The Teacher's Knapsack: Generative AI and Teacher Productivity

Samantha Keppler
Stephen M. Ross School of Business, University of Michigan, srmeyer@umich.edu

Wichinpong Park Sinchaisri
Walter A. Haas School of Business, University of California, Berkeley, parksinchaisri@berkeley.edu

Clare Snyder
Leonard N. Stern School of Business, New York University, clare.snyder@stern.nyu.edu

Much of the existing evidence on generative AI productivity studies settings in which tasks are assigned. We study K12 teachers, who choose which instructional tasks to prepare. We model teacher preparation as a time-minimizing knapsack in which the teacher chooses the least-time bundle of instructional tasks that meets a required learning value. We ground the model in a case study of 24 teachers over the 2023–2024 school year, including 360 minutes of observations, 29 interviews, and 34 surveys. Regarding the impact of generative AI, the evidence motivates a distinction between teachers' *task selection* and *task execution*. Teachers who use generative AI to execute familiar tasks in their legacy bundle of tasks can save time or increase the learning value of those tasks. Yet, moving away from legacy tasks can generate additional time savings when AI makes alternative tasks faster, more valuable for learning, or newly feasible. However, there is a cost to switching away from the legacy set of tasks, and for some teachers this may be too high, particularly when generative AI substantially improves the legacy tasks themselves. Environmental changes may also make selecting a new bundle of tasks necessary, even for teachers who do not personally use AI. Our case study evidence provides insight into these mechanisms: teachers commonly used AI for historically familiar tasks, whereas teachers reporting productivity gains increased their use of AI for jumpstarting and iterating on what tasks to do in the first place. The central implication is that AI productivity in discretionary work depends jointly on task acceleration, substitution, and the cost of evaluating alternatives.

Key words: generative AI; task selection; worker discretion; productivity; education operations; qualitative research

1. Introduction

Generative AI is expected to transform teacher productivity. With limited budgets, persistent teacher shortages, and a post-pandemic student body that remains behind on learning benchmarks, schools face growing instructional needs, and generative AI appears to offer a way to help (Klopfer et al. 2024). Yet productivity gains are not automatic. A tool that makes worksheets faster to produce may simply lead teachers to make more worksheets, even when worksheets are not the best way to support learning. A tool that makes new activities feasible may present a steep learning curve. These examples point to a common feature of teacher work that existing AI research overlooks: teachers have substantial task *discretion*. They choose both the format (hands-on activity,

worksheet, project, group work, individual assessment, presentation) and the content (level of difficulty, examples, problem context, open-ended versus multiple choice questions) of the materials they create for their students each week.

The productivity question is therefore not only whether AI helps teachers complete tasks faster, but whether it changes which tasks they choose to complete, for better or worse. This distinction is largely absent from existing evidence on generative AI and productivity. Prior studies typically examine settings where the task is assigned or arrives exogenously (Noy and Zhang 2023, Dell'Acqua et al. 2023, Brynjolfsson et al. 2023). Workers may decide whether or how to use AI, but the task itself is largely fixed. Discretionary work creates a different problem. Research in behavioral operations management shows that worker discretion over task selection and sequencing can harm productivity (Ibanez et al. 2018, Kagan et al. 2018, Kc et al. 2020), yet this literature predates generative AI and does not consider how AI changes the economics of task selection itself. In this paper, we ask: *how does generative AI affect K12 teacher productivity, given teachers have discretion over both what to do and how to do it?* We characterize the generative AI impact along two distinct channels: *task execution*, completing chosen tasks faster or better, and *task selection*, changing which tasks are worth doing in the first place.

We model K12 teachers' task discretion as a time-minimizing knapsack. The teacher's objective is to select the set of tasks that achieve student learning goals in the least amount of preparation time. This formulation fits settings in which the instructional requirement remains binding and preparation time is scarce. Holding the requirement fixed also prevents productivity gains from coming at the expense of instructional ambition. At the same time, we do not assume that teachers explicitly solve an optimization problem or care only about time. The model makes explicit what changes when AI affects task time, task learning value, and the teacher's ability to search over alternatives. We ground the model in a multiple case study of 24 K12 teachers. Our data include 29 in-depth interviews, 360 minutes of observed teacher use of ChatGPT, and 34 surveys collected over the 2023–2024 school year.

