Making ChatGPT Work For Me

SAMANTHA KEPPLER*, Ross School of Business, University of Michigan, Ann Arbor, USA
WICHINPONG PARK SINCHAISRI, Haas School of Business, University of California, Berkeley, USA
CLARE SNYDER, Ross School of Business, University of Michigan, Ann Arbor, USA

Increasingly, work happens through human collaboration with generative AI (e.g., ChatGPT). In this paper, we present a qualitative study of this collaboration for real-life work tasks. We focus our study on US K12 public school teachers ($N = 24$) who regularly design and complete text-generation tasks such as creating quizzes, slide decks, word problems, reading passages, lesson plans, classroom activities, and projects. In one-on-one video- and audio-recorded virtual sessions, we observe each teacher using ChatGPT-4 for work tasks of their choosing for 15 minutes, then debrief their experience. Analyzing 201 prompts inputted by the 24 teachers, we uncover four main modes with which the teachers request support from ChatGPT: (1) *make for me* (55% of prompts), (2) *find for me* (15%), (3) *jump-start for me* (10.5%), and (4) *iterate with me* (15.5%). The first three modes (make, find, and jump-start) are often requests of generative AI to *do something*, whereas the fourth mode (iterate) is a request of generative AI to *think*. In a follow-up survey of the same 24 teachers, most report using multiple modes for their work, but infrequently. Our study contributes new data and knowledge about how teachers are coming to understand whether and how to integrate generative AI into their teaching preparation routines.

CCS Concepts: • **Human-centered computing → Empirical studies in collaborative and social computing**; • **Applied computing → Education**; • **Social and professional topics → K-12 education**.

Additional Key Words and Phrases: K12 education, generative AI, human-computer interaction

ACM Reference Format:
Samantha Keppler, Wichinpong Park Sinchaisri, and Clare Snyder. 2024. Making ChatGPT Work For Me. In . ACM, New York, NY, USA, 24 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Generative AI has the potential to significantly transform people's work, by saving people time on a variety of text-generation tasks [8, 44, 45] and also helping them come up with new or better ideas for creative tasks [10, 12]. Given the emerging evidence, there is a belief that generative AI will be widely integrated into people's workflows and as a consequence, work quantity and quality will improve [42, 51].

The realized value of generative AI, however, depends critically on whether and how it is taken up in practice. Its theoretical value, as evidenced in controlled laboratory experiments or simulations, may look very different than what is realized in practice if people do not use generative AI (e.g., because of algorithm aversion [16]), or use it in unexpected ways. In every wave of technological innovation, people use new technologies differently than intended by the designers who make them or the managers who implement them [36, 38, 46]. For example, consider the widespread adoption of social networking platforms within organizations [17]. The vision of these new platforms was to grow the advice networks that employees maintained and increase knowledge sharing between people in different areas of

*All authors contributed equally to this research.

Manuscript submitted to ACM

the company [34]. While this outcome was realized in some cases, there is evidence that remote workers use these technologies for the opposite purpose: to increase the distance between themselves and coworkers [37].

To understand how generative AI will ultimately impact workers and work requires a close examination of how people use the technology in practice. In this paper, we explore this in the education context, asking: *How are K12 teachers beginning to use generative AI to support their work?* The education sector stands to be affected significantly by the emergence of generative AI [1, 23]. A large portion of teacher work involves designing and completing text-generation tasks like quizzes, slide decks, word problems, reading passages, lesson plans, classroom activities, projects, or parent emails. Teachers spend many hours on such tasks weekly (e.g., US teachers work 15 hours per week above what they are contracted to do—including on material creation, as well as grading and correspondence) [9, 56], to the point where only one quarter of US K12 teachers are satisfied with the number of hours they work per week—a rate much lower than the general working population's [56]. Teachers' high workloads deter people from the profession and are the main reason why one-third of US teachers reported planning to leave their role between the 2021–22 and 2022–23 school year [7], creating crisis-level teacher shortages [28, 58]. There is hope that the problem can be alleviated in this new AI era, but whether this vision will be realized depends on whether and how teachers modify their work practices by integrating generative AI.

We intentionally sample [21] a diverse group of 24 teachers working at different grade levels and different subject areas from public K12 schools in the Midwestern US (see Table 1 for relevant details about our sample). At the time (Fall 2023), none of the teachers in our sample had used generative AI as an integrated part of their workflows, even as they reported spending hours creating materials each week. This was consistent with the large-sample evidence that teacher uptake of generative AI was still in the early stages at this time [2]. For each teacher, we conducted a 15-minute observation of their ChatGPT use, and interviewed them about their use practices. We qualitatively analyze the 201 prompts that teachers inputted during the observation period, and uncover four modes of use: (1) make for me (55% of prompts), (2) find for me (15%), (3) jump-start for me (10.5%), and (4) iterate with me (15.5%). The first three modes (make, find, and jump-start) are generally requests of generative AI to *do something*, whereas the fourth mode (iterate) is a request of generative AI to *think*.

We analyze the data at the conversation level, where we define a conversation as one or more related prompts about one particular topic; many of the 24 teachers had several conversations with ChatGPT during their observation period. We find that 66 of the 80 teacher-ChatGPT conversations in our data employ only one mode of support. Teachers only sought out multiple modes of support, typically *make for me* plus another mode, in 14 conversations. From a follow-up survey of the same 24 teachers, we learn that while most teachers use multiple modes for their work "in the wild" after the observation period, they still do so infrequently. Together, our results suggest that using generative AI for different types of support is a more complex way to use generative AI, takes more time than our 15-minute observation would allow, or both. Our prompt- and conversation-level data, supplemented by interviews and survey results, reveal novel insights about how teachers are coming to understand what generative AI can do (to help *do* and/or *think*) and how they might begin to change their routines to integrate generative AI to support their preparation work. We discuss these insights further in Section 5 of the paper.

2 RELATED WORK

Generative AI's advances are recent and rapid, therefore the body of research about this technology is similarly recent and rapidly growing. In general, this existing work falls into two categories: 1) research testing generative AI's (specifically, ChatGPT's) performance across a large and diverse range of researcher-designed prompts, and 2) research

to understand human-generative AI collaboration. This paper falls into the latter category. Below, we expand on these categories in turn, then focus on the impact of generative AI for educators.

2.1 Prompting ChatGPT

ChatGPT is currently among the most widely known and used forms of generative AI [18]. Its performance has therefore been of particular interest to researchers studying generative AI's capabilities [e.g., 4]. What they have found from testing the AI's performance on a series of pre-determined, researcher-designed prompts is that ChatGPT is effectively the digital equivalent of a Swiss Army knife; it is good at many things, but it underperforms most state-of-the-art methods (which are very good at one thing) [29]. While ChatGPT is the "jack of all trades" [29], it is not equally good at them—for example, ChatGPT is inconsistent at some forms of reasoning [3]. On the other hand, GPT-3 (a large-language model—LLM—developed by OpenAI, the company behind ChatGPT) performs well at summarizing texts [22]. Some researchers have made efforts to refine the "Swiss Army" technology for specific activities [e.g., 32], such as ideation and iteration [15], with a view towards putting a structure on what has previously been unstructured: "the power of LLMs comes from their emergence paradigm, which makes systems work in ways that haven't been anticipated when training them" ([15] page 623). However, this exercise is difficult because of the changing nature of these technologies. For some tasks, ChatGPT's output quality has depreciated over time [11].

Of course, while it is true that ChatGPT and other generative AI are imperfect, it is likewise the case that people are good but imperfect at many tasks. ChatGPT's value as a substitute or complement to human effort depends on its relative performance. In controlled experiments, ChatGPT performs nearly as well as, and in some cases better than, its human counterparts for both creative, subjective tasks and objective tasks [19, 24, 25, 27, 30, 44]. For example, ChatGPT outperforms students at writing argumentative essays [24], and performs around the passing threshold for the Bar exam [27], the US Medical Licensing Exam [30], and for many LeetCode problems [44].

2.2 Human-(Generative) AI Collaboration

To fully understand generative AI's value in practice, we must understand its role and impact within human-computer interactions (HCI). Indeed, how generative AI (such as ChatGPT) *augments* human performance has been a topic of much recent research. People have long collaborated with AI—and often struggled with this collaboration because of algorithm aversion [16]. However, human-generative AI collaboration is different; researchers argue that collaboration with *generative* AI feels more like human-human collaboration [39]. Experimental studies have shown that in many cases, this collaboration can be productive for singular, predefined tasks [12, 45, 48, 54, 57, 62, 64]. For example, subjects in an experiment are better and faster at a professional writing task after being exposed to ChatGPT [45], because ChatGPT "restructures tasks towards idea-generation and editing and away from rough-drafting" (page 1).

Yet, people's workflows are often more diverse and complex than the applications tested by experimentalists. According to affordance theory [33], it is important as a complement to conduct field research that examines holistically workers' routines and the flexibility of technologies to be used in different ways by those workers [35]. What matters for practice is whether and how workers, such as teachers, come to understand the way generative AI helps them do their work and will change their work practices to integrate generative AI.

2.2.1 (Generative) AI for Education. Educators, particularly K12 educators, present an ideal subject for such research because of the highly complex and varied nature of their work processes [20]. Teachers and education have been a popular subject of CSCW-related research even before the advent of ChatGPT and other generative AI [5, 6, 31, 61].

Unfortunately, one major takeaway from this literature is that while teachers often express the desire to take advantage of the technology available to them for sharing and developing content with others, teachers' work practices can be hard to change. Teachers sometimes struggle with discovering and remembering that technologies exist [31, 61]. In other cases, they are aware of technology but believe it is not worth the effort to adopt it. It is an open question whether teachers will be willing to change their work practices to integrate generative AI. For now, "the rate of adoption of AI in education is still slow compared to other fields" (page 4) [26] (see also [43]).

