Backwards Planning with Generative AI: Case Study Evidence from US K12 Teachers

Samantha Keppler
Stephen M. Ross School of Business, University of Michigan

Wichinpong Park Sinchaisri
Hass School of Business, University of California Berkeley

Clare Snyder
Stephen M. Ross School of Busines, University of Michigan

Abstract. Backward planning is an effective and efficient operational process when working towards a goal: work backward from the desired outcome to figure out the steps needed to accomplish them in the time allowed. While many organizations and workers might use it, backward planning is a universal practice among US K12 teachers. The emergence of generative AI has stimulated many conversations about its impact on teacher work, but it is still unclear if and how generative AI fits within the backward planning approach adopted by most every teacher. Given backward planning is standard workflow process in K12 education, we ask: How are teachers using generative AI to support their teaching work? Our methodology is a case study of 24 US public school teachers, sampled to vary by subject area and grade level, during the 2023–2024 school year. We conduct interviews, observations, and surveys at different points in time to understand their evolving generative AI use. In fall 2023, all teachers were novice users or had never tried generative AI. By spring 2024, the teachers separate into three distinct groups: (1) those who seek generative AI input (i.e., thoughts or ideas about learning plans) and output (i.e., quizzes, worksheets), (2) those who only seek generative AI outputs, and (3) those not using generative AI. The teachers in the first group—but not the second group—report productivity gains in terms of workload and work quality. Our findings have implications for understanding how to integrate generative AI into backward, goal-oriented workflows.

Key words: generative AI; productivity; workflow; education operations; algorithm aversion

1. Introduction

Knowledge about how generative AI will impact people's work is quite limited. Most studies to date have examined how people's performance (work quality and speed) are affected by generative AI on a given task, such as creating press releases (Noy and Zhang 2023), resolving a customer problem (Brynjolfsson et al. 2023), or deriving business recommendations (Dell'Acqua et al. 2023). However, the reality is that work tasks do not exist in isolation. They are part of someone's *workflow*.

Workflows are very different from job to job. For some, like doctors and nurses in maternity wards, workflows evolve endogeneously as information is gathered and the tasks-to-be-done next are decided (Freeman et al. 2017). In other cases, workers know exactly the tasks-to-be-done from the start, but the workflow may vary with the sequence in which the tasks are done (Ibanez et al.

2018). Customer-facing workflows might be exogeneously determined by customer arrivals (Aksin et al. 2007). Another case still is that workers plan their own workflows by working backward from a specified goal or target (Wiggings and McTighe 2005).

The role and impact of generative AI on work may vary considerably in different workflows. In workflows that involve only task execution, for example, it seems intuitive to try to use generative AI to speed up task execution without sacrificing quality. Yet, in workflows that involve deciding which tasks to do, in addition to doing those tasks, the role of generative AI is not yet clear. We do not know whether and how generative AI might help with task planning. Moreover, there is little understanding of what it might mean for productivity if generative AI is used to help complete tasks that are poorly planned. As scholars and practitioners alike try to understand how generative AI will impact human work, it is critical to understand how people are using generative AI *within the context of their real workflows* (Jaffe et al. 2024).

In this paper, we explore generative AI use within the backward-planned workflows of US K12 teachers. Backward planning is the process of devising teaching plans that are well-aligned to learning objectives and goals in order to avoid aimless activities and "coverage" of topics (Wiggings and McTighe 2005). In other words, backward planning starts with objectives like "I want my students to understand..." and then moves to figuring out the learning activities needed to achieve that objective and the assessments that will be evidence of achieving that objective (or not). This is different than, for example, brainstorming activities and plans without formalizing and checking for alignment with learning objectives. Backward planning is the standard best practice in US K12 education, where the learning goals are defined by state learning standards, such as the Common Core State Standards (Common Core State Standards 2010). Given backward planning is standard workflow process in K12 education, we ask: *How are teachers using generative AI to support their teaching work?*

To explore this question, we closely follow 24 public school teachers from the upper Midwestern United States during the 2023–2024 school year. We selected these teachers for variation by subject area and grade level. They also vary in years of experience from 2 to 27 years. We interview these 24 teachers, expose them to a generative AI tool (ChatGPT Plus), and observe directly their early uses of it in Fall 2023. Then, we follow up with the same teachers with surveys during the winter and spring of 2024 to understand their evolving generative AI use. We end data collection with follow-up interviews of a subset of the original teachers in June 2024. The result is a rich longitudinal dataset of real teacher experiences with generative AI during the beginning of this transformational era. Each teacher represents one case of teacher-generative AI use and experiences. We analyze the data qualitatively through human coding of teacher-composed generative AI prompts and comparative

analyses about their task versus workflow level use to unearth rich insights about generative AI use, and non-use, among this subset of teachers.

We find that all teachers' use of and thoughts about generative AI were quite similar at the beginning of 2023–2024, but that they diverged significantly over the course of the school year. In fall 2023, all teachers were either very novice users or had never tried any generative AI tool. They had heard of generative AI, and ChatGPT in particular, but had not tried much yet to use it for their teaching work. They also all received the same exposure to ChatGPT through our study: a standardized practice session with 12 required prompts and then an unstructured practice where they could come up with their own prompts and use ChatGPT for a teaching task of their choice, under our observation. By spring 2024, the teachers separate into three distinct groups: (1) those who seek generative AI input (i.e., thoughts or ideas about learning plans) *and* output (i.e., quizzes, worksheets), (2) those who only seek generative AI outputs but not for input into their teaching plans, and (3) those not using generative AI. We find that the teachers in the first group—but not the second group— report productivity gains in terms of workload and work quality. The divergence among teachers seems to derive from different responses (attraction versus aversion) to the two capabilities of generative AI: the ability to *generate* content (create output) and the ability to learn something from its *intelligence* (provide input).

Our finding contribute rich qualitative evidence about task *and* workflow level use of generative AI to the emerging scholarship about how generative AI may impact human work (Brynjolfsson et al. 2023, Noy and Zhang 2023, Dell'Acqua et al. 2023, Jaffe et al. 2024). Our findings also have implications for practice. To school and education leaders, our findings suggest that within backward planning, the ability of generative AI to provide teachers with *outputs* may have limited productivity upside. Generating materials with support from generative AI may even counter the well-respected logic of backward planning as it makes it easier to create activities without checking their alignment with end goals. The productivity potential of generative AI for teachers may instead lie with leveraging the tool for *input* about teaching plans. When teachers seek input from generative AI they are seeking additional certainty and confidence that they are doing the right and best activities to achieve the desired learning objectives, by for example asking "What manipulatives besides a number line can I use to help my students understand how to add positive and negative numbers?" (T1).

For leaders and managers more broadly, our findings illuminate how the productive potential of generative AI does indeed depend on workflow. Fast and high-quality output generation is an advantage to productivity when the tasks-to-be-done are well-established, as shown by Brynjolfsson et al. (2019) and Noy and Zhang (2023). However, when tasks are steps towards a larger goal and those steps may vary from worker-to-worker, situation-to-situation, or over time, the value

of generative AI's content creation capabilities is less certain. The ability of generative AI to *provide input*, however, is more clear. Planning is hard work that requires expertise and experience (particularly when the success of those plans are dependent on children's behaviors, effort, and choices). Workflow planning can also be high stakes, and *input* from generative AI can help people consider alternative workflows than what they would have typically done and therefore help them gain confidence about their plans.

A final implication of our research is that aversion to generative AI is a major barrier to its adoption and thus productivity gains. Regardless of the eventual stance taken by companies and governments as to whether and how generative AI ought to be limited (Hacker et al. 2023), our evidence reveals aversion exists to generative AI's fundamental abilities, particularly its ability to create outputs. Moreover, some teachers in our study explain that the upside of generative AI to provide input cannot be technologically decoupled from its more problematic ability to create outputs. One may claim that this perspective will become outdated. People may change their views when there is a broader cultural shift toward acceptance of generative AI; the technology is still very new. Yet, another path forward is to think about how to design and transform this new technology so that at least in some cases—like when there are goals and objectives for learning–we can achieve a more desirable balance between generative AI's input and its outputs.

2. Literature Review

This research contributes to three bodies of research.

2.1. Backward Planning

Our study contributes broadly to operations management research about planning and designing workflows for creative tasks (Krishnan and Ulrich 2001). Backward planning is one effective approach to designing workflows; it is the gold standard within education contexts because it emphasizes student outcomes (Wiggings and McTighe 2005). The planning that goes into workflow design is non-trivial. People often struggle to adequately plan and allocate time for each stage of creative tasks (Kagan et al. 2018, Ibanez et al. 2018). A backward planning approach can address this, by helping people anticipate situational factors that might lead to delays (Wiese et al. 2016). Backward planning parallels the top-down constraints that companies like Lyft and Canvas implement for software updates (Allon et al. 2022). To this existing literature, we contribute an exploration of the role of generative AI within this planning process. Generative AI promises to change the way people do work, but it remains unknown what this means for teachers', and more generally, workers', workflow planning.

