workflows in higher education. In *Proceedings of the 30th International Conference on Intelligent User Interfaces*, IUI '25, page 1242–1255, New York, NY, USA, 2025. Association for Computing Machinery.
- [5] Gloria Milena Fernandez-Nieto, Zachari Swiecki, Yi-Shan Tsai, Lele Sha, Yinwei Wei, Jim Wen, Yuheng Li, Yueqiao Jin, Ivan Silva Feraud, Yuan-Fang Li, Weiqing Wang, Guanliang Chen, and Dragan Gasevic. Co-designing a knowledge management tool for educator communities of practice. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*, DIS '24, page 1970–1990, New York, NY, USA, 2024. Association for Computing Machinery.
- [6] Shuangshuang Liu and Geoffrey Phelps. Does teacher learning last? understanding how much teachers retain their knowledge after professional development. *Journal of Teacher Education*, 71(5):537–550, 2020.
- [7] Deishin Lee and Eric Van den Steen. Managing know-how. *Management Science*, 56(2):270–285, 2010.
- [8] Norhaiza Binti Ishak, Uchenna C Eze, and Lew Sook Ling. Integrating knowledge management and human resource management for sustainable performance. *Journal of Organizational Knowledge Management*, 2010:1–13, 2010.

- [9] Maryam Alavi and Dorothy E Leidner. Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS quarterly*, pages 107–136, 2001.
- [10] Zhenyu Fan and Loo-See Beh. Knowledge sharing among academics in higher education: A systematic literature review and future agenda. *Educational Research Review*, 42:100573, 2024.
- [11] Ghazala Bilquise and Khaled Shaalan. Critical success factors of knowledge management in higher education: A systematic review. In Mohammed A. Al-Sharafi, Mostafa Al-Emran, Mohammed Naji Al-Kabi, and Khaled Shaalan, editors, *Proceedings of the 2nd International Conference on Emerging Technologies and Intelligent Systems*, pages 228–239, Cham, 2023. Springer International Publishing.
- [12] Bhada Yun, Dana Feng, Ace S Chen, Afshin Nikzad, and Niloufar Salehi. Generative ai in knowledge work: Design implications for data navigation and decision-making. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pages 1–19, 2025.
- [13] Joseph Giacomini. What is human centred design? *The Design Journal*, 17(4):606–623, 2014.
- [14] Peter Sarka, Peter Heisig, Nicholas H.M. Caldwell, Anja M. Maier, and Christine Ipsen. Future research on information technology in knowledge management. *Knowledge and Process Management*, 26(3):277–296, 2019.
- [15] Dhruv Galgotia and Nirupa Lakshmi. Implementation of knowledge management in higher education: A comparative study of private and government universities in india and abroad. *Frontiers in Psychology*, 13, 2022.
- [16] Khalid Rehman, Petra Poulouva, Fakhra Yasmin, Syed Arslan Haider, and Samrena Jabeen. Empirical investigation of the impacts of knowledge management on organizational learning-a case study of higher education institutions. *Academy of Strategic Management Journal*, 20:1–15, 2021.
- [17] Mariano Corso, Andrea Giacobbe, and Antonella Martini. Designing and managing business communities of practice. *Journal of knowledge management*, 13(3):73–89, 2009.
- [18] Dhruv Galgotia and Nirupa Lakshmi. Implementation of knowledge management with artificial intelligence in higher education. In *2021 IEEE 6th International Conference on Computing, Communication and Automation (ICCCA)*, pages 832–836, 2021.
- [19] Assunta Di Vaio, Rosa Palladino, Alberto Pezzi, and David E. Kalisz. The role of digital innovation in knowledge management systems: A systematic literature review. *Journal of Business Research*, 123:220–231, 2021.
- [20] Nasir Koranteng Asiedu, Mercy Abah, and De-Graft Johnson Dei and. Understanding knowledge management strategies in institutions of higher learning and the corporate world: A systematic review. *Cogent Business & Management*, 9(1):2108218, 2022.
- [21] Andreia de Bem Machado, Silvana Secinaro, Davide Calandra, and Federico Lanzalonga and. Knowledge management and digital transformation for industry 4.0: a structured literature review. *Knowledge Management Research & Practice*, 20(2):320–338, 2022.