In this paper, we ask: *How are K12 teachers beginning to use generative AI to support their work?* Drawing on rich observational, qualitative, and survey data, we develop new theory about how teachers are coming to understand whether and how generative AI enables them to do their work.

3 METHODOLOGY

3.1 Participants

Our sample consists of 24 US public school teachers from the Midwestern US (Table 1). Prior evidence suggests teacher training and practices vary by subject area [e.g., 13] and grade level [e.g., 55], and moreover, ChatGPT may vary in its capabilities for different grades and subjects. Therefore, we recruited teachers from all grade levels (elementary, middle, high) and core subject areas (math, science, ELA, social studies, foreign language, general education). Five participants (indicated by * in Table 1) had pre-existing relationships with the research team. The remaining participants (19/24) were recruited through an email sent to 83 K12 public school teachers in southeast Michigan chosen based on subject area and grade level diversity goals (22% response rate). All teachers who responded to our recruitment email were included in our study.

3.2 Data Collection Procedure

Data collection happened over two phases: Phase 1, which was conducted from September through November of 2023, was a one-hour Zoom session with each teacher, that involved both observation and semi-structured interviewing. Teachers earned $50 for participating in Phase 1. Phase 2 was a five-to-ten-minute follow-up survey sent in January 2024 to the same 24 respondents. Teachers earned $10 for completing the survey (response rate 17/24, 71%).

Phase 1. We developed a three-part protocol (see Appendix A.1) to explore teachers' use of ChatGPT in practice. The first part was an open-ended conversation about teachers' material creation as part of their daily/weekly work process, and their previous exposure to generative AI such as ChatGPT. The second part was the observation stage; in this stage, teachers first logged into our ChatGPT Plus (GPT-4) account on their own computers, and then shared their screens. We requested all teachers enter the same 12 prompts (Table 2) one-by-one to ensure all teachers, even those without prior experience using the technology, had some exposure to its capabilities. We adapted these practice prompts in part from other recent studies of ChatGPT; specifically, Prompt 2—a quantitative question—comes directly from Chen et al. [11], and Prompts 6–8—writing questions—were motivated by the style of prompt designed by Noy and Zhang [45]. We developed the remainder of the prompts (1, 3–5, 9–12) to give teachers a broader view of ChatGPT's abilities, including fact-finding (e.g., Prompt 5) and planning (e.g., Prompt 11), in accordance with the open-ended nature of our simulation. After inputting the 12 prompts, we began the 15-minute, teacher-led observation period, instructing subjects to try to create relevant teaching materials with ChatGPT. The third and final part of the session was a debrief of the observation period using semi-structured questions about teachers' experience creating materials using ChatGPT.

Table 1. Teacher Participants

ID	State	Grade Level	Subject Area	Teaching Experience (years)	ChatGPT Version
T1*	Ohio	Middle School	General	12	Pre-Turbo
T2	Michigan	High School	ELA	22	Pre-Turbo
T3*	Pennsylvania	High School	Math	7	Pre-Turbo
T4	Michigan	High School	ELA	21	Pre-Turbo
T5*	Pennsylvania	High School	Foreign Language	12	Pre-Turbo
T6*	Pennsylvania	Elementary School	General	25	Pre-Turbo
T7	Michigan	Elementary School	General	6	Pre-Turbo
T8	Michigan	Elementary School	General	12	Pre-Turbo
T9	Michigan	Elementary School	General	3	Pre-Turbo
T10	Michigan	High School	Science	25	Pre-Turbo
T11*	Pennsylvania	Middle School	ELA	17	Pre-Turbo
T12	Michigan	High School	Math	22	Pre-Turbo
T13	Michigan	Middle School	Science	2	Pre-Turbo
T14	Michigan	High School	Foreign Language	11	Pre-Turbo
T15	Michigan	High School	Foreign Language	8	Pre-Turbo
T16	Michigan	High School	Foreign Language	14	Post-Turbo
T17	Michigan	High School	Science	8	Post-Turbo
T18	Michigan	High School	Math	12	Post-Turbo
T19	Michigan	Middle School	Social Studies	27	Post-Turbo
T20	Michigan	High School	ELA	20	Post-Turbo
T21	Michigan	High School	Social Studies	9	Post-Turbo
T22	Michigan	High School	Math	2	Post-Turbo
T23	Michigan	High School	Math	27	Post-Turbo
T24	Michigan	High School	Science	3	Post-Turbo

With subjects' consent, the audio and video of each Zoom call was recorded, and the audio transcribed using an automated online transcription service (Sonix). The research team watched each video recording to copy teachers' prompts into a spreadsheet, from which we created a data matrix [41]. In this data matrix, teachers were assigned an anonymous identifier that linked their ChatGPT prompts and their transcribed quotes.

Phase 1 of this study overlapped OpenAI's first DevDay, which took place November 6, 2023. On this day, OpenAI announced several new developments, among them the introduction of GPT-4 Turbo. ChatGPT Plus became an integrated multi-tool that could interpret more input formats and create more output formats. For example, after the update, ChatGPT Plus could search the internet via Bing and input information from attached documents (versus only plain texts prompts). It could output textual data in tabular format (versus only paragraphs and bullets) and also images[1]. This update was rolled out to users following the announcement, and our account updated on November 7, 2023 between 11:15am and 2:15pm. We did not have advance notice of the upgrade. As it happened, four teacher sessions (T14, T15, T16, and T17) were conducted on November 7. The first two (T14 and T15) took place in the morning and were unaffected. The later two (T16 and T17) represented our first exposure to ChatGPT Plus' new capabilities. In response, we slightly modified three of our practice prompts to give subsequent teachers (T18 through T24) exposure to its newest features. Table 2 shows the updated prompts, indicated with '.

Phase 2. In January 2024, we sent the 24 teachers who participated in Phase 1 a follow-up survey (see Appendix A.3). The survey was conducted *after* our qualitative analysis, described in the following section, was complete. The survey asked questions about the teachers' current ChatGPT use and how this compared to their ChatGPT use at the time of

[1]DALL-E was available to ChatGPT Plus users previously through a separate drop-down menu, but required switching between tools.

Table 2. ChatGPT Standardized Practice Prompts

No.	Prompt	Support Mode*
1.	What is GPT-4?	*Find for me*
2.	Is 17077 a prime number? Think step by step and then answer.	*Find for me*
3.	What are today's top news headlines?	*Find for me*
4.	What notable events happened on February 30, 2020?	*Find for me*
5.	What notable events happened on February 29, 2020?	*Find for me*
6.	Explain the economic impacts of the COVID-19 pandemic.	*Iterate with me*
7.	Help me write an introductory paragraph for an essay on this topic.	*Iterate with me*
8.	Rewrite the paragraph using simpler language.	Follow-up to *iterate with me*
9.	Summarize 'Pride and Prejudice' in one paragraph.	*Make for me*
9'.	Summarize this text in one paragraph. *(upload PDF - Chapter 43 of 'Pride and Prejudice')*	*Make for me*
10.	Please give the same summary as a rhyme.	*Make for me*
11.	Design a simple workout plan for beginners.	*Jump-start for me*
11'.	Design a simple workout plan for beginners and present it in table form.	*Jump-start for me*
12.	Design a simple workout plan for beginners with limited free time.	*Jump-start for me*
12'.	Give a diagram of the proper form for one of these exercises.	*Make for me*

*See Section 3.3 and Table 3 for details about this prompt categorization.

the interview. Specifically, the survey included Likert and open-response questions about teachers' use of ChatGPT for different purposes (as uncovered in our qualitative analysis). The survey had a response rate of 71% (17/24).

3.3 Data Analysis

We analyzed the qualitative interview and observational data from Phase 1 in four stages. In the first stage, we engaged in open coding [60], meaning we looked in the data matrix for descriptors that could point us to an emerging categorization of the way teachers sought support from ChatGPT. We examined the prompts that teachers inputted holistically and created categories like "do-for-me" and "idea generation." Finding the holistic approach was muddled; we decided in the second analysis stage to break the data down by prompt, consistent with a disassembling data approach [60]. The 24 teachers in our sample inputted 201 prompts into ChatGPT during the time we observed them. As different prompts may be seeking different kinds of support, we adjusted our data matrix to the prompt level and analyzed the data at this level. In multiple iterations, we refined our categorizations of how teachers sought support from ChatGPT through a process of repeated comparison [14]. That is, we would take pairs of prompts and discuss their similarities and differences, and their appropriate categorization. In the end, we settled on a four-category coding scheme: (1) *make for me*, (2) *find for me*, (3) *jump-start for me*, and (4) *iterate with me*. We used a fifth category, *show me what you can do*, for non-teaching-relevant requests testing ChatGPT's capabilities (4.5% of prompts). Table 3 reports our coding scheme. We describe each type of code in detail with supporting examples in our findings section.

In the third stage, we went again through the 201 prompts and systematically re-applied our four-category coding scheme. Specifically, two members of the research team independently coded all 201 prompts. The inter-rater reliability for the coding was 91% (see Table 4 in the Appendix for details about the prompts with initial rater misalignment). Through discussion among the coders, different coding categorizations were resolved and in the end, all authors were in agreement. In the fourth and final stage, we grouped the prompts into conversations or clusters of one or more prompts related to the same main idea or topic. For example, if a teacher asked multiple prompts related to teaching positive and negative numbers, these conversations would be marked as part of the same conversation. The 201 prompts cluster into

Table 3. Prompt Coding Scheme

Prompt Support Mode	Description
Make for me	Requests for fully-developed content (e.g., problems, quizzes, essays, poems, images) within well-defined parameters; user engages ChatGPT as a task executor
Find for me	Informational requests seeking pre-existing, factual information (e.g., existing facts, quotes, resources, or examples); user engages ChatGPT as a search engine
Jumpstart for me	Requests to initiate the development of often-lengthy and complex materials like activities, projects, lessons, or unit plans; user engages ChatGPT as a catalyst
Iterate with me	Requests for advice, or to understand/refine/re-think concepts or teaching approaches; user engages ChatGPT as a sounding board like they would a teaching colleague; additionally these prompts are often identifiable by words like "explain," "discuss," or "describe"

80 unique conversation (mean of 3.24 conversations per teacher and mean of 3.6 prompts per conversation). This final step enabled comparison at the teacher, prompt, and conversation level.