The knapsack model together with case study evidence yields three implications. First, generative AI can change the teacher's task selection problem even before the teacher uses it. Student AI use, verification burdens, grading effort, and assignment redesign can alter the time or learning value of legacy tasks, so preserving the same task bundle need not preserve the pre-AI instructional environment. Second, the productivity value of AI depends on which tasks improve. Faster legacy tasks make the current bundle cheaper to retain and reduce the additional gain from task selection, whereas making tasks outside the legacy bundle faster, increasing task learning value, or broadening the candidate set (tasks "unlocked" due to generative AI) can increase that gain. Third, task reselection due to generative AI is not automatic. Searching over candidate tasks, evaluating AI task

suggestions, adapting materials, and replacing familiar routines is burdensome and we model it as a one-time cost. The teacher re-selects the bundle only when the cumulative saving across expected uses of the materials created through the task exceeds that cost. AI can lower the cost of figuring out which tasks are the best to do or can expand the set of tasks even considered, but it can also crowd out search when execution gains are concentrated on legacy tasks. The same logic extends to when AI use is budget-constrained: teachers must decide not only which tasks to do, but also where AI use is worth the verification, policy, ethical, or attention costs it creates.

The case study evidence gives further insight into the model findings. Teachers who reported productivity gains were more likely to increase their use of generative AI for jumpstarting and iterating on teaching plans, uses more consistent with task selection. We also examine generative AI prompts written by teachers as they search for new tasks. The prompts' goal clarity and connectedness are consistent with more structured task selection, while output dispersion is useful only when aligned with the teacher's objective. What this shows is observable traces of whether AI use appears organized around executing known tasks or searching over candidate tasks. Finally, non-use and selective use of generative AI are consistent with the model's frictions: teachers may avoid AI when search costs are high, when student AI use disrupts legacy assignments, when verification burdens increase preparation time, or when ethical concerns limit the tasks for which AI assistance is acceptable.

The paper contributes to research on generative AI, productivity, and discretionary work by shifting the unit of analysis from the task to the task bundle, similar to other emerging research on AI and work re-design (Savva 2026). For operations management, this shift matters because productivity in discretionary work depends not only on how quickly a worker executes a given task, but also on whether the worker reallocates effort across tasks after a technology changes their relative time and value. For research on AI in education, the paper shows why teacher-facing AI tools should not be evaluated only as artifact generators, important given the research that generative AI might harm teaching practice (Sungu et al. 2026). A tool can be valuable because it helps teachers make a worksheet faster, but it can also be valuable because it helps teachers decide whether a worksheet belongs in the bundle at all. For school and district leaders, the implication is that AI training should not frame productivity only as doing the same work faster. It should help teachers identify when AI changes which work is worth doing, while recognizing that search, evaluation, verification, and ethical acceptance are real constraints on adoption.

2. The K12 Context: Teacher Work and AI Use

Decisions about task selection and execution arise in many kinds of work. They are especially consequential in K12 classrooms, where high workloads coexist with substantial discretion over what teachers prepare and how they prepare it.

2.1. The K12 Classroom

There are nearly 4 million K12 teachers working in the United States today (Schaeffer 2024). Their work varies by grade level, subject, state, and experience, yet it is shaped by common constraints. Federal law, state standards, and district policies define learning objectives and accountability measures (Virudachalam et al. 2024). These requirements evolve with instructional guidance, administrations, and technologies, but they do not fully determine how teachers reach the goals set for them (Spillane et al. 2019). Teachers must therefore update and evaluate classroom materials and lesson plans as requirements and student needs change. They work, on average, 15 hours beyond their contracts. Most feel overworked (Steiner et al. 2023), and resource constraints contribute to turnover (Keppler et al. 2025b).

Business leaders such as Sam Altman, the CEO of OpenAI, and Sal Khan, the CEO of Khan Academy, have emphasized the potential for generative AI to help teachers and schools meet instructional needs (Gates 2024, Olster 2023). Both OpenAI and Anthropic have dedicated considerable effort to create teacher-friendly AI tools for the K12 level (i.e., FERPA-compliant and useful for generating a range of common materials), but evidence from practice remains limited (Klopfer et al. 2024).

2.2. Generative AI for Education and Work

Research about generative AI in education is emerging and growing, but it largely focuses on student use. Several studies raise concerns that generative AI harms student learning by taking away struggle and cognitive effort that is essential for true understanding (Baek et al. 2024, Bastani et al. 2025, Kulesa et al. 2025). Unrestricted AI access can let students bypass the *zone of proximal development*, in which problems are challenging enough to stretch their thinking but still accessible (Poulidis et al. 2025). With appropriate oversight, however, generative AI can support learning (Bastani et al. 2025). Personalized practice problems, for example, can help students spend more time working within that zone (Chung et al. 2026).