2.2. Productivity-Enhancing Technologies

With our study, we also build on research about productivity-enhancing technologies more generally. It is natural to assume that the emergence of new technologies, which are created to enhance worker productivity, would do so. However, counter to this expectation, productivity growth has slowed in the last decade (Brynjolfsson et al. 2019, Greenberg et al. 2024). In other words, the relationship between new technologies and productivity is complicated. For example, new technologies might make people more productive over the long run, but at the cost of lower productivity in the short run as workers learn to integrate it into their workflow (Bhargava and Mishra 2014, Ramdas et al. 2018). New technologies are also unlikely to affect all workers equally. Information technology (IT) appears to be an equalizer in some cases, as it boosts the productivity of marginalized workers the most (Ding et al. 2010).

New technologies might also affect outcomes other than productivity. For example, it is possible that AI will allow workers to focus on the parts of the job they enjoy most (Paschkewitz and Patt 2020). This is particularly important within the education context, where teachers have flexibility about their workflows (Wiggings and McTighe 2005). To this stream of literature, we contribute a particular focus on generative AI as a potential productivity-enhancing technology, which differs from previous technologies in how flexible it is for users. Its flexibility may mean generative AI offers even more productivity benefits for teachers, but it might also create even steeper barriers to integration than other, more standardized, technologies.

2.3. AI in Education

Although ChatGPT and other technologies are relatively new, there are a growing number of studies about developing and implementing other forms of AI for educators (Ng et al. 2023). While AI technology has been promising, teachers have been slow to adopt it. They often have difficulty finding relevant AI, and remembering to use it if they have found it (Leake and Lewis 2016, Zagalsky et al. 2015). It remains to be seen exactly how educators will use newer generative AI tools. While there is emerging research in this area (Lo 2023), much of this work has focused on teaching strategies for student users. For students, generative AI can harm learning, unless it is carefully deployed (Bastani et al. 2024, Abbas et al. 2024). Our paper contributes to this area the perspective of teacher users, specifically for their backward planning work.

3. Methodology

3.1. Sample

In Fall 2023, we recruited 24 US public school teachers from the Midwestern US. As shown in Table 1, the sample includes teachers from different grade levels (elementary, middle, high) and subject areas (math, science, ELA, social studies, foreign language, elementary education). There is also

Table 1 Teacher Participants

ID	State	Grade Level	Subject Area	Experience Teaching (yrs)	T_1	T_2	T_3	T_4
T1*	Ohio	Middle School	General	12	Yes	No	No	No
T2	Michigan	High School	ELA	22	Yes	Yes	Yes	Yes
T3*	Pennsylvania	High School	Math	7	Yes	Yes	Yes	No
T4	Michigan	High School	ELA	21	Yes	Yes	Yes	Yes
T5*	Pennsylvania	High School	Foreign Language	12	Yes	Yes	Yes	No
T6*	Pennsylvania	Elementary School	General	25	Yes	Yes	Yes	No
T7	Michigan	Elementary School	General	6	Yes	Yes	No	No
T8	Michigan	Elementary School	General	12	Yes	Yes	Yes	No
T9	Michigan	Elementary School	General	3	Yes	No	Yes	No
T10	Michigan	High School	Science	25	Yes	Yes	Yes	No
T11*	Pennsylvania	Middle School	ELA	17	Yes	Yes	Yes	Yes
T12	Michigan	High School	Math	22	Yes	Yes	Yes	No
T13	Michigan	Middle School	Science	2	Yes	No	No	No
T14	Michigan	High School	Foreign Language	11	Yes	Yes	No	No
T15	Michigan	High School	Foreign Language	8	Yes	Yes	Yes	No
T16	Michigan	High School	Foreign Language	14	Yes	No	No	No
T17	Michigan	High School	Science	8	Yes	No	No	No
T18	Michigan	High School	Math	12	Yes	No	No	No
T19	Michigan	Middle School	Social Studies	27	Yes	Yes	Yes	No
T20	Michigan	High School	ELA	20	Yes	Yes	Yes	No
T21	Michigan	High School	Social Studies	9	Yes	Yes	Yes	Yes
T22	Michigan	High School	Math	2	Yes	Yes	Yes	Yes
T23	Michigan	High School	Math	27	Yes	No	Yes	No
T24	Michigan	High School	Science	3	Yes	Yes	Yes	No
Total Number of Teachers Per Round of Data Collection:					**24**	**17**	**17**	**5**

variation in their years of experience (from 2 to 27). Our sample is theoretical to capture variation among teachers, and not random (Glaser and Strauss 1967, Eisenhardt 1989). We identified this sample through, first, reaching out to five participants (indicated by * in Table 1) through pre-existing relationships with the authors of this paper (personally known public school teachers). We then recruited the remaining participants (19/24) through emails sent to 83 K12 public school teachers in southeast Michigan based on subject area and grade level sampling goals. All teachers who responded to our recruitment email were included in our study. Consistent with the logic of case study methodologies, our theoretical sample is designed to maximize the breadth and richness of information gathered across "polar types" meaning teachers of different grades, subjects, and experience levels (Eisenhardt 1989). Our sample is not representative and cannot be used to infer general behavior or patterns of the wider population. Evidence from our small sample can be valuable to the operations management field for idea generation, novel theorizing, and innovation in a way parallel to how case studies are valuable for the generation of knowledge in the medical field (Fisher 2007).

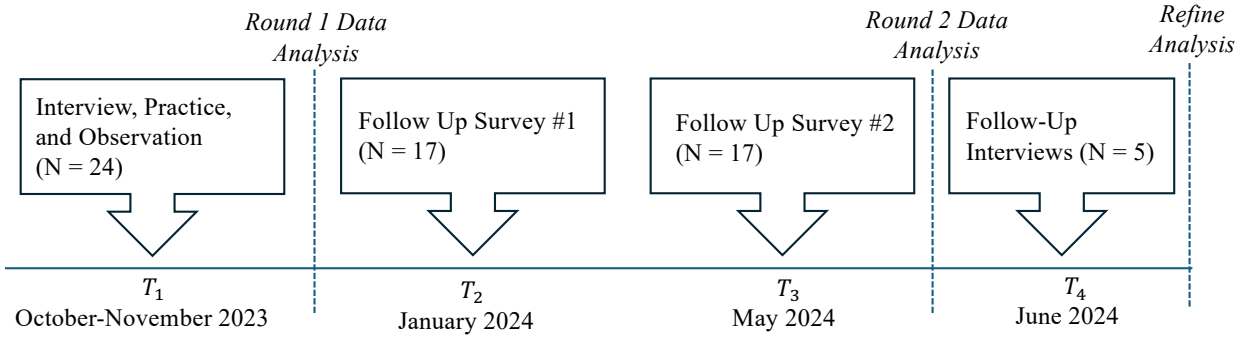
3.2. Choice of Generative AI Tool

We focus primarily on teacher use of ChatGPT. ChatGPT is one of many productivity-enhancing generative AI tools, but it is "by far the most widely recognised" one (Fletcher and Nielsen 2024). In Fall 2023, publicly available generative AI tools were still nascent technologies. Across the US, the majority of teachers had never used any form of AI for teaching, and of those teachers who had, most used virtual learning platforms like Google Classroom and Kahn Academy, which belong to an older wave of AI. Of the new wave of generative AI, general-purpose chatbots such as ChatGPT were the most common (Diliberti et al. 2024). None of the teachers in our sample worked at a school or in a district that formally adopted a generative AI tool during the 2023–2024 school year. Therefore, the teachers in our sample could only choose a publicly available tool, such as ChatGPT, for their personal use. The teachers in our sample reflected broader trends, in that ChatGPT was the most familiar form of generative AI to them. All of the teachers in our sample had heard about ChatGPT, even if they had never used it. Teacher 6 shared "I've never used it" but that "I hear about it everywhere." For these reasons, we opted to conduct our initial data collection during Fall 2023 on teachers using ChatGPT (specifically a paid ChatGPT Plus account). At the same time, we used our open-ended interview questions to ask teachers about their generative AI use more generally. Further, as education-tailored tools like SchoolAI and MagicSchool became more widespread over the school year, we adapted the language in our surveys to ask teachers the other generative AI tools they use, in addition to or instead of ChatGPT.

3.3. Data Collection

We collect multiple types of data at multiple points in time. There are four points of data collection—T_1, T_2, T_3, T_4—as shown in Figure 1. Below, we discuss each data collection effort in detail. We contacted all teachers in T_1, T_2, and T_3. Which teachers opted to participate at that time is indicated in Table 1.