After this qualitative analysis was complete, we designed the follow-up survey to send to the teachers (Phase 2 of data collection). With a new typology to describe the variable usage of ChatGPT in practice, we asked teachers specifically about their use of ChatGPT to *find*, *make*, *jump-start*, and *iterate*. Importantly, only 35.29% of teachers used ChatGPT for their work more than once per month and all described themselves as novice users at the time of our Zoom session. At the time of the follow-up survey that number of monthly users increased to 64.71%. These numbers indicate that we are catching teachers at the beginning of their experience with ChatGPT while they are still learning and trying to understand how to use this new tool. We report some descriptive findings from our survey in Appendix A.4. The final stage of data analysis involved integrating the survey and interview data for a more holistic picture of these teachers' ChatGPT use patterns. In particular, we augmented our data matrix to include the survey responses, to map the usage we observe in our observations to the way that same teacher is using ChatGPT in the months following.

4 RESULTS

Over the following sections, we present results related to each of the four support modes (as described in Table 3), and the conversations teachers had with ChatGPT employing them.

4.1 Make For Me

Of the 201 prompts entered in Phase 1, 55% were *make for me* prompts and 79% of the 24 teacher participants inputted at least one *make for me* prompt. This makes it the most common type of support sought from ChatGPT during our observation period. In the follow-up survey, 69.23% of people indicated they regularly use ChatGPT for this purpose, and 46.15% of people indicated this was their favorite way to use ChatGPT.

One example of a *make for me* prompt is "Make 10 multiple choice questions using the subjunctive for the story *Mi Proprio Auto* with an answer key" (T5, High School Spanish). About the ChatGPT output to this prompt, the teacher shared, "Wow....It's crazy. I literally am flabbergasted...I'm excited. I'm going to use this for their quiz on Thursday." Another teacher (T11, Middle School ELA) asked ChatGPT for help with an ELA quiz: "Create a common and proper noun review for 5th graders." She was similarly impressed with the immediate transferrability of the output, saying "I could literally cut and paste that or screenshot it, plug it into the slides that I'm talking about and use that." A high school ELA teacher (T20) asked ChatGPT to "Write a vocabulary question with four answer selections of the meaning

of the word mendacious." The teacher expressed how hard it is for him to make these questions because it is hard to think of non-correct multiple choice options, saying "it takes me a long time to think about what those distractors would be...that was really an easy thing that [ChatGPT] can start doing for me right away."

In other cases, however, *make for me* prompts did not immediately generate usable materials. A different Spanish teacher (T14) asked ChatGPT to "Create comprehension questions in Spanish about the following text. Questions can include multiple choice, short answer, true and false, fill in the blank, and personal response. [copied text from article in Spanish]." About the questions, the teacher said "I wouldn't be able to just give that to kids right away. I'd have to look at it because [the questions] had some stuff that they don't know. I have to go back through and be like, 'oh, they don't know this vocab word, this would be really hard for them.' Or, 'oh, this is a weird question.' " A history teacher (T21) asked ChatGPT "Can you design a twenty-question, multiple-choice quiz in the AP US History style that focuses on Progressive era politics and figures?" In this case, the teacher shared, "this is not close to what an history quiz looks like, which are mostly the presentation of short passages and questions that ask students about the content of the passage, or about how the passage is related to other historical events around the time period. Almost none of this is helpful."

4.2 Find For Me

Of the 201 prompts, 15% were *find for me* prompts and 37.5% of the 24 teacher participants inputted at least one *find for me* prompt. In the follow-up survey, 53.85% of people indicated they regularly use ChatGPT for this purpose, and 23.08% of people indicated this was their favorite way to use ChatGPT.

An example of a *find for me* prompt is one inputted by a high school global studies teacher (T2): "What are examples of oracular statements from ancient Delphi?" The teacher explained she is planning an activity where she pretends she is the oracle and students ask questions, like in ancient Delphi. She shared that "The prophecies themselves were very riddle-like and unclear, which is why you had to have someone else interpret them for you. So, we [are] kind of looking for vague riddle-like statements." The teacher then followed-up with an additional prompt, "What are less famous examples?" This revealed an important feature of the search and find capabilities of ChatGPT: it can find less common results, different than other search engines that typically generate the most common results. The teacher noted this herself, "I did get a lot from ChatGPT. On Wikipedia there were certainly more. Wikipedia had lots of good stuff. It was just, you know, I'm copying from here and then scroll, scroll, scroll. Sort through all the information. I'm really just looking for again, these sort of direct quotes. So it was helpful to be able to ask ChatGPT."

A science teacher (T10) also used ChatGPT to find information, asking "How does dialysis work?" with a goal to get a brief summary she could share with her students. She was surprised by the output: "I didn't even know that there were two kinds of dialysis. I only talk about the hemodialysis. I didn't know that there was even peritoneal dialysis..." This exemplifies how *find for me* prompts can sometimes generate unexpected information in addition to the desired information.

4.3 Jump-Start For Me

Of the 201 prompts, 10.5% were *jump-start for me* prompts and 46% of the 24 teacher participants inputted at least one *jump-start for me* prompt. In the follow-up survey, 69.23% of people indicated they regularly use ChatGPT for this purpose, and 30.77% of people indicated this was their favorite way to use ChatGPT.

For example, an elementary teacher (T9) first asked ChatGPT about place values and addition. Then, she inputted the following *jump-start for me* prompt: "Help to create a lesson for first and second graders based on this topic." He explained, "I got a 35 minute math place value lesson that I don't have anything for tomorrow." Different than for *make*

for me and *find for me*, teachers do not have a specific desired output in mind when inputting *jump-start for me* prompts. Rather, they are seeking help getting started for something they are stuck on. He shared, "I'm a pretty, pretty creative guy and I'm a pretty creative thinker. But sometimes I get bogged-down in the idea and I get really scatterbrained when I'm trying to think of exactly what I want the kids to learn and what I want them to focus on." For that reason, "getting some brainstorming ideas from ChatGPT, I think would be a really useful tool for me."

Feeling stuck was a common issue among the teachers we interviewed. A middle school teacher (T19, social studies) explained she feels stuck when developing over-arching unit plans. She shared, "for our school we don't do individual day lesson plans. We do a unit planner and they have to have the key concepts and the related statement of inquiry." She inputted the following *jump-start for me* prompt: "Create an MYP IB [middle-years-program International Baccalaureate] unit planner on this PDF. [attached an outlined planning document with notes]" About the ChatGPT output she shared, "I would have to flesh this out a lot more to make it meet our administrators' standards. But this is a great start and definitely takes away all the writer's block." Moreover, from the output she got new ideas, "Oh, we should do an oral presentation. I haven't done that! And a creative project. I would take these inquiry questions. I would take the global context [part]...I would take all of this as an awesome skeleton."

In these examples, it seems that ChatGPT is saving teachers time, but perhaps not as expected. The theorized use case is that generative AI will save teachers' time by helping them finish administrative tasks, but here it saves teachers' time by helping them start creative tasks.

4.4 Iterate With Me

Of the 201 prompts, 15.5% were *iterate with me* prompts and 54% of the 24 teacher participants inputted at least one *iterate with me* prompt. In the follow-up survey, 38.46% of people indicated they regularly use ChatGPT for this purpose, but none indicated this was their favorite way to use ChatGPT (30.77% indicated this was their second favorite way).

A middle school special education teacher (T1) leaned on ChatGPT for new ideas about how to explain a math topic to her students. She inputted into ChatGPT the following *iterate with me* prompts: "Can you explain how to add a negative number and a positive number", "What manipulatives besides a number line can I use?", and then later following up with "what are key words my students should know for this?" This iteration with ChatGPT helped her in the following way: "There is a struggling student. How would I explain it? A lot of times that's where I get stuck, right? Sometimes, I explained it once and sometimes in that moment I can explain it a different way, but it's sometimes hard. This would be helpful before teaching a lesson to be like hey, this is how I normally explain it. What's another way? Because we all get stuck in our own ways of explaining things, and sometimes it's hard to get our brains out of that."

A middle school science teacher (T13) also iterated with ChatGPT to think through a science lab she was planning. She asked, "How can I use s'mores to demonstrate the law of conservation of mass?" and then later, "What are 4 different group member roles that I could use for this activity?" The teacher explained, "I just want to get a layout of how I want to do the slides and how I want to give [the students] instruction and all of that." She was impressed by ChatGPT's output: "The roles I was actually impressed with because I usually make roles for each thing. I give descriptions of what each person does. So using [ChatGPT] for roles, I didn't even think about before, but I thought it was pretty good." She explained how the iteration processes helped her: "I think it saves me some time. But for the most part, I think it saves me a stress instead. It's not like I'm all done. I can't just copy and paste this into a document. I still have to make it my own. However, once I have a clear outline, then I feel like I'm able to move more efficiently and also feel more confident about it as well." In these examples, it is clear having ChatGPT's help to think through ideas helped teachers feel more effective and confident.