- [22] Mohammad Fakhra Manesh, Massimiliano Matteo Pellegrini, Giacomo Marzi, and Marina Dabic. Knowledge management in the fourth industrial revolution: Mapping the literature and scoping future avenues. *IEEE Transactions on Engineering Management*, 68(1):289–300, 2020.
- [23] Rayees Farooq. A review of knowledge management research in the past three decades: a bibliometric analysis. *VINE Journal of Information and Knowledge Management Systems*, 54(2):339–378, 2024.
- [24] Obed Figueron. The influences impacting staff turnover in higher education. *J. Mgmt. & Sustainability*, 5:86, 2015.
- [25] Gunjan Sharma and DS Syal. A conceptual analysis of employee turnover intention: The causes of turnover intention in the faculties in higher educational institutions. *International Journal of Research in Finance and Management*, 5(2), 2022.
- [26] H Arun. 360 degree knowledge management. *Journal of Knowledge Management Practice*, 2005.
- [27] Rejitha Ravikumar, Abdelkarim Kitana, Abdallah Taamneh, Ahmad Aburayya, Fanar Shwedeh, Said Salloum, and Khaled Shaalan. Impact of knowledge sharing on knowledge acquisition among higher education employees. *Comput. Integr. Manuf. Syst*, 28(12):827–845, 2022.
- [28] John Carroll, Chun Choo, Dan Dunlap, Philip Isenhour, Stephen Kerr, Allan MacLean, and Mary Beth Rosson. Knowledge management support for teachers. *Educational Technology Research and Development*, 51:42–64, 12 2003.
- [29] Sangeeta Namdev Dhamdhare. Importance of knowledge management in the higher educational institutes. *Turkish Online Journal of Distance Education*, 16(1):162–183, 2015.

- [30] I. Suazo. Modelos de gestión del conocimiento en educación superior. *European Journal of Education and Psychology*, page 1–23, 2024.
- [31] Marina Du Plessis. The role of knowledge management in innovation. *Journal of knowledge management*, 11(4):20–29, 2007.
- [32] Mostafa Al-Emran, Vitaliy Mezhuyev, Adzhar Kamaludin, and Maryam ALSinani. Development of m-learning application based on knowledge management processes. In *Proceedings of the 2018 7th international conference on software and computer applications*, pages 248–253, 2018.
- [33] Mostafa Al-Emran, Vitaliy Mezhuyev, Adzhar Kamaludin, and Khaled Shaalan. The impact of knowledge management processes on information systems: A systematic review. *International Journal of Information Management*, 43:173–187, 2018.
- [34] Ranjit Bose and Vijayan Sugumaran. Application of knowledge management technology in customer relationship management. *Knowledge and Process Management*, 10(1):3–17, 2003.
- [35] Kuan Yew Wong and Elaine Aspinwall. Knowledge management implementation frameworks: a review. *Knowledge and Process Management*, 11(2):93–104, 2004.
- [36] Christian Dagenais, Didier Dupont, Frédéric N. Brière, Diego Mena, Gabrielle Yale-Soulière, and Esther Mc Sween-Cadieux. Codifying explicit and tacit practitioner knowledge in community social pediatrics organizations: Evaluation of the first step of a knowledge transfer strategy. *Evaluation and Program Planning*, 79:101778, 2020.
- [37] Tatiana Andreeva and Aino Kianto. Knowledge processes, knowledge-intensity and innovation: a moderated mediation analysis. *Journal of knowledge management*, 15(6):1016–1034, 2011.
- [38] B Rubenstein-Montano, J Liebowitz, J Buchwalter, D McCaw, B Newman, and K Rebeck. A systems thinking framework for knowledge management. *Decision Support Systems*, 31(1):5–16, 2001. Knowledge Management Support of Decision Making.
- [39] Luciana Debs and Bryan Hubbard. Gathering and disseminating lessons learned in construction companies to support knowledge management. *Construction Economics and Building*, 23, 07 2023.
- [40] Andreas Kaplan and Michael Haenlein. Collaborative projects (social media application): About wikipedia, the free encyclopedia. *Business Horizons*, 57(5):617–626, 2014.
- [41] Brad Hinton. Knowledge management and communities of practice: An experience from rabobank australia and new zealand. *International Food and Agribusiness Management Review*, 5(1030-2016-82596), 2003.
- [42] Ibrahim Arpaci, Mostafa Al-Emran, and Mohammed A. Al-Sharafi. The impact of knowledge management practices on the acceptance of massive open online courses (moocs) by engineering students: A cross-cultural comparison. *Telematics and Informatics*, 54:101468, 2020.