Figure 1 2023–2024 School Year: Data Collection and Analysis Timeline



3.3.1. *Initial Interview and Exposure (T_1)*

We collected multiple types of data from all 24 teachers during a 1-hour one-on-one Zoom session conducted in September, October, or November 2023 (T_1). With a shared semi-structured protocol (Appendix A) one member of the research team did 13 of the interviews and another member did 11 interviews, which is a best practice in qualitative data collection that among other things can enhance validity when common patterns are found from different interviewers (Patton 2014). All teachers earned $50 for participating. With each teacher's consent, the audio and video of each Zoom call was recorded and the audio transcribed using an automated online transcription service (Sonix). The hour-long session involved four distinct parts.

Background Interview (15 minutes): Each session began with a background interview where teachers were asked about their workflows for material creation (colloquially described as "prep"). We asked about how they spend their days or weeks preparing for teaching, and the types of tasks involved. We also asked about whether they do their preparation alone, or with other teachers. We did ask about their experience with ChatGPT or other generative AI tools at this time, and all had very limited to no experience using the tool for their work. Only 2 teachers (Teachers 13 and 22) mentioned receiving any training on generative AI before our initial session with them. Thus, all teachers in our sample should be understood as new generative AI users at T_1.

Structured Practice with ChatGPT (10 minutes): The second part began when we asked teachers to log into our ChatGPT Plus (GPT-4)[1] account on their own computers and share their screen. After this, each teacher engaged in a structured practice session with ChatGPT where they entered the exact 12 prompts (Table 2) one-by-one in our prescribed order (we sent each prompt by chat individually, after which the teacher copied it into their ChatGPT window). The practice prompts were created in-line with contemporary studies of ChatGPT (Chen et al. 2023, Noy and Zhang 2023). The rationale of this standardized exposure was to make sure all teachers in our sample have a shared foundational understanding of what ChatGPT can and cannot do.

[1] Our data collection effort overlapped with OpenAI's first DevDay, which took place November 6, 2023. At that time, ChatGPT Plus was upgraded and 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 also output textual data in tabular format (versus only paragraphs and bullets) and also images. 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' altered capabilities. The dashed line in Table 1 indicates the separation between the before-versus-after interview and observation sessions. In response to the update, we slightly modified three of our practice prompts to give subsequent teachers (T18 through T24) practice with its newest features. Table 2 marks the updated prompts with '. For example, we updated prompt 11 from "Design a simple workout plan for beginners" to "Design a simple workout plan for beginners and present it in table form" to evidence the new feature that ChatGPT Plus could output text organized in a table.

Table 2 Required ChatGPT Prompts During the Structured Practice

No.	Prompt
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.
9'.	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.
11'.	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.
12'.	Give a diagram of the proper form for one of these exercises.

Observation of Teachers' ChatGPT Use (15 minutes): The third part of the Zoom session was a 15-minute observation of each teacher exploring and trying to use ChatGPT to support their teaching preparation work. We prompted each teacher to "Pick one of the things you mentioned earlier 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." We then set a timer for the teacher, and remained silent as they worked (some teachers narrated as they worked and in that case we engaged minimally). This observational design is an established approach in education research to examine how teachers create and modify materials on their own, without ChatGPT (Silver 2022).

Debrief Interview (10 minutes): After the observation period was completed, we debriefed the teachers' experience with ChatGPT. We asked specific questions about the intention behind the prompts inputted and their evaluation of the ChatGPT output during the observation. When time allowed, we asked teachers to share their thoughts about ChatGPT and how it might affect their work more generally.

3.3.2. *Generative AI Use Survey (T_2)*

The second wave of data collection was done in January 2024 (T_2) through a survey sent to the same 24 teachers that participated at T_1. The survey included both closed (Likert-style) questions about their *frequency* and *mode of* ChatGPT use in the months since T_1. We also asked about other generative AI tools the teachers were using at that time. The full survey is provided in Appendix B.

We asked specifically about how teachers were now using ChatGPT or another generative AI tool to *make*, *find*, *jumpstart*, and/or *iterate* on their planning. We derived this four-type categorization from an analysis of the evidence from our unstructured observation period at T_1 (See 3.4.1 for details of this analysis). This means that we conducted one round of analysis of our data before all data collection was complete. The iteration between data collection and data analysis is a common and in fact recommended best practice for case study research because it allows iterative refinement of emerging ideas (Yin 2016). The objective of the T_2 survey was to assess whether and how the *potential* uses of generative AI identified during our initial T_1 exposure and observation period materialized into real use *in practice*. Statistical analysis and extrapolation to a wider population was not an objective. For completion of the survey, teachers were compensated $10. 17 of the 24 (71%) of the original teachers responded to the January survey (See Table 1).

3.3.3. *Generative AI Use Survey (T_3)*

The third wave of data collection was done in May 2024 (T_3) through another survey sent to the same 24 teachers that participated at T_1. The survey contained the same questions as the January 2024 survey, but also included additional Likert-style questions about how generative AI is impacting their own learning, stress, number of tasks completed, hours working per week, and work quality. We asked separately about the impact of generative AI on their teaching, preparation, grading, and emailing. The additional May 2024 survey questions are provided in Appendix C. These additional questions aimed to uncover some qualitative insights about whether these teachers feel productivity gains, or losses, from generative AI so far. As in T_2, we were also generally interested in whether and how the *potential* uses of generative AI identified during our initial T_1 exposure and observation period were yet materializing into real use *in practice*. For completion of the slightly-longer survey, teachers were compensated $15. 17 of the 24 (71%) of the original teachers responded to the May survey (See Table 1).

3.3.4. *End-of-Year Interviews (T_4)*

We conducted five teacher interviews in June 2024 (T_4). These five teachers indicated in their May 2024 survey (T_3) that they have more they would like to share with us about generative AI. For example, at one teachers' school (Teacher 4), there was recently a training done on generative AI and a lot of discussion about student use of generative AI at the end of the school year. Another teacher (Teacher 21) wanted to talk to us more about their conscious decision to *not* use any generative AI. Among the five teachers interviewed at T_4, one reported productivity improvements in the May 2024 survey, two were strong non-users, and two were in the middle. We customized our interview questions based on these differences, as shown in Appendix D. These final interviews provided us an opportunity to refine our understanding of the different types of teacher-users and

why within our group teachers there came to be very different in-practice use of generative AI despite consistent initial exposure at T_1.

3.4. Data Analysis

We conducted two rounds of data analysis, as shown in Figure 1. We iterated between data collection and analysis, a best practice for case study research because it allows iterative refinement of emerging ideas (Yin 2016). Our emerging understanding of how teachers are beginning to integrate generative AI into their workflows, and specifically the differences among teachers, came from revisiting and reanalyzing earlier data in light of new data over the course of the 2023–2024 school year.

3.4.1. *Round 1 Data Analysis (Between T_1 and T_2)*

Our first round of data analysis focused narrowly on coding the prompts teachers wrote and submitted during the unstructured observation period in T_1. The objective was to define a use typology grounded in teachers' own prompting choices. We recorded all prompts written and sent by each teacher during the 15-minute unstructured observation period in a large spreadsheet. We did this by watching the recorded videos back, and copying the prompts verbatim into the spreadsheet. The 24 teachers inputted 201 prompts in total (8.4 per teacher, on average). We then iteratively derived a coding scheme of different generative AI use cases, first through open coding prompts and then refining descriptions of similar prompts to match (Corbin and Strauss 2008). 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 prompt coding scheme: (1) *make for me* (55% of prompts), (2) *find for me* (15% of prompts), (3) *jumpstart for me* (10.5% of prompts), and (4) *iterate with me* (15.5% of prompts). We used a fifth category, *show me what you can do*, for non-teaching-relevant requests testing ChatGPT's capabilities (4% of prompts). Table 3 gives descriptions of each code. Two members of the research team independently coded all 201 prompts. Even prompt was assigned only one code. The inter-rater reliability for the coding was 91%. Through discussion among the coders, different coding categorizations were resolved and in the end, all authors were in agreement. More details on this coding can be found in Keppler et al. (2024).

This coding scheme was used to create the survey questions used at T_2 and T_3. That is, in addition to overall generative AI use, we asked teachers specifically how often they used generative AI to *find for me*, *make for me*, *jumpstart for me*, and *iterate with me in practice* (See Appendix B for exact questions). The prompts inputted during the observation period represent potential use, whereas the survey questions examine whether teachers are adapting their workflows to integrate generative AI.

Table 3 Prompt Coding Scheme

Code	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"

3.4.2. *Round 2 Data Analysis (After T_3)*

After T_3, we conducted our primary data analysis effort. In this round of analysis, we analyze data at the teacher level and treat each teacher as one case of exploring whether and how to use generative AI. We follow the standard analysis steps of multiple case study research: (i) within-case analysis, (ii) across-case analysis, and (iii) theory-case analysis (Eisenhardt 1989).