4.5 Multi-Mode versus Single-Mode Conversations

Teachers entered an average of 8.4 prompts during the 15-minute observation period (min = 2, max = 17). When we group the prompts as part of the same conversation, we observe whether and how teachers move between the four modes (*make for me, find for me, jump-start for me, iterate with me*) within the same ChatGPT conversation. Five teachers (20.83%) used only one mode across all of their conversations. The average years of teaching experience for single-moders (multi-moders) are 15.2 (13.37). Single-moders also tend to have less experience with ChatGPT use than multi-moders. Of the 80 conversations in our dataset, in 66 (83%) teachers employed only one of the four modes. 33 out of 80 (41%) were exclusively *make for me* conversations (average length 2.24 prompts), 12 out of 80 (15%) were exclusively *iterate with me* conversations (average length 1.25 prompts), 8 of 80 (10%) were exclusively *jump-start for me prompts* (average length 1.25 prompts), and 8 out of 80 (10%) were exclusively *find for me* prompts (average length 1.75 prompts). *Make for me* conversations tend to be longer because teachers clarified and refined their creation requests. Among the 14 multi-mode conversations, there is large variation in the combination of modes. Four involve three modes (*jump-start, iterate, make* and *make, jump-start, find*). Ten combined two modes, and those are roughly evenly distributed among each pair-wise combination of modes. Multi-mode conversations with ChatGPT more closely resemble how someone might converse with a colleague. For example, in one three-mode conversation, Teacher 9 (Elementary School) moved from *iterate with me*, to *jump-start for me*, and finally to *make for me*. Here are the sequence of his prompts:

(1) "Explain the place value to the thousands" (*iterate with me*)
(2) "How does place value help with addition" (*iterate with me*)
(3) "Help to create a lesson for first and second graders based on this topic" (*jump-start for me*)
(4) "Make this lesson more challenging (follow-up to *jump-start for me*)
(5) "Make this lesson 35 minutes" (follow-up to *jump-start for me*)
(6) "Create large place value charts" (*make for me*)

This set of prompts exemplifies how teachers can move from a high-level idea to more specific requests for material creation, leveraging insights from ChatGPT. After this multi-mode set of prompts, the teacher shared, "I got a 35 minute math place value lesson for tomorrow, so great."

Two-mode conversations can also be dynamic. Teacher 13 (Middle School Science) oscillated between prompting ChatGPT to *make* and *iterate*.

(1) "Write a simple definition for the law of conservation of mass for my middle school science class" (*make for me*)
(2) "What do you mean by closed system? Rewrite the definition explaining that term as well" (follow up to *make for me*)
(3) "How can I use s'mores to demonstrate the law of conservation of mass? Write a list of simple instructions for middle schoolers to follow in order to complete this activity." (*iterate with me*)
(4) "What if I don't have a microwave or anything to heat it up? Will it still work?" (follow-up to *iterate with me*)
(5) "Give me 3 post-activity questions that we could discuss as a class based on this activity" (*make for me*)
(6) "What are some potential answers to these questions?" (follow-up to *make for me*)
(7) "Write a list of classroom expectations for this activity in order to manage classroom behaviors." (*iterate with me*)
(8) "Make it shorter" (follow-up to *iterate with me*)
(9) "They're not supposed to eat them tho" (follow-up to *iterate with me*)
(10) "What are 4 different group member roles that I could use for this activity?" (*iterate with me*)
(11) "How could I distribute these roles so no one fights over them?" (*iterate with me*)

This set of prompts again exemplifies how some teachers used outputs from ChatGPT to inform their future requests for support, moving from one mode of support to another, and sometimes back again.

5 DISCUSSION

This research explores how teachers are beginning to use generative AI, specifically ChatGPT, to support their significant workloads related to material design and creation. We observe 24 teachers use ChatGPT to prepare materials for their classroom, and then follow-up with these same teachers to learn about their ongoing generative AI use. We identify four modes of support by analyzing the 201 prompts from our observations: *make for me, find for me, jumpstart for me*, and *iterate with me. Iterate with me*, and in some cases *jumpstart for me*, stand apart as requests for generative AI to *think*, whereas the other modes tend to be requests for generative AI to *do something*.

Moreover, we find that in our observation period, most conversations (83%) that teachers have with ChatGPT involve only one of the four modes, suggesting that multi-mode support related to a single work objective takes significant time (more than our 15 minute observation would allow), is a more complex use of generative AI, or both. This is supported by evidence from a follow-up survey of the same 24 teachers, from which we learn that most teachers report using multiple modes for their work "in the wild"—but infrequently. Based on this evidence, we discuss the implications of our emerging *think/do* framework of generative AI reliance for how teachers, and workers more generally, are coming to understand the affordances of generative AI.

5.1 A Think/Do Framework for Worker Collaboration with Generative AI

Generative AI stands apart among technologies due to its ability to do a large variety of tasks [29]. While technologies have long been understood as flexible to workers' different intentions and designs [33], generative AI is particularly flexible and requires users to grapple with what the technology affords them to do. Our analysis helps capture this flexibility in terms of whether people are asking generative AI to *think* or to *do something*. Thinking and doing are associated with different work routines for teachers, and likely for other workers as well. In the planning stage of work, thinking is needed to determine the tasks to-be-done, while in the execution stage of work, tasks are done.

Figure 1 depicts the emerging think/do framework of generative AI within the context of work planning and execution. The top row shows how workers integrate generative AI to *do something*: the human worker does the planning stage on their own, and comes to generative AI *after* they have defined the work to-be-done (prompting generative AI to make or find, for example). The worker will need to at least verify the generative AI output. In the second row, workers integrating generative AI to support *thinking* involves different human-AI interaction. The human prompts the generative AI to *iterate with me*, or think through how to accomplish a goal (i.e., getting students to understand how to add and subtract positive and negative numbers). If the worker is using generative AI only to help them think, then once the tasks-to-be-done are decided, they complete the execution stage on their own without generative AI. The third row shows that the human and AI might work together in both the planning and execution stages. Importantly, there are different measures of success in each stage (shown at the bottom of Figure 1). In the planning stage the objective is to gain *task clarity* about the work to-be-done, whereas in the execution stage the objective is *output usability*.

According to affordance theory [33], people construct perceptions about whether a technology constrains their ability to achieve their goals, or affords them the possibility of achieving new goals, through their experiences with the technology. Our study shows that generative AI for *thinking* and *doing* are two different perceptions that teachers are developing of generative AI affordances. The ability of generative AI to think and specifically help *plan* work has implications for knowledge creation, notions of expertise and knowledge ownership, and knowledge relationships

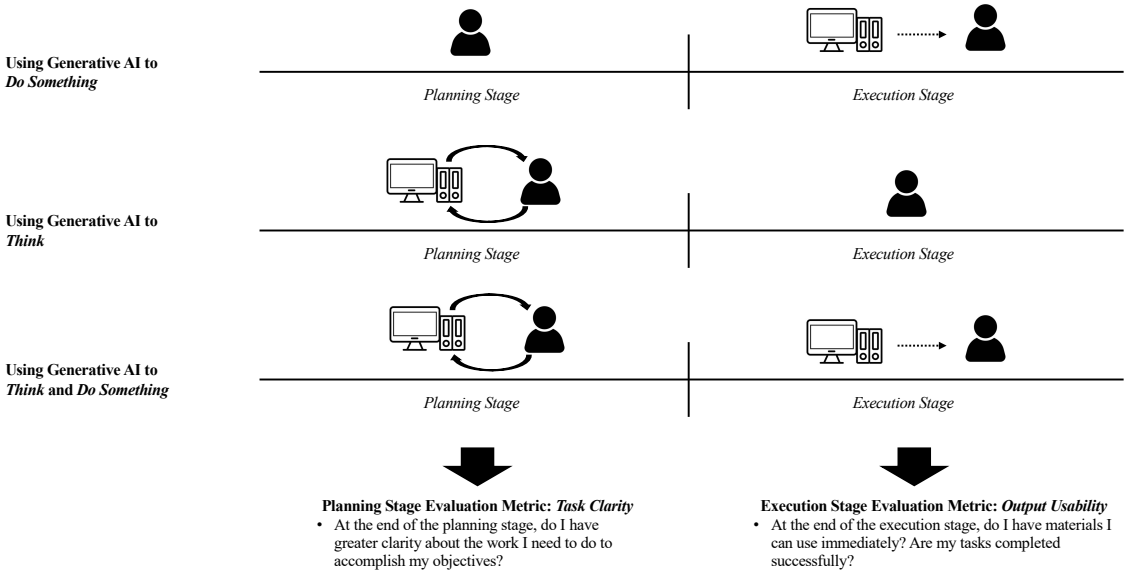


Fig. 1. Thinking and Doing with Generative AI

between workers, as highlighted in parallel research on generative AI affordances [49, 50]. The ability of generative AI to *do something*, however, elevates different affordances, such as the ability to work more quickly or even automate tasks, but also the constraint of unsatisfactory output that requires re-work. Moreover, another way to think of generative AI's affordances is to consider its role for *both thinking and doing*. Using generative AI in multiple work stages requires, for example, workers to use the technology in different ways and for different purposes at different times. Overall, our emerging think/do framework contributes to the nascent effort to understand generative AI affordances [40, 49, 50].

5.2 Teacher Work and Thinking/Doing with Generative AI

Our emerging think/do framework of human collaboration with generative AI has specific implications for teacher work. There is a widespread belief that generative AI will dramatically affect teacher work by making teachers more productive or create more value [1, 23]. The findings from this study motivate a closer examination of how teacher work and productivity might be impacted differently depending on the way teachers integrate generative AI into their workflows. Teachers have the choice to integrate generative AI in any of the three ways shown in Figure 1, and how they do it has implications for educational organizing and outcomes.

Comparing our research to recent studies about university faculty using generative AI for their work [47, 63], we seem to identify an important difference between university and K12 teaching work: K12 teachers quickly iterate between planning and executing, often revising their teaching plans weekly or even daily, whereas for university faculty, the planning and execution stages are longer and less iterative. Thus, the uses and productivity potential of generative AI may look different in K12 versus higher education. Moreover, given the recent interest in UI about designing generative AI tools [59], our findings suggest a need for different generative AI tools for K12 versus higher education. More than that, our framework suggests it might be important to have different UIs for *thinking* versus

doing. Right now, however, it is unclear if different generative AI tools for planning versus doing would help or hinder workflow integration.