Recent evidence suggests teacher use of generative AI also needs guardrails: students are commonly less satisfied with AI-created teaching materials (Sungu et al. 2026). At the same time, there is considerable variation in how generative AI can be used. Qualitative studies document uses ranging from lesson planning to material creation and feedback, while also showing that districts, administrators, and teachers are still learning how to integrate generative AI into classrooms (Van den Berg and Du Plessis 2023, Keppler et al. 2025a). Evidence from other forms of knowledge work provides a useful comparison. Generative AI can improve the speed and quality of task execution in writing, consulting, and customer service (Noy and Zhang 2023, Brynjolfsson et al. 2023, Dell'Acqua et al. 2023, Chen and Chan 2024), suggesting similar potential for teacher preparation.

These settings differ from K12 teaching in one important respect: the task is externally defined, so the central margin is task execution. Even in these constrained settings, workers' engagement with AI can be difficult to predict. Research on discriminative AI for prespecified tasks documents this difficulty in service (Bastani et al. 2026, Snyder et al. 2026), pricing (Caro and de Tejada Cuenca 2023), and inventory decisions (Balakrishnan et al. 2026). K12 teachers must also determine which tasks and materials will advance students toward their goals. Their productivity therefore depends on both task execution and *task selection*. We study these decisions from the perspective of individual workers and complement an emerging body of research on how AI reshapes task organization and the division of labor at the firm level (Savva 2026).

2.3. Task Discretion and Productivity

Discretion around task selection is not unique to K12 education. Many knowledge workers face high-level goals with substantial freedom over how to achieve them, and this discretion has well-documented consequences for productivity. Studies in behavioral operations management show that moving from planning a work process to executing the tasks within it is not costless (e.g., Ibanez et al. 2018). Constraints on planning can therefore improve productivity. In product development, for example, the time workers spend deciding when to switch from ideation to execution can exceed the benefits of that flexibility (Kagan et al. 2018). Similarly, discretionary sequencing and selection can lead workers to prioritize easier tasks over longer ones, reducing productivity and throughput (Kc et al. 2020). Yet removing autonomy is neither always possible nor desirable. Discretion has operational value when services are personalized or workers possess local knowledge. In labor and delivery units, for example, midwives respond to patient needs by deciding when to escalate care (Freeman et al. 2017).

This literature treats discretion as an operational lever that improves fit to local conditions, while also introducing frictions. K12 teachers face both sides of this tradeoff. We next introduce a framework to capture their preparation work.

2.4. The Teacher's Knapsack

Teachers' discretion encompasses the task pacing and sequencing studied in prior work (Kagan et al. 2018, Kc et al. 2020), but both decisions occur after a more fundamental choice: which instructional tasks belong in the plan. We represent this upstream bundle decision as an inverse knapsack problem. Knapsack models describe choices among discrete items subject to a constraint (Kellerer et al. 2004, Pinedo 2012). The structure fits K12 preparation because teachers choose among non-divisible instructional tasks that collectively must meet a learning requirement, while seeking to limit the time required to prepare them.

For a given lesson, week, or unit, the teacher faces a learning requirement L, such as 80% student mastery of a standard. The teacher chooses among preparation tasks: explanations to develop, examples to prepare, activities to run, materials to create, assessments to write, and revisions to make for particular students or classes. These choices often concern instructional modality or format, such as whether to use a worksheet, guided discussion, in-class activity, personalized project, or assessment. They are not limited to choosing individual questions after a worksheet has already been selected. The alternatives can require substantially different amounts of preparation and can contribute differently to the learning goal.

The learning requirement is not a choice variable for most K12 teachers. Teachers are responsible for student mastery and accountable to standards and assessments. While some teachers work with students who have already exceeded the requirement, and thus would adopt a different objective, our time-minimizing formulation applies when the instructional requirement remains binding. Moreover, teachers with students who are not already above grade level are the same teachers who routinely work beyond their contracted hours (Steiner et al. 2023), making preparation time an outcome to reduce rather than a fixed budget. For these teachers, productivity means reducing preparation time while continuing to meet the instructional goal so that more standards can be taught in the given academic year. Unlike a standard knapsack that maximizes value subject to a time budget, we model the teacher's problem as minimizing time subject to L.

The choice of orientation is not essential to the distinction we develop. Appendix J.4 considers the standard formulation in which a teacher maximizes learning value subject to a preparation-time budget. The same distinction remains between evaluating a legacy bundle under post-AI conditions and reopening the bundle choice.