Within-case analysis: The objective of within-case analysis is to develop an understanding of each of the 24 teachers' generative AI trajectories. We constructed a longitudinal database that compiled the data across multiple time points (T_1, T_2, and T_3). This allowed us to connect, for example, each teacher's prompts inputted during the observation at T_1 to their ongoing use at T_2 and T_3. In particular, the outcomes at T_3 related to their productivity, stress, and work quality we connected to their generative AI use originally and over time. The database also included each teacher's qualitative descriptions of their work and generative AI use. We also add columns for each teacher's attributes such as years of experience, grade level, subject area, and whether they worked independently or as part of a teaching group.

Across-case analysis: The objective of across-case analysis is to compare pairs or subsets of cases with one another with a goal to first identify and then refine emerging patterns. First, we separate the teachers by their productivity outcomes in the May 2024 survey. We code a teacher as reporting improved productivity if they indicate that as a result of generative AI they are (i) doing more tasks in less time, (ii) doing the same amount of tasks in less time, (iii) doing fewer tasks in less time, *or* (iv) doing more tasks in the same amount of time in one or more of the following areas of their work: teaching, preparation, grading, or emailing. Six teachers met this criteria. Note that it was possible for teachers to report productivity gains even if they do not use generative AI themselves, this was not the case for these six teachers. All were using generative AI at least one per month in May 2024. All six of these teachers also reported greater work *quality* as a result of using generative AI. The teachers who only reported increased *quality* and did not indicate improved

Table 4 Teacher Group Categorization at T_3 (May 2024)

ID	Productivity at T_3	Work Quality at T_3	User or Non-User at T_3	**Categorization at T_3**
T2	Decreased	No Change	Non-User	Non-User
T3	Increased	Increased	User	Improved Productivity
T4	No Change	Increased	User	No Change to Productivity
T5	Increased	Increased	User	Improved Productivity
T6	No Change	No Change	Non-User	Non-User
T8	Increased	Increased	User	Improved Productivity
T9	Increased	Increased	User	Improved Productivity
T10	Increased	Increased	User	Improved Productivity
T11	Increased	Increased	User	Improved Productivity
T12	No Change	No Change	User	No Change to Productivity
T19	No Change	Increased	User	No Change to Productivity
T20	No Change	No Change	Non-User	Non-User
T21	No Change	No Change	Non-User	Non-User
T22	Decreased	No Change	Non-User	Non-User
T23	No Change	Increased	User	No Change to Productivity
T24	No Change	Increased	User	No Change to Productivity

productivity were not included in this group of six "more productive with generative AI" teachers. Among the other 11 teachers (recall, we only have May 2024 survey responses from 17 of the 24 original teachers), 5 were non-users (they reported never using generative AI for their work) and 6 were users of generative AI but without reported productivity gains. This formed three distinct groups of teachers for our comparative, across-case analysis: improved-productivity teacher users, no-change teacher users, and non-users. The teacher grouping is reported in Table 4.

The next step was to compare the three groups of teachers to understand the different outcomes. It was curious to us that teachers who were very similar in fall 2024 when we began our study—novice users, little knowledge of generative AI, curious about the technology—and who all got a similar intervention from us in terms of the standard prompting and practice period could evolve over the same time period into distinct groups. We made conjectures about why, and tested them through a "logic of replication" (Eisenhardt 1989) with our teacher cases where we verified if a hypothesized relationship consistently held true for all teachers in our sample. For example, one conjecture was that teachers' experience in the classroom could explain this divergence. However, in the improved productivity group the experience level ranges from 3 to 25 years, in the no change group ranges from 3 to 29 years, and in the non-user group ranges from 2 to 22 years. While we cannot conclude if the hypothesized pattern might exist more generally, it does not seem to explain the diverge we observe. We also examine differences by grade level and subject area. Each user group included within it teachers of all grade levels (elementary, middle, and high school) and a variety of subject areas spanning both ELA/social studies and STEM disciplines. Thus, this also did not seem to explain the differences in productivity.

We next compared teachers by *how* they used of generative AI, starting with the observation period at T_1. Perhaps teachers who reports productivity improvements and those that do not

Table 5 Change in Teacher Generative AI Use from T_2 to T_3 (January to May 2024)

Frequency of Use Mode *(Scale 1 to 5, 1 = Never, 5 = Weekly)*	**Improved Productivity (N = 6)**	**No Change (N = 6)**
Make For Me	Increase (Avg. + 0.2)	Increase (Avg. + 1.0)
Find For Me	Increase (Avg. + 0.2)	Decrease (Avg. - 0.2)
Jumpstart For Me	Increase (Avg. + 1)	Increase (Avg. 0.4)
Iterate With Me	Increase (Avg. + 1.2)	Decrease (Avg. -0.2)
Overall*	Increase (Avg. + 0.6)	Increase (Avg. + 0.4)

**Overall change in frequency of use is a separate question asking teachers their overall use frequency, and is not derived as a total of the change in usage of each type (make for me, find for me, jumpstart for me, and iterate with me).*

approach generative AI differently. This was not the case at T_1. All groups—improved-productivity users, no-change users, and non-users—had inputted a similar variety of prompts at T_1 as defined by the coding in Round 1 (make, jumpstart, find, and iterate). The average number of different prompts entered by teachers in each of the groups was 2.5, 2.3, and 2.6, respectively. Thus, it seemed that early on in our controlled observation setting, all teachers were exploring how to generative AI. We looked therefore closely at their use in January and May to see how it evolved in practice (T_2 and T_3). In doing so, we uncovered that there was a notable divergence in use between improved-productivity and no-change users in January and May. As shown in Table 5, both improved-productivity and no-change users similarly increased the frequency with which they used generative AI *overall* (but 0.6 and 0.4, respectively). However, that increase derived from different modes of use. No-change teacher users largely increased their use of generative to *make*, whereas improved-productivity users largely increased their use of generative AI to *iterate* and also *jumpstart*. In fact, *iterate* evidenced the largest divergence between the two groups in terms of their evolving usage. This provided suggestive initial evidence of why some teachers were reporting productivity gains whereas others were not.

We then interrogated the data further. We went back to the *iterate with me* and *jumpstart for me* prompts teachers entered into ChatGPT during our observation period to identify what about these types of requests could stimulate productivity improvements. We also read through the open-ended examples from the surveys about how teachers were using generative AI to iterate, make, find, or jumpstart. Through this process, we came to understand that relying on generative AI to iterate or jumpstart was similar to asking generative AI for *input* into their teaching plans, whereas using generative AI to make or find (but mostly to make) largely involved creating *outputs*. For example, Teacher 1 prompted ChatGPT, "Can you explain how to add a negative number and a positive number" and then "What manipulatives besides a number line can I use?" As a result of the *input* from ChatGPT, she decided on an output she could make:"maybe I make this a word bank and print it off on a half sheet, and they can keep that in their math notebook right there doing their practice problems." The word bank was not her initial plan, but was devised

with input from generative AI. She explained that when it comes to deciding how to achieve a learning goal when something has not been working is "a lot of times where I get stuck." This approach is different *from a workflow perspective* than prompting ChatGPT for *outputs* after the teacher has already identified what needs to be created, for example prompting "Make a word bank for students for word problems so they know what sign (positive or negative) to use." With this insight, we moved to theory-case analysis.

Theory-case analysis: Theory-case analysis is the integration of emerging insights from case study data with the existing knowledge and theory. With the insight that the improved-productivity teacher users in our sample seemed to be using generative AI for input into workflow plans as well as for specific outputs, we turned to the literature on workflow planning and design (Ibanez et al. 2018). Comparing our own evidence to the literature, we identified an important feature teacher workflows: they stem backward from state-mandated learning objectives (See Appendix E for examples of state learning objectives; as you can see, there is no direction on *how* or *what* to teach, just what students need to be able to do in the end.) This pointed us to a related literature on backward planning (Wiggings and McTighe 2005, Wiese et al. 2016). We adopted a backward planning perspective to help situate our evidence within the wider scholarly knowledge, and generate knowledge about the role of generative AI in teaching work.