It is important to note the ethical concerns about generative AI in general, and in education in particular [40, 52, 53]. Our think/do framework is useful for theorizing AI ethics. The ability of generative AI to *do* tasks, perhaps with minimal to no human involvement, for example, raises concerns about cheating or avoiding responsibilities. Meanwhile, the ability of generative AI to *think* raises different concerns, such as about safety and security. Therefore, there may be an entanglement of affordances and ethical concerns when it comes to generative AI: the better the technology is at thinking and doing, the more ethically problematic it is to use. Overall, as education scholars, education leaders, and policymakers navigate the productivity and ethical issues related to generative AI, the three workflows shown in Figure 1 and the four use cases identified through our research can be a useful guiding framework.

5.3 Theoretical versus In-Practice Use of Generative AI

Our emerging think/do framework also points to how use of generative AI in practice may be similar and different from what we observe in simulations or experimentally-controlled studies of generative AI use. Our study reinforces the notion that users will implement generative AI to directly create materials [12, 45]. *Make for me* prompts were the most common use case in our observation period, and are most highly rated as the favorite use case in the months following our observations. Eight of the 24 teachers mentioned how the creation feature of ChatGPT would save them considerable time in their work. Yet, no one reported in Phase 2 using ChatGPT frequently, or on a weekly basis. This is true although every teacher reported in Phase 1 spending multiple hours per week on material creation. The teachers in our study, for the most part, are not yet adapting their work practices to integrate generative AI.

A key insight our study reveals about generative AI in practice is the *multi-modal* nature of human-generative AI interactions. As we describe above in Section 4.5, in our observations, some teachers in our study use ChatGPT for one type of support, and then immediately afterward prompt ChatGPT for an entirely different type of support (e.g., Teacher 9 moves from *iterate with me* prompting to *jump-start for me* prompting to *make for me* prompting within one conversation). This variety touches on the unique flexibility of ChatGPT's capabilities as a "Swiss Army knife." Yet, our finding that teachers during the observation period typically employ only one mode per conversation suggests it might be difficult for them to fully realize these advantages without sufficient experience or time. This is supported by evidence from a follow-up survey of the same 24 teachers, from which we learn that most teachers report using multiple modes for their work "in the wild"—but still infrequently.

5.4 Limitations and Future Directions

Like any study about an emerging technology, our research has several limitations related to the particular generative AI we study (ChatGPT Plus) and subjects' familiarity with this technology. ChatGPT Plus was the most cutting-edge version of ChatGPT and the most well-known generative AI technology at the time of data collection, and this motivated our choice to study it. While OpenAI remains an industry leader, there are more alternatives (e.g., Google's Gemini) and more specialized tools (e.g., SchoolAI) emerging. Generative AI tools trained on more relevant corpora may be better for K12-specific needs like lesson planning, and teachers interactions with these technologies could be very different. As generative AI tools evolve, so too do the ethical concerns about their use, particularly in education [40, 52, 53]. While we have non-users who describe their ethical concerns, a limitation of this paper is that it does not fully explore reasons for non-use.

It is also important to keep in mind that our sample is not representative. We conduct a small-scale qualitative study, which was most appropriate for our research question. However, the findings cannot be generalized to the wider population of K12 teachers. Specifically, most teachers in our sample had minimal experience using ChatGPT or any generative AI technology, and none are power-users. It is possible that teachers would shift their conversation styles or modes of use with more experience. The value in our sample is that we capture a diverse group of teachers by grade and subject area who help illuminate the way that at least some teachers are beginning to use generative AI. Overall, our study motivates future studies that examine a broader range of K12 teachers and further examine the ethical issues related to AI use, especially those raised by non-users in our study, in K12 education.

6 CONCLUSION

We studied 24 US K12 teachers' early interactions with ChatGPT to understand how they are beginning to use generative AI to support their work. Analyzing these inputs at the prompt- and conversation-level, we develop a think/do framework that captures an important dichotomy for how teachers might come to integrate generative AI into the planning and execution stages of their workflows. We see that teachers use four modes of prompting: *iterate with me*, and in some cases *jump-start for me*, as they plan the work they need to do, and then *find for me*, *jump-start for me*, and *make for me* as they complete tasks at hand. The distinction between *thinking* versus *doing* with generative AI has important implications for how to design generative AI tools to support teacher work and how to support teachers as they adapt their routines and work practices to integrate generative AI.

REFERENCES

[1] Tufan Adiguzel, Mehmet Haldun Kaya, and Fatih Kürşat Cansu. 2023. Revolutionizing education with AI: Exploring the transformative potential of ChatGPT. *Contemporary Educational Technology* 15, 3 (2023), ep429.

[2] Kara Arundel. 2023. Teachers increasingly embrace ChatGPT — students not so much. *K–12 Dive* (2023). https://www.k12dive.com/news/teachers-use-of-ChatGPT-increases/688129/

[3] Yejin Bang, Samuel Cahyawijaya, Nayeon Lee, Wenliang Dai, Dan Su, Bryan Wilie, Holy Lovenia, Ziwei Ji, Tiezheng Yu, Willy Chung, Quyet V. Do, Yan Xu, and Pascale Fung. 2023. A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity. arXiv:2302.04023 [cs.CL]

[4] David BAÏDOO-ANU and Leticia OWUSU ANSAH. 2023. Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning. *Journal of AI* 7, 1 (2023), 52–62. https://doi.org/10.61969/jai.1337500

[5] Nadia Bouz-Asal, Rieko Inaba, and Toru Ishida. 2011. Analyzing Patterns in Composing Teaching Materials from the Web. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work* (Hangzhou, China) *(CSCW '11)*. Association for Computing Machinery, New York, NY, USA, 605–608. https://doi.org/10.1145/1958824.1958925

[6] Thomas Breideband, Jeffrey Bush, Chelsea Chandler, Michael Chang, Rachel Dickler, Peter Foltz, Ananya Ganesh, Rachel Lieber, William R Penuel, Jason G Reitman, et al. 2023. The Community Builder (CoBi): Helping Students to Develop Better Small Group Collaborative Learning Skills. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing*. 376–380.

[7] Jake Bryant, Samvitha Ram, Doug Scott, and Claire Williams. 2023. K–12 Teachers Are Quitting. What Would Make Them Stay? *McKinsey & Company, March* (2023).

[8] Erik Brynjolfsson, Danielle Li, and Lindsey R Raymond. 2023. *Generative AI at work.* Technical Report. National Bureau of Economic Research.

[9] Daniel A. Camacho, Sharon A. Hoover, and Hazel S. Rosete. 2021. Burnout in urban teachers: The predictive role of supports and situational responses. *Psychology in the Schools* 58, 9 (2021), 1816–1831. https://doi.org/10.1002/pits.22561

[10] Tuhin Chakrabarty, Vishakh Padmakumar, Faeze Brahman, and Smaranda Muresan. 2023. Creativity support in the age of large language models: An empirical study involving emerging writers. *arXiv preprint arXiv:2309.12570* (2023).

[11] Lingjiao Chen, Matei Zaharia, and James Zou. 2023. How is ChatGPT's behavior changing over time? arXiv:2307.09009 [cs.CL]

[12] Zenan Chen and Jason Chan. 2023. Large language model in creative work: The role of collaboration modality and user expertise. *Available at SSRN 4575598* (2023).

[13] Julie Cohen, Erik Ruzek, and Lia Sandilos. 2018. Does teaching quality cross subjects? Exploring consistency in elementary teacher practice across subjects. *AERA Open* 4, 3 (2018), 2332858418794492.

[14] Juliet Corbin and A Strauss. 2008. Basics of Qualitative Research 3e (V. Knight, S. Connelly, L. Habib, K. Wiley, & G. Treadwell Eds.).

[15] Giulia Di Fede, Davide Rocchesso, Steven P. Dow, and Salvatore Andolina. 2022. The Idea Machine: LLM-Based Expansion, Rewriting, Combination, and Suggestion of Ideas. In *Proceedings of the 14th Conference on Creativity and Cognition* (Venice, Italy) *(C&C '22)*. Association for Computing Machinery, New York, NY, USA, 623–627. https://doi.org/10.1145/3527927.3535197

[16] Berkeley J Dietvorst, Joseph P Simmons, and Cade Massey. 2015. Algorithm aversion: people erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General* 144, 1 (2015), 114.

[17] Nicole B Ellison, Jennifer L Gibbs, and Matthew S Weber. 2015. The use of enterprise social network sites for knowledge sharing in distributed organizations: The role of organizational affordances. *American Behavioral Scientist* 59, 1 (2015), 103–123.

[18] Dennis Foung, Linda Lin, and Julia Chen. 2024. Reinventing assessments with ChatGPT and other online tools: Opportunities for GenAI-empowered assessment practices. *Computers and Education: Artificial Intelligence* 6 (2024), 100250. https://doi.org/10.1016/j.caeai.2024.100250

[19] Catherine A. Gao, Frederick M. Howard, Nikolay S. Markov, Emma C. Dyer, Siddhi Ramesh, Yuan Luo, and Alexander T. Pearson. 2023. Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. *npj Digital Medicine* 6 (2023), 1–5. https://doi.org/10.1038/s41746-023-00819-6

[20] Sukhpal Singh Gill, Minxian Xu, Panos Patros, Huaming Wu, Rupinder Kaur, Kamalpreet Kaur, Stephanie Fuller, Manmeet Singh, Priyansh Arora, Ajith Kumar Parlikad, Vlado Stankovski, Ajith Abraham, Soumya K. Ghosh, Hanan Lutfiyya, Salil S. Kanhere, Rami Bahsoon, Omer Rana, Schahram Dustdar, Rizos Sakellariou, Steve Uhlig, and Rajkumar Buyya. 2024. Transformative effects of ChatGPT on modern education: Emerging Era of AI Chatbots. *Internet of Things and Cyber-Physical Systems* 4 (2024), 19–23. https://doi.org/10.1016/j.iotcps.2023.06.002

[21] Barney Glaser and Anselm Strauss. 1967. *Discovery of grounded theory: Strategies for qualitative research*. Routledge: London, UK.