- [43] Erik Brynjolfsson, Danielle Li, and Lindsey Raymond. Generative ai at work. *The Quarterly Journal of Economics*, page qjae044, 2025.
- [44] Maryam Alavi, Dorothy Leidner, and Reza Mousavi. Knowledge management perspective of generative artificial intelligence (genai). *Journal of the Association for Information Systems*, pages 1–12, 2024.
- [45] Justin D. Weisz, Jessica He, Michael Muller, Gabriela Hofer, Rachel Miles, and Werner Geyer. Design principles for generative ai applications. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery.
- [46] Waymond Rodgers, Jesus A. Cardenas, Leopoldo A. Gemoets, and Robert J. Sarfi. A smart grids knowledge transfer paradigm supported by experts' throughput modeling artificial intelligence algorithmic processes. *Technological Forecasting and Social Change*, 190:122373, 2023.
- [47] Amandine Pascal, Catherine Thomas, and A. Georges L. Romme. Developing a Human-centred and Science-based Approach to Design: The Knowledge Management Platform Project. *British Journal of Management*, 24(2):264–280, 2013.
- [48] Oleg Horban. The problem of the quality of higher education in the context of knowledge management. In Paweł Poszytek, Dariusz Brakoniecki, Rostyslav Romaniuk, Aleksandra Kordonska, and Roman Kordonski, editors, *Contemporary Challenges in Education*, pages 51–66. Foundation for the Development of the Education System, 2022.
- [49] Judith Odili Uchidiuno, Jaemarie Solyst, Erik Harpstead, and Ross Higashi. “i’m a little less inclined to do it”: How afterschool programs’ culture impact co-design processes and outcomes. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, DIS '23, page 1263–1276, New York, NY, USA, 2023. Association for Computing Machinery.

- [50] Meryl Reis Louis and Jean M. Bartunek. Insider/outsider research teams: Collaboration across diverse perspectives. *Journal of Management Inquiry*, 1(2):101–110, 1992.
- [51] Stephen James Krol, Maria Teresa Llano Rodriguez, and Miguel J Llor Paredes. Exploring the needs of practising musicians in co-creative ai through co-design. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, New York, NY, USA, 2025. Association for Computing Machinery.
- [52] Hien Nguyen Ngoc, Ganix Lasa, and Ion Iriarte. Human-centred design in industry 4.0: case study review and opportunities for future research. *Journal of Intelligent Manufacturing*, pages 1572 – 8145, 2022.
- [53] Carine Edith Touré, Christine Michel, and Jean-Charles Marty. Re-designing knowledge management systems - towards user-centred design methods integrating information architecture. *ArXiv*, abs/1601.08032, 2014.
- [54] Mohammad Fakhar Manesh, Massimiliano Matteo Pellegrini, Giacomo Marzi, and Marina Dabic. Knowledge management in the fourth industrial revolution: Mapping the literature and scoping future avenues. *IEEE Transactions on Engineering Management*, 68(1):289–300, 2021.
- [55] M.N. Almunawar, M.Z. Islam, and P. Ordóñez de Pablos. *Digitalisation and Organisation Design: Knowledge Management in the Asian Digital Economy*. Routledge, 1st edition, 2022.
- [56] Lauren Wojciechowski, Emily Himes, Nelson Cheruiyot, Abigail Hamwada Turinayo, Sheila Niyonsaba, and Julia Kramer. Evaluating and developing human-centered design workshops for engagement of co-designers living and working in sub-saharan africa. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, volume 87349, page V006T06A043. American Society of Mechanical Engineers, 2023.
- [57] Hanlie Smuts and Alta Van der Merwe. Knowledge management in society 5.0: A sustainability perspective. *Sustainability*, 14(11), 2022.
- [58] Nurnaninsih A, Ahmad Mukhtar B, and Hanifah Nurul Muthmainah. Knowledge management in the digital age: Harnessing information and innovation with knowledge management systems. *The Eastasouth Journal of Information System and Computer Science*, 1(01):25–34, Aug. 2023.
- [59] Luca Nannini. Explainability in process mining: A framework for improved decision-making. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '23, page 975–976, New York, NY, USA, 2023. Association for Computing Machinery.