3.4.3. *Additional Refinement (T_4)*

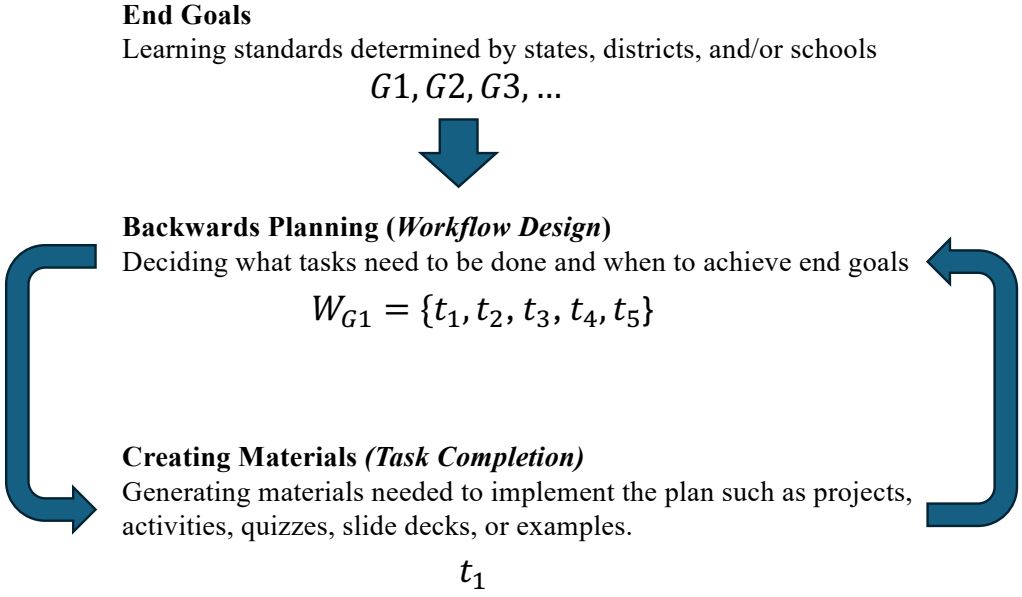
With the insight from our data collection that there was heterogeneity among users of generative AI, and that some teachers were non-users of generative AI, we leveraged a small set of additional follow-up interviews to refine our understanding of *why*. Recall, we conducted five follow-up interviews with the teachers. Those teachers were: improved productivity user (T11), no change user (T4), two always non-users (T2 and T21), and one who initially used generative AI in January but then became a non-user by May (T22). We integrated the evidence from these interviews into our teacher database. The interviews were particularly instructive about non-use of generative AI, and learning why a subset of teachers in our sample were choosing that option at this time. From interviews with the two teachers who were using generative AI, we identified differences mostly consistent with our emerging theory. The improved-productivity teacher explained how she was using generative AI for both input and outputs, yet the no-change teacher user spoke at length about the increased stress and frustration about student use of generative AI since January 2024 (T4 is a high school ELA teacher, a subject and grade level particularly affected by unauthorized student use). This helped us refine our emerging ideas to incorporate contextual factors that may amplify or hamper each teachers' motivation and ability to use generative AI for their work. Both the way teachers are (and are not) using generative AI *and* how that usage interacts with features of their work context that explains different outcomes. We go into the evident contextual factors in Section 4.3.

4. Findings

4.1. Using Generative AI *For* versus *Within* Backward Planning

Backward planning is a process shown in Figure 2. It begins with end goals $G1, G2, G3, ...$, such as the learning standards in Appendix E. With these goals in mind, a teacher works backward by first planning the set of tasks necessary to accomplish that goal. This set of tasks is a *workflow* i.e., $W_{G1} = \{t_1, t_2, t_3, t_4, t_5\}$. Once the workflow is planned, teachers then move to complete specific tasks t_i within the workflow. For teaching, tasks often (but not always) involve material creation, such as activities, lessons, discussions, quizzes, or slide decks. There is a return arrow between creating materials (task-level execution) and backward planning (workflow planning) because teachers regularly make revisions to their original plans. As tasks are completed, teachers gather information about progress toward goals. This can happen, for example, when teachers grade a quiz. At that point, teachers may adapt their original backward plan by adding more tasks (i.e., $W'_{G1} = \{t_1, t_2, t_3, t_4, t_5, t_6, t_7\}$ and/or modifying original tasks (i.e., $W'_{G1} = \{t_1, t_2, t_3, t'_4, t'_5\}$. The iterative nature of planning-then-doing in K12 teacher work means that teachers are frequently (re-)planning workflows. As Teacher 6 put it, *"I'm constantly on my feet, revising mentally or having to take a look at the upcoming common assessments."* Thus, the cycle shown in Figure 2 is short and tends to occur weekly, or even every few days, in K12 teaching work.

Figure 2 Backward Planning Workflow



In our initial interviews with teachers at T_1, we learned that 92% of the 24 teachers in our sample are involved in *backward planning*, 96% are involved in *material creation*, and 100% are involved in *modifying workflows*. The few teachers that are not involved with original planning and creation

are ones that get materials from other teachers, and then modify as needed for their students. 42% of the 24 teachers in our sample *provide* their plans and created materials to other teachers at their school. Overall, this evidence indicates that the teachers we study have a high degree of ownership and involvement in their backward planning workflow processes.

Common Use: For Creating *Outputs*. All teachers in our study that report using generative AI to create outputs necessary for completing their work tasks. When teachers use generative AI in this way, they already have the workflow in mind (i.e., $W_{G1} = \{t_1, t_2, t_3, t_4, t_5\}$) and ask for generative AI support on a particular task (i.e., t_1). Consider the prompting shown in Table 6 from our initial observation period where teachers asked generative AI for task-level support. Teacher 3, for example, requests generative AI to make an in-out table he can use in an in-class activity. Teacher 5 requests multiple choice questions. In the observation period, 80% (19/24) teachers requested generative AI support specifically for making something.

Table 6 Examples of Generative AI Prompts For Outputs

Teacher ID	**Output Prompting Examples from T_1 Observation Period**
3 (HS Math)	Can you give me a puzzle where I have to find the next thing in a visual pattern. Give me an in-out table where the input is not a number but the output is a number. Make it a little more complicated. Make the input not words. Make it less complicated.
5 (HS Spanish)	Make 10 multiple choice questions for chapter 7 for the story *Mi Proprio Auto*. Make 10 multiple choice questions using the subjunctive for the story *Mi Proprio Auto* with an answer key.
6 (Elementary)	Create a 5 question quiz for 3rd grade on the topic of forces and motion.
12 (HS Math)	Write a calculus question that would make it clear that a student understands how to find an absolute maximum
24 (HS Science)	Make a lab for 9th grade students using the following criteria: [attached a long list of criteria copied from course notes about a lab launching marshmallows] Give more guidance on the data collection and organizing section in day 3. Make slides for the teacher to present this all to the class. Add diagrams to the presentation. Instead of diagrams, add drawings of students doing the lab to the slides. Give a diagram of a student participating in this lab. Give a diagram of an adult participating in this lab. What about this is not aligned with the content policy for images? Give a diagram of a person demonstrating this lab.

Using generative AI to help create outputs was also evident in the follow-up surveys. *Make for me* support was the most frequently reported way teachers were using generative AI in practice in both the January and May 2024 surveys, on average. One teacher we interviewed in June (T11) explained why generative AI's ability to support teachers for particular tasks within workflows is helpful: *"The way we usually teach, we introduce the skill on day one. And then we do practice*

throughout the week. So with [the curriculum] only providing us one general worksheet, it's not very helpful. So I had to use [ChatGPT] quite frequently, like, 'oh, create a worksheet on this for fifth grade'. And then if I didn't like it, I just ask it in a different way." On the May 2024 survey, teacher (T15) shared, *"I have [ChatGPT] make vocab lists from readings or to generate questions to intro topics. I can use it to make images. I also ask it to make questions from chapters or readings."*

Using generative AI to create outputs within backward-planned workflows is similar in process to using generative AI for task completion in other workflows, though its implications may be different. Much of the existing research on generative AI has compared worker productivity and quality on particular tasks with and without generative AI (i.e., Dell'Acqua et al. 2023, Brynjolfsson et al. 2023, Chen et al. 2023) and found a significant improvement in workers using generative AI. While all the teachers in our study who use generative AI use it in a similar way to how participants use it in other studies, only some of the teachers we study are reporting productivity gains. No-change teacher users are using generative AI to *make* something at almost the same exact frequency as improved-productivity teacher users (3.5 versus 3.7 on a scale of 1 to 5, respectively). The similarity in the way teachers are using generative AI for this purpose, but different productivity outcomes, suggests that the productivity outcomes may be explained by a different use case.

Different Use: For *Input* into Work Plans. A subset of teachers in our sample sought generative AI for *input* into their work plans, in addition to asking for *outputs*. Table 7 gives examples from the observation period at T_1 where teachers came to generative AI with learning objectives in mind and not yet a set of tasks to complete. As the examples show, teachers sometimes shifted from planning to material creation (such as when Teacher 18 asks for help creating a problems set). The important difference compared to examples in Table 6 is the initiation of the conversation without a particular task in mind and instead with learning standards and objectives.