[22] Tanya Goyal, Junyi Jessy Li, and Greg Durrett. 2023. News Summarization and Evaluation in the Era of GPT-3. arXiv:2209.12356 [cs.CL]

[23] Simone Grassini. 2023. Shaping the future of education: exploring the potential and consequences of AI and ChatGPT in educational settings. *Education Sciences* 13, 7 (2023), 692.

[24] Steffen Herbold, Annette Hautli-Janisz, Ute Heuer, Zlata Kikteva, and Alexander Trautsch. 2023. A large-scale comparison of human-written versus ChatGPT-generated essays. *Scientific Reports* 13 (2023). https://doi.org/10.1038/s41598-023-45644-9

[25] Fan Huang, Haewoon Kwak, and Jisun An. 2023. Is ChatGPT Better than Human Annotators? Potential and Limitations of ChatGPT in Explaining Implicit Hate Speech *(WWW '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 294–297. https://doi.org/10.1145/3543873.3587368

[26] Enkelejda Kasneci, Kathrin Sessler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeier, Stephan Krusche, Gitta Kutyniok, Tilman Michaeli, Claudia Nerdel, Jürgen Pfeffer, Oleksandra Poquet, Michael Sailer, Albrecht Schmidt, Tina Seidel, Matthias Stadler, Jochen Weller, Jochen Kuhn, and Gjergji Kasneci. 2023. ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences* 103 (2023), 102274. https://doi.org/10.1016/j.lindif.2023.102274

[27] Daniel Martin Katz, Michael James Bommarito, Shang Gao, and Pablo Arredondo. 2023. Gpt-4 passes the bar exam. *Available at SSRN 4389233* (2023).

[28] Alyson Klein. 2023. What's Keeping People From Becoming Teachers? An Eye-Popping To-Do List, for One. *Education Week* (2023). https://www.edweek.org/leadership/whats-keeping-people-from-becoming-teachers-an-eye-popping-to-do-list-for-one/2023/09

[29] Jan Kocoń, Igor Cichecki, Oliwier Kaszyca, Mateusz Kochanek, Dominika Szydło, Joanna Baran, Julita Bielaniewicz, Marcin Gruza, Arkadiusz Janz, Kamil Kanclerz, Anna Kocoń, Bartłomiej Koptyra, Wiktoria Mieleszczenko-Kowszewicz, Piotr Miłkowski, Marcin Oleksy, Maciej Piasecki, Łukasz Radliński, Konrad Wojtasik, Stanisław Woźniak, and Przemysław Kazienko. 2023. ChatGPT: Jack of all trades, master of none. *Information Fusion* 99 (2023), 101861. https://doi.org/10.1016/j.inffus.2023.101861

[30] Tiffany H Kung, Morgan Cheatham, Arielle Medenilla, Czarina Sillos, Lorie De Leon, Camille Elepaño, Maria Madriaga, Rimel Aggabao, Giezel Diaz-Candido, James Maningo, et al. 2023. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLoS digital health* 2, 2 (2023), e0000198.

[31] Mackenzie Leake and Colleen Lewis. 2016. Designing a New System for Sharing Computer Science Teaching Resources. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion* (San Francisco, California, USA) *(CSCW '16 Companion)*. Association for Computing Machinery, New York, NY, USA, 321–324. https://doi.org/10.1145/2818052.2869109

[32] Mina Lee, Percy Liang, and Qian Yang. 2022. Coauthor: Designing a human-ai collaborative writing dataset for exploring language model capabilities. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–19.

[33] Paul M Leonardi. 2011. When flexible routines meet flexible technologies: Affordance, constraint, and the imbrication of human and material agencies. *MIS quarterly* (2011), 147–167.

[34] Paul M Leonardi. 2013. When does technology use enable network change in organizations? A comparative study of feature use and shared affordances. *MIS quarterly* (2013), 749–775.

[35] Paul M Leonardi. 2017. Methodological guidelines for the study of materiality and affordances. In *The Routledge Companion to Qualitative Research in Organization Studies*. Routledge, 279–290.

[36] Paul M Leonardi and Stephen R Barley. 2010. What's under construction here? Social action, materiality, and power in constructivist studies of technology and organizing. *The Academy of Management Annals* 4, 1 (2010), 1–51.

[37] Paul M Leonardi, Jeffrey W Treem, and Michele H Jackson. 2010. The connectivity paradox: Using technology to both decrease and increase perceptions of distance in distributed work arrangements. *Journal of Applied Communication Research* 38, 1 (2010), 85–105.

[38] Alex Jiahong Lu, Tawanna R Dillahunt, Gabriela Marcu, and Mark S Ackerman. 2021. Data work in education: Enacting and negotiating care and control in teachers' use of data-driven classroom surveillance technology. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–26.

[39] Xiaoyue Ma and Yudi Huo. 2023. Are users willing to embrace ChatGPT? Exploring the factors on the acceptance of chatbots from the perspective of AIDUA framework. *Technology in Society* 75 (2023), 102362. https://doi.org/10.1016/j.techsoc.2023.102362

[40] Nigel P Melville, Lionel Robert, and Xiao Xiao. 2023. Putting humans back in the loop: An affordance conceptualization of the 4th industrial revolution. *Information Systems Journal* 33, 4 (2023), 733–757.

[41] Matthew B Miles and A Michael Huberman. 1994. *Qualitative data analysis: An expanded sourcebook*. sage.

[42] Ethan Mollick. 2022. ChatGPT is a tipping point for AI. *Harvard Business Review* (2022).

[43] Fatemeh Mosaiyebzadeh, Seyedamin Pouriyeh, Reza Parizi, Nasrin Dehbozorgi, Mohsen Dorodchi, and Daniel Macêdo Batista. 2023. Exploring the Role of ChatGPT in Education: Applications and Challenges. In *Proceedings of the 24th Annual Conference on Information Technology Education* (Marietta, GA, USA) *(SIGITE '23)*. Association for Computing Machinery, New York, NY, USA, 84–89. https://doi.org/10.1145/3585059.3611445

[44] Nascimento Nathalia, Alencar Paulo, and Cowan Donald. 2023. Artificial Intelligence vs. Software Engineers: An Empirical Study on Performance and Efficiency using ChatGPT. In *Proceedings of the 33rd Annual International Conference on Computer Science and Software Engineering*. 24–33.

[45] Shakked Noy and Whitney Zhang. 2023. Experimental evidence on the productivity effects of generative artificial intelligence. *Available at SSRN 4375283* (2023).

[46] Wanda J Orlikowski and Susan V Scott. 2015. The algorithm and the crowd. *MIS quarterly* 39, 1 (2015), 201–216.

[47] Gustavo Pinto, Isadora Cardoso-Pereira, Danilo Monteiro, Danilo Lucena, Alberto Souza, and Kiev Gama. 2023. Large language models for education: Grading open-ended questions using chatgpt. In *Proceedings of the XXXVII Brazilian Symposium on Software Engineering*. 293–302.

[48] Aditya kumar Purohit, Aditya Upadhyaya, and Adrian Holzer. 2023. ChatGPT in Healthcare: Exploring AI Chatbot for Spontaneous Word Retrieval in Aphasia. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing* (Minneapolis, MN, USA) *(CSCW '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 1–5. https://doi.org/10.1145/3584931.3606993

[49] Laavanya Ramaul, Paavo Ritala, and Mika Ruokonen. 2024. Creational and conversational AI affordances: How the new breed of chatbots are revolutionizing the knowledge industries. *Business Horizons* (2024).

[50] Jana Retkowsky, Ella Hafermalz, and Marleen Huysman. 2024. Managing a ChatGPT-empowered workforce: Understanding its affordances and side effects. *Business Horizons* (2024).

[51] David Rotman. 2023. ChatGPT is about to revolutionize the economy. We need to decide what that looks like. *MIT Technol. Rev.* (2023).

[52] Nazmus Sakib, Fahim Islam Anik, and Lei Li. 2023. ChatGPT in IT Education Ecosystem: Unraveling Long-Term Impacts on Job Market, Student Learning, and Ethical Practices. In *Proceedings of the 24th Annual Conference on Information Technology Education*. 73–78.

[53] Daniel Schiff. 2022. Education for AI, not AI for education: The role of education and ethics in national AI policy strategies. *International Journal of Artificial Intelligence in Education* 32, 3 (2022), 527–563.

[54] Hanieh Shakeri, Carman Neustaedter, and Steve DiPaola. 2021. SAGA: Collaborative Storytelling with GPT-3 *(CSCW '21 Companion)*. Association for Computing Machinery, New York, NY, USA, 163–166. https://doi.org/10.1145/3462204.3481771

[55] David J Shernoff, Suparna Sinha, Denise M Bressler, and Lynda Ginsburg. 2017. Assessing teacher education and professional development needs for the implementation of integrated approaches to STEM education. *International Journal of STEM Education* 4 (2017), 1–16.

[56] Elizabeth D. Steiner, Ashley Woo, and Sy Doan. 2023. *All Work and No Pay — Teachers' Perceptions of Their Pay and Hours Worked: Findings from the 2023 State of the American Teacher Survey*. RAND Corporation, Santa Monica, CA. https://doi.org/10.7249/RRA1108-9

[57] Xin Sun, Emiel Krahmer, Jan De Wit, Reinout Wiers, and Jos A Bosch. 2023. Plug and Play Conversations: The Micro-Conversation Scheme for Modular Development of Hybrid Conversational Agent. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing*. 50–55.

[58] Leib Sutcher, Linda Darling-Hammond, and Desiree Carver-Thomas. 2016. A Coming Crisis in Teaching? Teacher Supply, Demand, and Shortages in the US Research Brief. *Learning Policy Institute* (2016).