- [60] Jitipon Somboonpatarakit, Norranut Saguansakdiyotin, Poohridate Arpasat, and Wichian Premchaiswadi. Identifying and addressing inefficiencies in customer complaints: A process mining study with generative ai. In *2024 22nd International Conference on ICT and Knowledge Engineering (ICT&KE)*, pages 1–5, 2024.
- [61] Simon Buckingham Shum, Rebecca Ferguson, and Roberto Martinez-Maldonado. Human-centred learning analytics. *Journal of Learning Analytics*, 6(2):1–9, Jul. 2019.
- [62] Rayaguru Akshaya Kumar Das and A. B. Khan. A study of agile iterative development methodology on web application quality. In Prafulla Kumar Behera and Purna Chandra Sethi, editors, *Digital Democracy – IT for Change*, pages 19–26, Singapore, 2021. Springer Singapore.
- [63] Chirag Shah and Ryen W. White. *Task Intelligence for Search and Recommendation*. Springer Cham, 2022.
- [64] Sören Dreano, Derek Molloy, and Noel Murphy. Embed_llama: Using llm embeddings for the metrics shared task. In *Proceedings of the Eighth Conference on Machine Translation*, page 738–745, 2023.
- [65] Xiaoyang Chen, Yanjiang Liu, Ben He, Le Sun, and Yingfei Sun. Understanding differential search index for text retrieval. *Findings of the Association for Computational Linguistics: ACL 2023*, 2023.
- [66] Hans-Jörg Schulz, Marc Streit, Thorsten May, and Christian Tominski. Towards a characterization of guidance in visualization. In *IEEE Inf. Visualization Conf.*, page 1, Baltimore, MD, USA, Oct. 13-18 2013.
- [67] Brent Dykes. Data storytelling: What it is and how it can be used to effectively communicate analysis results. *Applied Marketing Analytics*, 1(4):299–313, oct 2015.
- [68] Lindy Ryan. *The visual imperative : creating a visual culture of data discovery*. Kaufmann, Morgan, Massachusetts, 2016.
- [69] Cole N. Knaflitz. *Storytelling with data: A data visualization guide for business professionals*. John Wiley & Sons, New Jersey, 12 edition, 2017.
- [70] Vanessa Echeverria, Roberto Martinez-Maldonado, Roger Granda, Katherine Chiluiza, Cristina Conati, and Simon Buckingham Shum. Driving data storytelling from learning design. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, pages 131–140. Association for Computing Machinery, mar 2018.

- [71] Sakinah S. Alhadad. Visualizing data to support judgement, inference, and decision making in learning analytics: Insights from cognitive psychology and visualization science. *Journal of Learning Analytics*, 5(2):60–85, Aug. 2018. doi: <https://doi.org/10.18608/jla.2018.52.5>.
- [72] Anja Tjärnhage, Ulrik Söderström, Ole Norberg, Mattias Andersson, and Thomas Mejtoft. The impact of scrollytelling on the reading experience of long-form journalism. In *Proceedings of the European Conference on Cognitive Ergonomics 2023, ECCE '23*, New York, NY, USA, 2023. Association for Computing Machinery.
- [73] AJMM Weijters and Joel Tiago S Ribeiro. Flexible heuristics miner (fhm). In *2011 IEEE symposium on computational intelligence and data mining (CIDM)*, pages 310–317. IEEE, 2011.
- [74] Jochen De Weerd, Manu De Backer, Jan Vanthienen, and Bart Baesens. A multi-dimensional quality assessment of state-of-the-art process discovery algorithms using real-life event logs. *Information systems*, 37(7):654–676, 2012.
- [75] Peter Ingwersen and Kalervo Järvelin. *The Turn: Integration of Information Seeking and Retrieval in Context (The Information Retrieval Series)*. Springer-Verlag, Berlin, Heidelberg, 2005.
- [76] Kassem Moustafa, Zhen Hu, Zissimos P. Mourelatos, Igor Baseski, and Monica Majcher. System reliability analysis using component-level and system-level accelerated life testing. *Reliability Engineering & System Safety*, 214:107755, 2021.
- [77] John Rooksby. Wild in the laboratory: A discussion of plans and situated actions. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 20(3):1–17, 2013.
- [78] Jocelyn Spence, Benjamin Bedwell, Michelle Coleman, Steve Benford, Boriana N Koleva, Matt Adams, Ju Row Farr, Nick Tandavanitj, and Anders Sundnes Løvlie. Seeing with new eyes: designing for in-the-wild museum gifting. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1–13, 2019.