Let I^0 denote the finite set of tasks the teacher actively considers before the emergence of generative AI[1]. Each task $i \in I^0$ has preparation time $t_i > 0$ and learning value $v_i \geq 0$, where v_i captures the task's contribution toward L, such as developing understanding or preparing students for an external assessment. Let $x_i = 1$ if task i is selected and $x_i = 0$ otherwise. The baseline preparation problem is

$$\begin{aligned} \min_{x} \quad & \sum_{i \in I^0} t_i x_i \\ \text{s.t.} \quad & \sum_{i \in I^0} v_i x_i \geq L, \qquad x_i \in \{0,1\} \quad (i \in I^0). \end{aligned} \tag{1}$$

Let S^L denote the support of an optimal solution to (1). We refer to S^L as the teacher's legacy bundle: the set of tasks the teacher would choose under the pre-AI task environment and active consideration set. For any bundle $S \subseteq I^0$, define $T(S) = \sum_{i \in S} t_i$ and $V(S) = \sum_{i \in S} v_i$. Additivity

[1] We discuss changes in the consideration set in §4.

provides a transparent benchmark, but it does not require classroom activities to be pedagogically independent. Activities that create value only when used together can be treated as a composite task. More generally, the execution-selection distinction, the feasibility boundary, and the planning-cost threshold extend to bundle-level time and learning-value functions. The exact replacement representation and task-level value comparative statics in §4.3 use additivity to show how particular tasks generate bundle-level gains. Teachers are unlikely to solve (1) explicitly, and researchers generally cannot observe the task-level values of t_i and v_i. The formulation instead makes the structure of the preparation decision precise. It separates the learning goal, which the teacher must meet, from the task bundle, over which the teacher retains discretion. Generative AI can then affect productivity by changing task time, learning value, or the set of tasks under consideration.

3. Teacher Case Study: Data, Coding, and Observations

We conduct a multiple case study to understand how generative AI enters teacher preparation and to identify the decisions that the model should represent. Consistent with inductive research, we began without specific hypotheses (Eisenhardt 1989). We describe the sample and data collection before presenting three observations that emerged from the case analyses.

3.1. Sample

We recruited 24 US public school teachers in fall 2023. Each teacher represents one case. All teachers work in public schools in Michigan, Ohio, or Pennsylvania (see Table 1). Teachers vary in grade level (elementary, middle, high), subject area (math, science, ELA, social studies, foreign language, elementary education), and years of experience (from 2 to 27). Random samples are not appropriate for case study research (Eisenhardt 1989). Evidence from small samples is valuable to the operations management field for idea generation, novel theorizing, and innovation (Fisher 2007). Our sample is not representative and cannot be used to infer general behavior or patterns in the wider population. Our empirical goal is instead to observe qualitatively how teachers are beginning to use generative AI for their preparation work.

3.2. Data Collection

We collected multiple types of data across four periods during the 2023–2024 school year.

T_1**: Teacher Interview, Exposure, and Observation.** In fall 2023, each teacher participated in a one-hour one-on-one Zoom session. With a shared semi-structured protocol (Appendix B), one member of the research team conducted 13 interviews and another conducted 11 interviews. The call started with open-ended questions about how they prepare for class, what kinds of materials they create, whether they plan alone or with other teachers, and whether they had previously used any generative AI tools. All teachers had very limited to no experience using generative AI for

Table 1 Teacher Participants

ID	State	Grade Level	Subject Area	Years Teaching	T_1	T_2	T_3	T_4
1*	Ohio	Middle	General	12	Yes	No	No	No
2	Michigan	High	ELA	22	Yes	Yes	Yes	Yes
3*	Pennsylvania	High	Math	7	Yes	Yes	Yes	No
4	Michigan	High	ELA	21	Yes	Yes	Yes	Yes
5*	Pennsylvania	High	Foreign Language	12	Yes	Yes	Yes	No
6*	Pennsylvania	Elementary	General	25	Yes	Yes	Yes	No
7	Michigan	Elementary	General	6	Yes	Yes	No	No
8	Michigan	Elementary	General	12	Yes	Yes	Yes	No
9	Michigan	Elementary	General	3	Yes	No	Yes	No
10	Michigan	High	Science	25	Yes	Yes	Yes	No
11*	Pennsylvania	Middle	ELA	17	Yes	Yes	Yes	Yes
12	Michigan	High	Math	22	Yes	Yes	Yes	No
13	Michigan	Middle	Science	2	Yes	No	No	No
14	Michigan	High	Foreign Language	11	Yes	Yes	No	No
15	Michigan	High	Foreign Language	8	Yes	Yes	Yes	No
16	Michigan	High	Foreign Language	14	Yes	No	No	No
17	Michigan	High	Science	8	Yes	No	No	No
18	Michigan	High	Math	12	Yes	No	No	No
19	Michigan	Middle	Social Studies	27	Yes	Yes	Yes	No
20	Michigan	High	ELA	20	Yes	Yes	Yes	No
21	Michigan	High	Social Studies	9	Yes	Yes	Yes	Yes
22	Michigan	High	Math	2	Yes	Yes	Yes	Yes
23	Michigan	High	Math	27	Yes	No	Yes	No
24	Michigan	High	Science	3	Yes	Yes	Yes	No
Total Number of Teachers Per Round of Data Collection:					**24**	**17**	**17**	**5**