[59] Justin D Weisz, Jessica He, Michael Muller, Gabriela Hoefer, Rachel Miles, and Werner Geyer. 2024. Design Principles for Generative AI Applications. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–22.

[60] Robert K Yin. 2015. *Qualitative research from start to finish*. Guilford publications.

[61] Alexey Zagalsky, Joseph Feliciano, Margaret-Anne Storey, Yiyun Zhao, and Weiliang Wang. 2015. The Emergence of GitHub as a Collaborative Platform for Education. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) *(CSCW '15)*. Association for Computing Machinery, New York, NY, USA, 1906–1917. https://doi.org/10.1145/2675133.2675284

[62] Mingyuan Zhang, Zhaolin Cheng, Sheung Ting Ramona Shiu, Jiacheng Liang, Cong Fang, Zhengtao Ma, Le Fang, and Stephen Jia Wang. 2023. Towards Human-Centred AI-Co-Creation: A Three-Level Framework for Effective Collaboration between Human and AI *(CSCW '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 312–316. https://doi.org/10.1145/3584931.3607008

[63] Yong Zheng. 2023. ChatGPT for teaching and learning: an experience from data science education. In *Proceedings of the 24th Annual Conference on Information Technology Education*. 66–72.

[64] Qingxiaoyang Zhu and Hao-Chuan Wang. 2023. Leveraging Large Language Model as Support for Human Problem Solving: An Exploration of Its Appropriation and Impact. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing* (Minneapolis, MN, USA) *(CSCW '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 333–337. https://doi.org/10.1145/

3584931.3606965

A APPENDIX

A.1 Phase 1 Session Protocol

Part 1 (10-15min): Introduction

Tell me about the types of materials you create on your own for your class every week.

- Every day? Every month? At the start of the school year?
- How long does it take you?
- How do you feel about creating this stuff? (probe: Love? Hate? Scale of 1- 10?)
- Do you get materials from anyone else – TeachersPayTeachers? Another teacher?
- Have you ever used ChatGPT/generative AI to help you create any of these materials?
- Is there a school policy about ChatGPT use by teachers?

Part 2 (25 min): Observation

ChatGPT practice (10 min)

(log in to ChatGPT)

Please copy the following prompts (one at a time) into ChatGPT.

(1) What is GPT-4?
(2) Is 17077 a prime number? Think step by step and then answer.
(3) What are today's top news headlines?
(4) What notable events happened on February 30, 2020?
(5) What notable events happened on February 29, 2020?
(6) Explain the economic impacts of the COVID-19 pandemic.
(7) Help me write an introductory paragraph for an essay on this topic.
(8) Rewrite the paragraph using simpler language.
(9) Summarize 'Pride and Prejudice' in one paragraph. [Update: Summarize this text in one paragraph. *(upload PDF - Chapter 43 of 'Pride and Prejudice')*]
(10) Please give the same summary as a rhyme.
(11) Design a simple workout plan for beginners. [Update: Design a simple workout plan for beginners and present it in table form.]
(12) Design a simple workout plan for beginners with limited free time. [Update: Give a diagram of the proper form for one of these exercises.]

Open-Ended Observation (15 min)

Pick one of the things you mentioned earlier (in Part 1) for which you might use ChatGPT to help, and create whatever it is from scratch. Work as if you are trying to create the "finished product" in 15 minutes. You are welcome to use other technology in addition to ChatGPT such as Google Docs, Word, Excel, a web browser, etc. It's okay if you are unable to finish, just work like you'd typically work. Remember, the final product may be included in a publication as an example of how teachers use ChatGPT, so please try your best. [give 3-minute warning when time is almost up]

Part 3 (10-15 min): Debrief

- Tell me about were you creating.
- Describe what you were thinking about before using ChatGPT.

- Describe what you were thinking while using ChatGPT.
- What was the quality of ChatGPT's output? (probe: Love? Hate? Scale of 1- 10?)
- (If not finished) Describe what else you would do to finish.
- Would you use ChatGPT in practice for something like this? How similar/different is simulation from reality?
- Based on your experience, how useful would ChatGPT be for you in practice?
- Any further reflections / thoughts / questions?

A.2 Misalignment in Rater Prompt Coding

As we describe in Section 3.3 of the paper, two raters coded all 201 prompts, and 9% of their codes were initially misaligned. In this table, Table 4, we present all the prompts for which the two raters' initial coding were misaligned, as well as the final support mode coding—chosen following discussion between the coders—and the alternative support mode considered in that discussion.

Table 4. Support mode coding for instances of initial rater misalignment, with details about the final and alternative coding.

ID	Prompt	Final Support Mode	Alternative/Overridden Support Mode
T4	Describe novels comtemporary to Adventures of Huckleberry Finn that reflect similar social and cultural issues.	*Iterate with me*	*Find for me*
T4	Describe novels comtemporary to The Great Gatsby that reflect similar social and cultural issues.	*Iterate with me*	*Find for me*
T4	Describe essays, pamphlets, and books contemporary to The Great Gatsby that could inform a student's understanding of ther Jazz Age and/or class stratification in the United States in the early 20th century.	*Iterate with me*	*Find for me*
T8	visual resource for oham	*Find for me*	*Make for me*
T10	Explain why strawberries get wet when they are sprinkled with sugar.	*Iterate with me*	*Find for me*
T10	change the percent to molarity	Follow-up to *jumpstart for me*	*Find for me*
T12	What are the key steps to finding global extrema using calculus?"	*Iterate with me*	*Find for me*
T12	What are the key steps for the concavity of a graph using calculus?"	*Iterate with me*	*Find for me*
T13	Write a list of classroom expectations for this activity in order to manage classroom behaviors.	*Iterate with me*	*Make for me*
T13	Make it shorter	Follow-up to *iterate with me*	Follow-up to *make for me*
T13	They're not supposed to eat them tho	Follow-up to *iterate with me*	Follow-up to *make for me*
T16	*In Mandarin:* Give me some pictures about artificial intelligence	*Make for me*	*Show me what you can do*
T16	*In Mandarin:* Can you give me an article based on these pictures?	*Find for me*	*Make for me*
T18	Can you expand on week 3? What are some good examples?	*Iterate with me*	*Find for me*
T22	Create an activity on financial literacy for high school students	*Make for me*	*Jumpstart for me*
T22	How do I keep students off their phones?	*Iterate with me*	*Jumpstart for me*
T23	Solve the inequality |4x -1| > 2x +5, showing steps.	*Make for me*	*Show me what you can do*
T24	what is this not aligned with in the content policy for images?	*Show me what you can do*	*Make for me*

A.3 Phase 2 Survey Questions

(1) What is your name? (First and Last)
(2) At the time of our interview, how often did you use ChatGPT for your work?
    - Never
    - Occasionally (about once every few months)
    - Sometimes (about once a month)
    - Often (a few times times per month)
    - Always (about weekly or more frequently)
(3) During our interview, did you have clear goals or visions of what you would use ChatGPT for during the open-ended material creation stage? For example, were you preparing for materials that you'd soon have to make anyways?
    - No - I did not have clear goals of what to make with ChatGPT; just wanted to try
    - I had some ideas but there were no specific or concrete materials I was trying to make
    - Yes - I was trying to make / prepare materials that I could use in my upcoming class / the near future
    - I don't remember.
(4) How would you rate the outputs generated by ChatGPT during our interview?
    - Likert scale from 1–5, where 1 indicates "Really bad/not useful" and 5 indicates "Really good/useful."
(5) Did that (the quality of the outputs) match your expectation?
    - Likert scale from 1–5, where 1 indicates "Much worse than my expectation" and 5 indicates "Much better than my expectation."
(6) Currently, how often do you use ChatGPT for your work?
    - Never
    - Occasionally (about once every few months)
    - Sometimes (about once a month)
    - Often (a few times times per month)
    - Always (about weekly or more frequently)

Display logic: if "Never" is selected for Question 6.

(7) Please describe why you do not use ChatGPT for your work.
(8) Are there any functions/features you wish it has that would encourage you to use ChatGPT?
(9) Please rate how useful you think ChatGPT would be for each of these four common functions people use ChatGPT for:
    (a) to jumpstart your new project/task
    (b) to make or write things for you
    (c) to iterate and work through ideas
    (d) to search for and find information
    - For each function, a Likert scale from 1–5, where 1 indicates "Not useful" and 5 indicates "Very useful."
(10) "Jumpstart for me": Please describe what you think of the use of ChatGPT to jumpstart your new project or task.
(11) "Make for me": Please describe what you think of the use of ChatGPT to make or write things for you.
(12) "Iterate for me": Please describe what you think of the use of ChatGPT to iterate and work through ideas for your project or task.

(13) "Find for me": Please describe what you think of the use of ChatGPT to search for and find information.

*Display logic: if "Never" is **not** selected for Question 6.*

(7) How often do you use ChatGPT to...
   (a) ....jumpstart your new project/task
   (b) ...make or write things for you
   (c) ...iterate and work through ideas
   (d) ...search for and find information
   - For each function, a Likert scale from 1–5, where 1 indicates "Never" and 5 indicates "Always"

(8) If there are functions of ChatGPT that you use but missing here, please state them and describe how/how often you use ChatGPT for those functions.

(9) Rank your favorite "functions" of ChatGPT (1 = most favorite/useful and 4 = least favorite/useful)
   (a) ....jumpstart your new project/task
   (b) ...make or write things for you
   (c) ...iterate and work through ideas
   (d) ...search for and find information

(10) "Jumpstart for me": Please describe how you may have used or what you think of the use of ChatGPT to jumpstart your new project or task.

(11) "Make for me": Please describe how you may have used or what you think of the use of ChatGPT to make or write things for you.

(12) "Iterate for me": Please describe how you may have used or what you think of the use of ChatGPT to iterate and/or work through ideas.