During the observation period, 54% (13/24) of teachers sough input from generative AI in at least one instance. Yet, in January 2024, only 9 teachers reported using it in this way. For example, in January 2024 a teacher (T19) shared she used ChatGPT to help her figure out *"how to teach dividing fractions."* Yet, another teacher shared that while she used it for input during the observation period, *"I don't use ChatGPT in this way"* in practice (T8). By May 2024, there is even more divergence. For example, only 8 teachers reported using generative AI to iterate, with four teachers reporting that they never use generative AI to iterate on their work plans and four teachers reporting they nearly always do. One of the frequent users of generative for this purpose, Teacher 3, explained how he *"typically starts a prompt with 'create a high level plan for…' so it will give me an outline of steps required for a project and things to consider at a high level."* Overall, while asking generative AI for outputs became common among all teacher users over the 2023–2024

Table 7 Examples of Generative AI Prompts For Input

Teacher ID	Input Prompting Examples from T_1 Observation Period
1 (MS Special Ed)	Can you explain how to add a negative number and a positive number? Create real world math problems within 100 that uses this concept. Add in multi-step word problems. Real word examples using numbers within 20. What manipulatives besides a number line can I use? My student doesn't understand how a positive and a negative number added can still be negative. Can you help me explain?
4 (HS ELA)	Describe standards by which a "great American novel" is determined. Describe novels contemporary to *Adventures of Huckleberry Finn* that reflect similar social and cultural issues. Describe novels contemporary to *The Great Gatsby* that reflect similar social and cultural issues. Describe essays, pamphlets, and books contemporary to *The Great Gatsby* that could inform a student's understanding of the Jazz Age and/or class stratification in the United States in the early 20th century. Discuss how Thorstein Veblen's *Theory of the American Leisure Class* illuminates social mores shown in *The Great Gatsby*.
18 (HS Math)	Construct a unit plan for teaching the binomial theorem at the Math HL level (use the IB syllabus). List in a table form including key concepts. Can you expand on week 3? What are some good examples? You say 'Example: Using the binomial theorem in problems involving physics, like calculating the potential energy in a spring system.' Can you give an example of such a problem Can you come up with an application where n is far greater than 2? Can you write a problem set with 4 questions related to the binomial theorem. One should be a word problem. Can you write a question that involves both trig identities and the binomial theorem used in conjunction? Can you rewrite that question in a 'show that' form? Produce a solution key to your problem

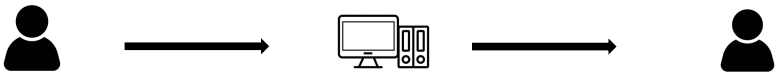
school year, asking generative AI for input became frequent for some teachers but never happened for others.

The teachers who report using generative AI for both input and outputs are the same group of teachers who report productivity gains. In May 2024, improved-productivity teachers report using generative AI to iterate on teaching plans and ideas nearly twice as frequently as no-change teachers (average of 3.2 versus 1.7, on a 1 to 5 scale), even as they report more similar overall use (average of 3.8 versus 3.3, respectively). Taken together with the results that using generative AI for material creation alone does not seem to produce productivity gains, our findings stimulate two hypotheses. The first is that generative AI's input—but not its output—generates productivity improvements for teachers. The second is perceptual: that teachers do not *feel* more productive when using generative AI to create outputs. Either way, there are implications of our findings for understanding how to integrate generative AI into backward, goal-oriented workflows.

4.2. How Teachers Seek Generative AI's Input and Outputs

Figure 3 shows the steps taken by teachers in our study when trying to use generative AI in the two different ways: for input versus for outputs.

Figure 3 The Way Teachers Seek Generative AI Support



Seeking Input:	Through reflection, the teacher identifies a potential that there may be a better way---that they do not know about---to achieve the target learning goals.	Teachers prompt generative AI to iterate on and explore new ideas about how to teach and get students to understand target learning goals.	Teachers either: 1. Feel more confident that what they are doing is an effective way forward. 2. Have ideas for a new workflow plan that they can be more confident about, which they then create.
Seeking Outputs:	The teacher identifies materials that they need but do not have even a draft of, and so they would like help creating it.	Teachers prompt generative AI to make the materials they need.	Teachers either: 1. Have a draft of the needed materials, which they then modify and improve upon. 2. Have a useable, final version of the materials they need.

Teachers seek input from generative AI when they identify a potential that there is a more effective way to reach the target learning goals. Teacher 1 explained, *"I know a lot of my students really struggle with word problems, and a lot of times like those are hard to create on your own to make sure they align with the actual objective, or you don't use the same example over and over to be like with fractions…it's nice to have other ideas."* Teachers seeking input from generative AI on their plans use it to come up with alternative ideas, not to make an entire plan for their course without incorporating their own expertise and input. As we shared earlier, Teacher 1, for example, came up with the idea to make a word bank for her students. A consequence of generative AI's input is that teachers can feel more confident. Teacher 13 explained, *"I think it saves me some time. But for the most part, I think it saves me a stress instead. It's not like this is like I'm all done. 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."*

Teachers using generative AI to create outputs are asking generative AI to make materials for them. Yet, in our data, teachers explain that they almost always still do a lot of material creation work themselves modifying or improving upon what generative AI creates. For example, Teacher 6 asked ChatGPT to make a forces-in-motion quiz during our observation period, and she explained, *"I would absolutely need to insert illustrations for each of the questions, since a lot of the students*

at the age that I have are very visual with regards to their learning." A high school calculus teacher also mentioned the gap between what he can use and what ChatGPT creates: *"that's the issue with math - I'm going to spend most of my time trying to rewrite this notation...it's a bummer."* The follow-up work when using ChatGPT for content creation is necessary in part because the tool is not specialized for teachers. New generative AI specifically for educators (i.e., MagicSchool) may be able to get closer to the final product teachers desire. Even with the limitations of ChatGPT, in a few examples in our study, teachers reported getting usable materials immediately, *"I think this is ready to go. This two pages is a perfect assignment for my kids. It's probably what I'm going to give them on tomorrow."* (T17).

4.3. Why Do Teachers Come to Use Generative AI Differently?

Why do teachers come to use generative AI differently over the course of the 2023–2024 year, despite similar baseline experience and initial exposure? There are many reasons. We describe three major categories of influence on teachers' generative AI use evident in our data.

The Features of Generative AI Technology: Generative AI, and ChatGPT in particular, is an evolving technology (García-Peñalvo and Vázquez-Ingelmo 2023). An important observation is that all six of the improved-productivity users in May 2024 were initially exposed to ChatGPT *before* its upgrade in November 2023. The upgrade added image generation and online searches to ChatGPT Plus, among other things, and therefore enhanced generative AI's ability to create outputs compared to before the upgrade. It is possible, therefore, that the teachers exposed before the upgrade discerned an initial potential of generative AI for both providing input and creating outputs, and this initial perception was carried throughout the 2023–2024 school year. Meanwhile, after the upgrade directed teachers' attention primarily to the output creation capabilities. This of course is not possible to test with our data, but it is important to note that no teachers in our study who were exposed to ChatGPT after the upgrade reported productivity gains by the end of the year.

Parallel to changes to the features of the technology itself were growing public rhetoric around generative AI's output capabilities. For example, on New Years Day 2024, Sal Khan (Founder of Khan Academy and education generative AI tool Khanmigo) shared *"Looking ahead to 2024, I see generative AI tools cutting 90% of teachers' admin tasks."* This quote suggests an expectation that teachers will use generative AI for outputs, rather than for seeking input. This prominent rhetoric may have contributed to some teachers' decisions to focus on using generative AI for output creation. In fact, given the societal and technological emphasis on output creation with generative AI, it is impressive that one-fourth of teachers in our sample identified an alternative, and complementary, potential of generative AI to provide input.

Competing Demands on Teachers' Time and Energy: Another factor that seems to drive different teacher use of generative AI is the extent to which teachers faced competing demands on their time and energy. Integrating generative AI requires adapting routines and the way work gets done (Leonardi 2011). Even if this is productive in the long-run, there is a short-term cost. For some teachers, the cost of adapting their routines and behaviors to integrate generative AI was too high, at least for the current school year. For example, two non-users (Teachers 20 and 22) were both fearful of district layoffs that were announced in January. This was part of the reason Teacher 22 stopped using generative AI: *"I was one of the potential layoffs. And I was recalled. And just all this emotional stuff. Learning new stuff was not on my docket."* Regarding ChatGPT she said, *"I will figure that out one day, you know. Then next month, or in the summer when there's more time."*

Another issue affecting teachers' time and energy was increasing non-sanctioned *student* use of generative AI over the 2023–2024 school year. In June 2024, Teacher 4 (high school English) explained, *"We just were alarmed at the incredible pace with which it seemed to take over content within this last school year. There was use of it last year [2022-2023]. But it was a far smaller number of students. And the students who did use it last year seemed to use it on more or less one off occasions in order to basically, you know, shortcut work. Whereas this year there are kids who were basically using it as a substitute for written language. Or their own written language, I guess I should say."* More than that, he shared, "*we were just shocked at a how little interest the [school and district] leadership seemed to have and in really having that conversation [about student use].*"

A number of teachers in our sample—Teachers 2, 22, and 24— reported spending *more* time on grading as a result of growing generative AI use in their classes. Overall, based on our survey results, teachers reported an *increase* in stress related to student use of generative AI over the 2023–2024 school year. Teacher 4 was one of the teachers who reported a higher stress level. He was using generative AI, but he did not report productivity gains, in part because he was alarmed by student use. He concluded with this statement, "*This feels to me bigger than cell phones in schools. People freak out about cell phones in school. There's all kinds of research that shows you know how detrimental they can be. And so there's slow progress, progress to like trying to limit students access to social media in school. [Generative AI] feels much bigger. This feels much potentially more disruptive to like the academic model that we've used, especially."* Thus, no-change teachers are not necessarily seeing no change with generative AI simply because they are not using it for input on teaching plans, but rather because they are significantly influenced by other environmental events related and unrelated to generative AI.