- [79] Nick Taylor, Keith Cheverst, Peter Wright, and Patrick Olivier. Leaving the wild: lessons from community technology handovers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1549–1558, 2013.
- [80] Selma Šabanović, Sarah M Reeder, and Bobak Kechavarzi. Designing robots in the wild: In situ prototype evaluation for a break management robot. *Journal of Human-Robot Interaction*, 3(1):70–88, 2014.
- [81] James R Lewis. The system usability scale: past, present, and future. *International Journal of Human-Computer Interaction*, 34(7):577–590, 2018.
- [82] Sandra G. Hart and Lowell E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In Peter A. Hancock and Najmedin Meshkati, editors, *Human Mental Workload*, volume 52 of *Advances in Psychology*, pages 139–183. North-Holland, 1988.
- [83] Jamil Hussain, Wajahat Ali Khan, Taeho Hur, Hafiz Syed Muhammad Bilal, Jaehun Bang, Anees Ul Hassan, Muhammad Afzal, and Sungyoung Lee. A multimodal deep log-based user experience (ux) platform for ux evaluation. *Sensors*, 18(5), 2018.
- [84] Hoo Meei Hao and Azizah Jaafar. Tracing user interface design pre-requirement to generate interface design specification. In *2009 International Conference on Electrical Engineering and Informatics*, volume 01, pages 287–292, 2009.
- [85] John Brooke et al. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996.
- [86] Prokopia Vlachogianni and Nikolaos Tselios and. Perceived usability evaluation of educational technology using the system usability scale (sus): A systematic review. *Journal of Research on Technology in Education*, 54(3):392–409, 2022.
- [87] Susana Rubio, Eva Díaz, Jesús Martín, and José M Puente. Evaluation of subjective mental workload: A comparison of swat, nasa-tlx, and workload profile methods. *Applied psychology*, 53(1):61–86, 2004.
- [88] David Williamson Shaffer, Wesley Collier, and A. R. Ruis. A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics*, 3(3):9–45, Dec. 2016.
- [89] Dale Bowman, Zachari Swiecki, Zhiqiang Cai, Yeyu Wang, Brendan Eagan, Jeff Linderroth, and David Williamson Shaffer. The mathematical foundations of epistemic network analysis. In Andrew R. Ruis and Seung B. Lee, editors, *Advances in Quantitative Ethnography*, pages 91–105, Cham, 2021. Springer International Publishing.
- [90] Zezhong Wang, Shunming Wang, Matteo Farinella, Dave Murray-Rust, Nathalie Henry Riche, and Benjamin Bach. Comparing effectiveness and engagement of data comics and infographics. In *Proceedings of the 2019*

- CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–12, New York, NY, USA, 2019. Association for Computing Machinery.
- [91] Greg Waddell. Making quality improvement data more accessible and understandable: Analyst, designer, and storyteller. *Pediatric Anesthesia*, 32(11):1230–1237, 2022.
 - [92] Raymond Fok, Nedim Lipka, Tong Sun, and Alexa F Siu. Marco: Supporting business document workflows via collection-centric information foraging with large language models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery.
 - [93] Holly C Baxter. Specialized knowledge transfer: Accelerating the expertise development cycle. *Procedia Manufacturing*, 3:1465–1472, 2015.
 - [94] Nicole Ottersböck, Isabella Urban, Christian Cost Reyes, Sven Peters, and Caroline Boiteux. Employee acceptance for ai based knowledge transfer: Conception, realization and results of an elsi+ux workshop. *Procedia Computer Science*, 232:221–231, 2024. 5th International Conference on Industry 4.0 and Smart Manufacturing (ISM 2023).
 - [95] Hongbo Shao, Roberto Martinez-Maldonado, Vanessa Echeverria, Lixiang Yan, and Dragan Gasevic. Data storytelling in data visualisation: Does it enhance the efficiency and effectiveness of information retrieval and insights comprehension? In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery.
 - [96] Marta Larusdottir, Åsa Cajander, and Ruochen Wang. Digitalisation and the work environment: Insights from evaluating the contextual think-aloud method. In Anna Bramwell-Dicks, Abigail Evans, Marco Winckler, Helen Petrie, and José Abdelnour-Nocera, editors, *Design for Equality and Justice*, pages 31–46, Cham, 2024. Springer Nature Switzerland.