(13) "Find for me": Please describe how you may have used or what you think of the use of ChatGPT to search for and find information.

Final questions, for all responses.

(14) Do you have any closing thoughts on your experience and/or views about ChatGPT since the interview? If so, please describe them here.

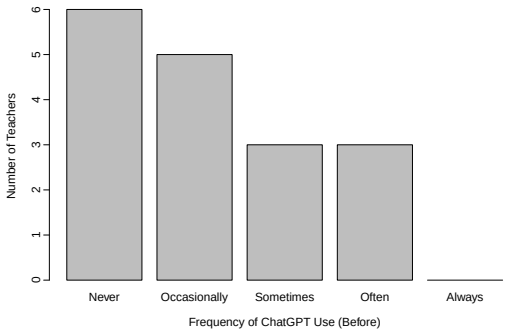
(15) Please list any other AI tools that you use for your work.

(16) *Questions regarding payment details.*

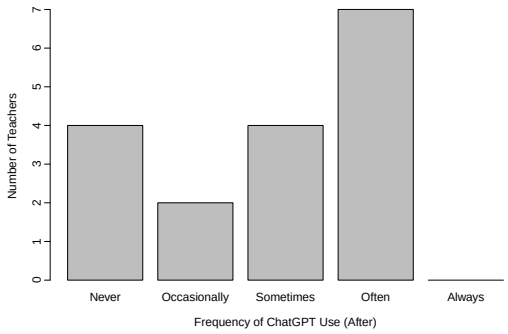
A.4 Phase 2 Survey Findings

17 teachers completed the follow-up survey. Figures 2a through 2c depict the comparison of before versus after frequencies of ChatGPT use. While there seems to be a slight increase in frequency, no one reached the "always" frequency. Four teachers self-reported that they had not been using ChatGPT at all since the interviews and offered different reasons. T6 wanted to use ChatGPT, but "the current policy in [their] school district discourages use with student or instructional related materials" and called for a more open policy. T22 stated that they "still feel somewhat at a loss on how to use [ChatGPT] effectively" and wished that the tool could perform "more creative lesson planning." T2 reported that, besides their lack of habit of using ChatGPT, the results generated by ChatGPT during the interview were not adequate to meet their needs: "I then looked at Wikipedia, and I found much more specific, concise information. So, in the side-by-side comparison, Wikipedia was a better resource." Finally, T21 decided not to use ChatGPT as they perceived it as "a broadly predatory system that steals content and credits no one"; they would consider using it if the

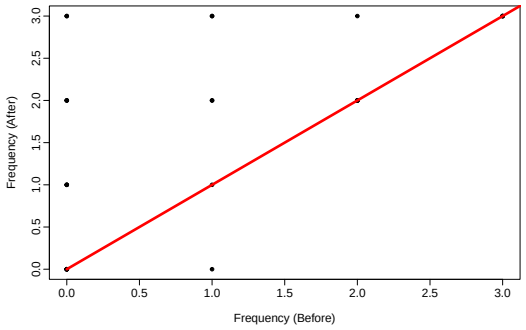
tool could consistently provide clear information sources. Figures 2e and 2e suggest that most teachers were satisfied with the quality of ChatGPT output, meeting or exceeding their expectations, in Phase 1.



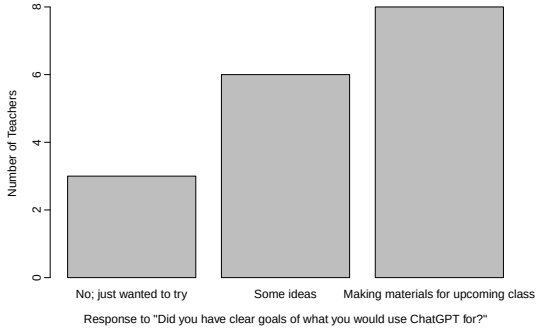
(a) Reported Frequency of ChatGPT Use (Before Phase 1)



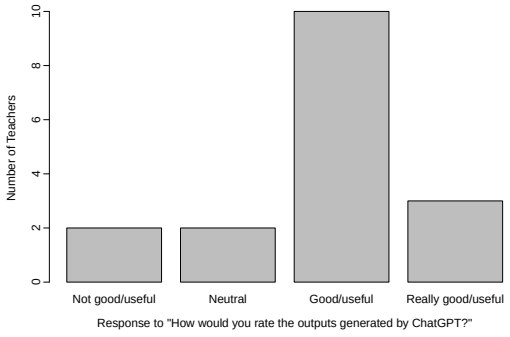
(b) Reported Frequency of ChatGPT Use (After Phase 1)



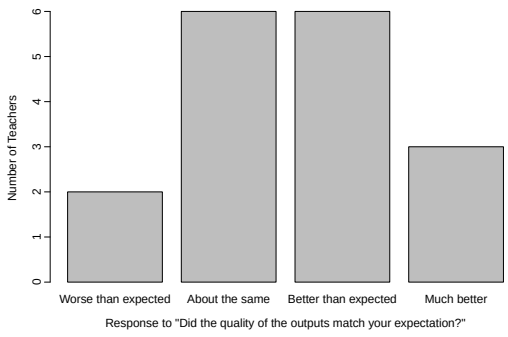
(c) Reported Frequency of ChatGPT Use (Before vs. After)



(d) Whether They Had Clear Goals



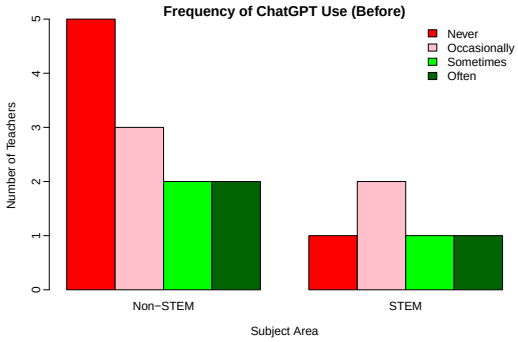
(e) Evaluation of ChatGPT Output in Phase 1



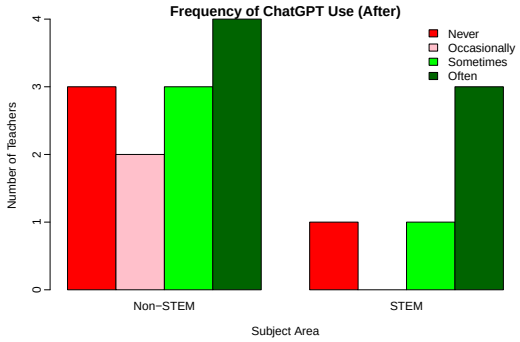
(f) ChatGPT Actual vs. Expected Quality

Fig. 2. Individual Teacher Responses to Phase 2 Survey.

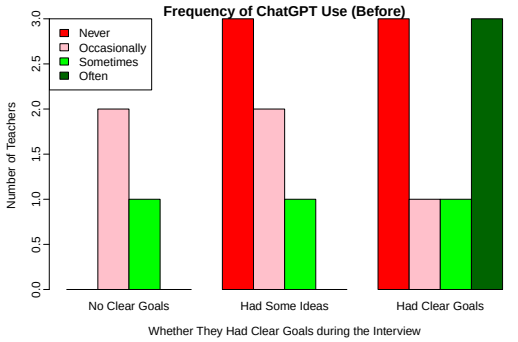
Figures 3 offer deeper insights into frequency of ChatGPT use and evaluation of ChatGPT output.



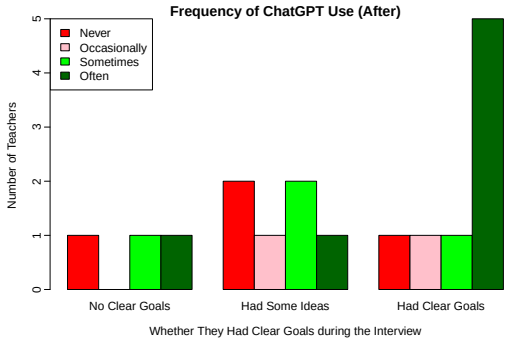
(a) Reported Frequency of ChatGPT Use (Before Phase 1) by Subject Area



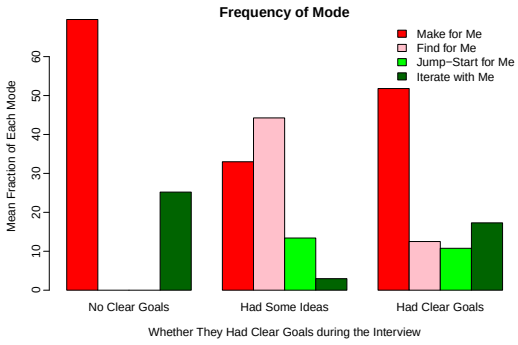
(b) Reported Frequency of ChatGPT Use (After Phase 1) by Subject Area



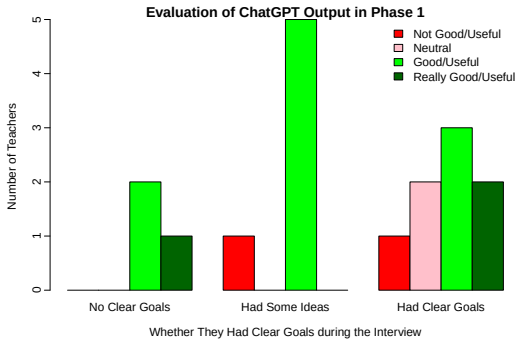
(c) Reported Frequency of ChatGPT Use (Before Phase 1) by Whether They Had Clear Goals



(d) Reported Frequency of ChatGPT Use (After Phase 1) by Whether They Had Clear Goals



(e) Average Fraction of Each Mode within Prompts by Whether They Had Clear Goals



(f) Evaluation of ChatGPT Output in Phase 1 by Whether They Had Clear Goals

Fig. 3. Individual Teacher Responses to Phase 2 Survey.

Received 16 July 2024