Ethical Aversions: Non-users, and some users, expressed varying levels and types of aversion to generative AI. In May 2024, Teacher 2 (a non-user) summarized her stance about teacher use: *"it*

feels like cheating." She went on: *"I mean, there's all these lawsuits, right? New York Times and all these other publications that are suing because the [AI companies] are profiting off of uncompensated writing by professional authors."* Another non-user, Teacher 21, similarly emphasized concerns with the way generative AI is designed and trained. In the follow-up interview in May 2024 he explained that *"I think ChatGPT kind of takes people's work... broadly, kind of scraping the internet, to be trained on. And then it seems like large companies like OpenAI, broadly profit off said work."* As a result, *"I think [ChatGPT, and similar genAI] is predicated on a pretty predatory and destructive model just off the base, and that's completely separate from my concerns as a teacher. Conceptually, it's bad."* Teacher 21 questioned what this model means for the future of generative AI. *" ...the current internet model is really based off of advertising, right? But it seems like a lot of generative AI ... you aren't even really going to these websites anymore where the original material is sourced from. So under our current sort of financial model, how are those websites to make money? And if they don't make money, how are they to survive? And if they don't survive, what will ChatGPT feed off of in the future to continue? It seems like a snake that eats itself with very little regulation. And it's already seemingly running low on materials to consume."* Even teachers who were using generative AI, such as Teacher 12, expressed these types of concerns: *"I use it to list sometimes (5 main reasons for revolutionary war) but I question sources."* In sum, ethical aversion also seemed to affect whether and how much teachers were using generative AI.

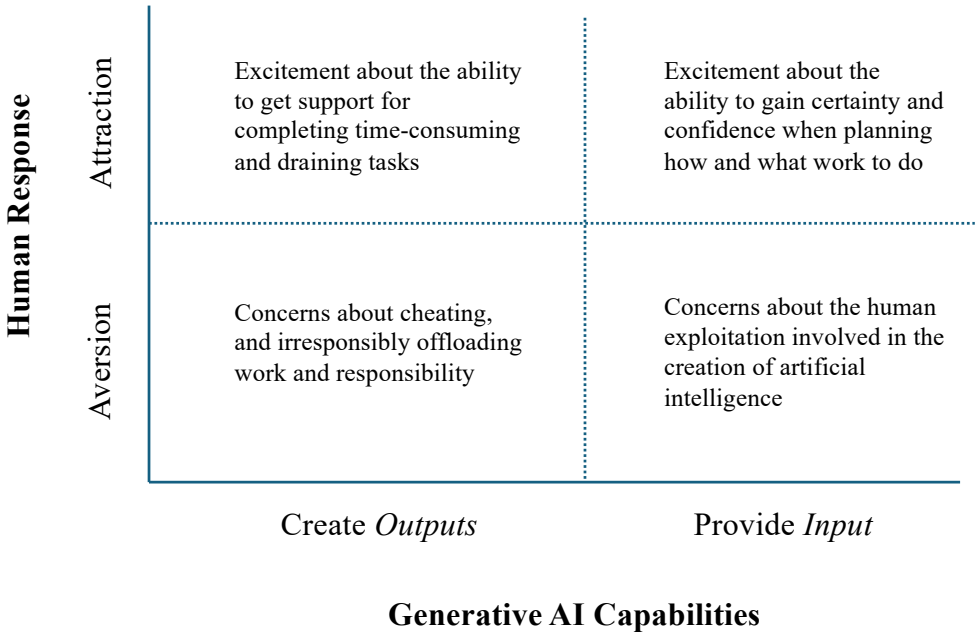
5. An Emerging Theory of Generative AI Use in Backward Planning

Figure 4 summarizes our emerging theory of generative AI use within backward planning, where it is necessary to both plan work and complete work. The horizontal axis is the two *Generative AI Capabilities*. One is the ability to create *outputs* and the other is the ability to provide *input*. The vertical axis is the *Human Response* to these different features of generative AI. One response is attraction. In our data, all teachers using generative AI were attracted to its generative capabilities to create outputs. Yet, only some teachers using generative AI were attracted to its ability to provide input. The teachers who were attracted to generative AI's ability to provide input were the ones who used it to ideate on backward teaching plans and gain certainty and confidence when planning how and what work to do.

Aversion to generative AI was another human response evident in out data. The same two features of generative AI that stimulated attraction in some cases stimulated aversion in other cases. Generative AI's ability to create outputs stimulated concerns about cheating and irresponsible offloading of work, whereas its ability to provide input stimulated concerns about human exploitation in the creation of generative artificial intelligence.

While there are contextual factors, like those discussed in section 4.3, that influence whether and how teachers use generative AI, it is harder to explain exactly why teachers *respond* differently

Figure 4 2-by-2 of Human Response to Generative AI for Backwards Planning



in terms of attraction versus aversion. Part of the reason why is that most teachers in our study cannot be easily placed within one cell of Figure 4's two-by-two. Improved-productivity teacher users, for example, are in the top two cells (top-left and top-right), as indicated by their dual usage of generative AI to create outputs and to seek input. Teacher 11 (T_4 interview), for example, explained her appreciation for the input generative AI— *"that's one of the biggest things—just to get the thoughts going. I'm thinking, okay, what are some things that the kids are going to be interested in? Because I'm talking about something that's going to bore them to death. They're not going to they're not going to be involved, you know"*—as well as the generation capabilities *"throughout the year, I've used it to make study guides, tests, quizzes.*" Meanwhile, non-users like Teachers 2 and 21 are in bottom two cells (bottom-left and bottom-right). They are averse to all of what generative AI can give (outputs or input) without attraction. The users of generative AI who are not reporting productivity gains, interestingly, seem to be mostly in the two cells on the left hand side of the two-by-two: experiencing a combination of attraction and aversion to the technology's ability to create outputs. Teacher 4 explained this struggle (June 2024 interview): *"The good things seem pretty cool, but the bad things seem really bad, right? It doesn't seem that the benefits are justifying some of the harms."*

6. Conclusion

With a longitudinal case study of 24 US public school teachers, this study provides rich insight into how generative AI is beginning to impact teacher work. Teachers work backward, meaning they

start with learning goals and then plan the set of tasks they (and their students) ought to do to achieve the goals. In this type of workflow, we find two ways teachers are using generative AI: (i) to create *outputs* necessary for specific tasks and (ii) for its *input* on their teaching plans. Teachers using generative AI in both ways are the ones that are reporting productivity gains from using generative AI. From this, we derive an emerging theory of different teacher responses (attraction versus aversion) to two capabilities of generative AI: the ability to *generate* content (create output) and the ability to learn something from its *intelligence* (provide input). There is still much more to understand about how generative AI will impact work in general, and teacher work in particular. Our study stimulates new directions to explore related to generative AI's capabilities for providing input versus creating outputs, and more generally how generative AI use and productivity potential varies based on workflow structure.

References

Abbas, M., Jam, F. A., and Khan, T. I. (2024). Is it harmful or helpful? examining the causes and consequences of generative ai usage among university students. *International Journal of Educational Technology in Higher Education*, 21(1):10.

Aksin, Z., Armony, M., and Mehrotra, V. (2007). The modern call center: A multi-disciplinary perspective on operations management research. *Production and operations management*, 16(6):665–688.

Allon, G., Askalidis, G., Berry, R., Immorlica, N., Moon, K., and Singh, A. (2022). When to be agile: Ratings and version updates in mobile apps. *Management Science*, 68(6):4261–4278.

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., and Mariman, R. (2024). Generative ai can harm learning. *Available at SSRN 4895486*.

Bhargava, H. K. and Mishra, A. N. (2014). Electronic medical records and physician productivity: Evidence from panel data analysis. *Management Science*, 60(10):2543–2562.

Brynjolfsson, E., Li, D., and Raymond, L. R. (2023). Generative ai at work. Technical report, National Bureau of Economic Research.

Brynjolfsson, E., Rock, D., and Syverson, C. (2019). Artificial intelligence and the modern productivity paradox. *The economics of artificial intelligence: An agenda*, 23:23–57.

Chen, L., Zaharia, M., and Zou, J. (2023). How is chatgpt's behavior changing over time?

Common Core State Standards (2010). *National Governors Association: Washington, DC*.

Corbin, J. and Strauss, A. (2008). Basics of qualitative research 3e (v. knight, s. connelly, l. habib, k. wiley, & g. treadwell eds.).

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., and Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of ai on knowledge worker productivity and quality. *Harvard Business School Technology & Operations Mgt. Unit Working Paper*, (24-013).

Diliberti, M., Schwartz, H. L., Doan, S., Shapiro, A. K., Rainey, L., and Lake, R. J. (2024). *Using Artificial Intelligence Tools in K-12 Classrooms*. RAND.

Ding, W. W., Levin, S. G., Stephan, P. E., and Winkler, A. E. (2010). The impact of information technology on academic scientists' productivity and collaboration patterns. *Management Science*, 56(9):1439–1461.

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of management review*, 14(4):532–550.

Fisher, M. (2007). Strengthening the empirical base of operations management. *Manufacturing & Service Operations Management*, 9(4):368–382.

Fletcher, R. and Nielsen, R. (2024). What does the public in six countries think of generative ai in news? *Reuters Institute for the Study of Journalism*.

Freeman, M., Savva, N., and Scholtes, S. (2017). Gatekeepers at work: An empirical analysis of a maternity unit. *Management Science*, 63(10):3147–3167.

García-Peñalvo, F. and Vázquez-Ingelmo, A. (2023). What do we mean by genai? a systematic mapping of the evolution, trends, and techniques involved in generative ai.

Glaser, B. and Strauss, A. (1967). *Discovery of grounded theory: Strategies for qualitative research*. Routledge: London, UK.

Greenberg, E., Padhi, A., and Smit, S. (2024). 2024 and beyond: Will it be economic stagnation or the advent of productivity-driven abundance? *McKinsey & Company*.

Hacker, P., Engel, A., and Mauer, M. (2023). Regulating chatgpt and other large generative ai models. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, pages 1112–1123.

Ibanez, M. R., Clark, J. R., Huckman, R. S., and Staats, B. R. (2018). Discretionary task ordering: Queue management in radiological services. *Management Science*, 64(9):4389–4407.

Jaffe, S., Shah, N., Butler, J., Farach, A., Cambon, A., Hecht, B., Schwarz, M., and Teevan, J. (2024). Generative AI in Real-World Workplaces: The Second Microsoft Report on AI and Productivity Research. Technical report, Microsoft.

Kagan, E., Leider, S., and Lovejoy, W. S. (2018). Ideation–execution transition in product development: An experimental analysis. *Management Science*, 64(5):2238–2262.

Keppler, S., Sinchaisri, P., and Snyder, C. (2024). Making chatgpt work for me. University of Michigan Ross Working Paper.

Krishnan, V. and Ulrich, K. T. (2001). Product development decisions: A review of the literature. *Management science*, 47(1):1–21.

Leake, M. and Lewis, C. (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, CSCW '16 Companion, page 321–324, New York, NY, USA. Association for Computing Machinery.

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

Lo, C. K. (2023). What is the impact of chatgpt on education? a rapid review of the literature. *Education Sciences*, 13(4):410.

Ng, D. T. K., Lee, M., Tan, R. J. Y., Hu, X., Downie, J. S., and Chu, S. K. W. (2023). A review of ai teaching and learning from 2000 to 2020. *Education and Information Technologies*, 28(7):8445–8501.

Noy, S. and Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654):187–192.

Paschkewitz, J. and Patt, D. (2020). Can ai make your job more interesting? *Issues in Science and Technology*, 37(1):74–78.

Patton, M. Q. (2014). *Qualitative research & evaluation methods: Integrating theory and practice*. Sage publications.

Ramdas, K., Saleh, K., Stern, S., and Liu, H. (2018). Variety and experience: Learning and forgetting in the use of surgical devices. *Management Science*, 64(6):2590–2608.

Silver, D. (2022). A theoretical framework for studying teachers' curriculum supplementation. *Review of Educational Research*, 92(3):455–489.

Wiese, J., Buehler, R., and Griffin, D. (2016). Backward planning: Effects of planning direction on predictions of task completion time. *Judgment and Decision Making*, 11(2):147–167.

Wiggings, G. P. and McTighe, J. (2005). *Understanding by Design (2nd Edition)*. Association for Supervision and Curriculum Development (ASCD): Alexandria, VA.

Yin, R. K. (2016). *Case Study Research and Applications: Design and Methods*. SAGE Publications: Los Angeles, CA, 6th edition.

Zagalsky, A., Feliciano, J., Storey, M.-A., Zhao, Y., and Wang, W. (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*, CSCW '15, page 1906–1917, New York, NY, USA. Association for Computing Machinery.

Appendix A: T_1 Semi-Structured 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?

Appendix B: T_2 (January 2024) 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.*

Appendix C: T_3 (May 2024) Additional Survey Questions

1. *(Scaled 1 to 5, where 1 = Strongly Disagree and 5 = Strongly Agree)* To what extent do you agree with the following statements: Using generative AI has helped me learn how to better do my....
   (a) prepping.
   (b) teaching.
   (c) grading.
   (d) emailing.
2. *(Scaled 1 to 5, where 1 = Strongly Disagree and 5 = Strongly Agree)* To what extent do you agree with the following statements: Generative AI has increased my stress about...
   (a) prepping.
   (b) teaching.
   (c) grading.
   (d) emailing.
3. For each aspect of your job, what is currently true about the effect of generative AI on the number of things/tasks you have to do each week? *(Do more, Do about the same/no change, or Do less)*
   (a) Prepping.
   (b) Teaching.
   (c) Grading.
   (d) Emailing.
4. For each aspect of your job, what is currently true about the effect of generative AI on the total number of hours you spend working each week? *(Do more, Do about the same/no change, or Do less)*
   (a) Prepping.
   (b) Teaching.
   (c) Grading.
   (d) Emailing.
5. For each aspect of student work, what is currently true about the effect of generative AI on the the quality of your students' work? *(Higher quality, About the same quality/no change, Lower quality, NA)*
   (a) Writing - formal
   (b) Writing - informal
   (c) Independent learning
   (d) Critical thinking
   (e) Problem solving
   (f) Classroom engagement
   (g) Scientific thinking
   (h) Quantitative/math skills

Appendix D: T_4 (June 2024) Follow-Up Survey Questions

Questions for Improved-productivity Teachers

1. You indicated that you are seeing some productivity boosts from using ChatGPT [verify they agree]. Can you about how you came to be able to use it in that way?
   (a) Did it happen right away?
   (b) Did it change over the course of the year?
   (c) What have you learned?
   (d) What advice would you give to district leaders knowing what you know now? 2-3 things.
2. How are you feeling about the next school year given this AI trend?
3. Anything else you want to share?

Questions for Minimal Users

1. You indicated that you are minimally using ChatGPT [verify they agree]. Can you talk about why you decreased/stopped/are minimally using it?
   (a) Is it that you think the technology is bad, or it's just hard to learn?
   (b) How did things change over the course of the year?
   (c) What have you learned?
   (d) What advice would you give to district leaders knowing what you know now? 2-3 things.
2. How are you feeling about the next school year given this AI trend?
3. Anything else you want to share?

Questions for Non-Users

1. You indicated that you are not using ChatGPT/generative AI [verify they agree]. Can you talk about why?
   (a) If/how has your perspective changed over the course of the year?
   (b) What have you learned?
   (c) What advice would you give to district leaders knowing what you know now? 2-3 things.
2. How are you feeling about the next school year given this AI trend?
3. Anything else you want to share?

Appendix E: Sample State Learning Standards

Table A1 Examples of 5th Grade Common Core State Standards in ELA and Mathematics

Standard Number	Standard
CCSS.ELA-LITERACY.L.5.1.A	Explain the function of conjunctions, prepositions, and interjections in general and their function in particular sentences.
CCSS.ELA-LITERACY.L.5.4.B	Use common, grade-appropriate Greek and Latin affixes and roots as clues to the meaning of a word (e.g., photograph, photosynthesis).
CCSS.ELA-LITERACY.L.5.3.B	Compare and contrast the varieties of English (e.g., dialects, registers) used in stories, dramas, or poems.
CCSS.MATH.CONTENT.5.NBT.A.1	Recognize that in a multi-digit number, a digit in one place represents 10 times as much as it represents in the place to its right and 1/10 of what it represents in the place to its left.
CCSS.MATH.CONTENT.5.MD.A.1	Convert among different-sized standard measurement units within a given measurement system (e.g., convert 5 cm to 0.05 m), and use these conversions in solving multi-step, real world problems.
CCSS.MATH.CONTENT.5.MD.C.5.C	Recognize volume as additive. Find volumes of solid figures composed of two non-overlapping right rectangular prisms by adding the volumes of the non-overlapping parts, applying this technique to solve real world problems.