Notes: We recruited the sample through 88 targeted emails based on grade and subject level, including all responding teachers in the study. Five were teachers previously known to the authors (indicated by *). T_1, T_2, T_3, T_4 indicate teachers' participation in that period of data collection. The dashed horizontal line indicates whether T_1 came before a major version update.

their work at the time of the initial session. Only two teachers mentioned receiving any training on generative AI before our initial session. Thus, teachers in our sample should be understood as new generative AI users at T_1.

Next, we gave teachers a structured exposure to ChatGPT Plus.[2] Teachers logged into our ChatGPT Plus account on their own computers, shared their screen, and entered a set of pre-designed prompts in a prescribed order. The purpose of this standardized exposure was to ensure that teachers had a shared baseline understanding of what ChatGPT could and could not do. The structured exposure was followed by the key observational component of the study: teachers were asked to use ChatGPT for 15 minutes for their actual teaching work. We asked each teacher to pick something they had mentioned earlier and work toward a finished product, using any other software she would normally use, and told them it was fine not to finish (Appendix B gives the

[2] We chose ChatGPT because it was the most widely recognized generative AI tool at the time. Our sampled teachers confirmed all were aware of ChatGPT. Teacher 6, for example, shared, "I hear about it everywhere."

full instruction). We then remained silent as they worked, except when teachers narrated their process and invited minimal engagement. At the end of the call, we debriefed each teacher about the intention behind their prompts, their evaluation of the ChatGPT output, and their views about whether and how they might use ChatGPT in practice. The entire session was video and audio recorded, and participants were paid with a gift card for their time.

T_2 and T_3: Teacher Surveys. At T_2 (January 2024), we surveyed teachers (Appendix C) about their ongoing use of ChatGPT or other generative AI tools since T_1. The survey included open and closed questions about frequency and mode of use. At T_3 (May 2024), we surveyed the same teachers again (Appendix D). This survey repeated the generative AI use questions used in T_2, and added questions about how generative AI affected teachers' own learning, stress, number of tasks completed, hours worked, and work quality. These questions were asked separately for teaching, preparation, grading, and emailing. For both surveys, we received 17 out of 24 responses.

T_4: Targeted Follow-Up Interviews. In June 2024, we conducted five follow-up interviews (Appendix E) with teachers who indicated at T_3 that they had more to share with us about the impact of generative AI on their work. The set of five teachers ranged from avid users to non-users.

In total, we gathered 360 minutes of recorded observation of teachers using generative AI for their own work, 201 teacher-entered prompts and associated responses, 29 in-depth interviews, and 34 generative AI use surveys.

3.3. Data Analysis

We analyzed the data in two phases. In phase 1 of analysis, we focused on the observed teacher prompts inputted into generative AI at T_1. We transcribed each prompt verbatim into a spreadsheet. The 24 teachers entered 201 prompts in total, or 8.4 prompts each on average. We then iteratively derived a coding scheme of different generative AI use, first through open coding and then by refining descriptions of similar use cases. Two members of the research team independently coded all 201 prompts. Every prompt was assigned one code. The coders agreed on 91% of prompt classifications. Disagreements were resolved through discussion, and all authors agreed on the final coding.

The coding yielded four main prompt categories: *make for me*, *find for me*, *jumpstart for me*, and *iterate with me*. We used a fifth category, *show me what you can do*, for meta requests testing ChatGPT's capabilities. Table 2 summarizes the coding scheme. In the T_1 observation, *make for me* was the most common, accounting for 55% of prompts. *Find for me* accounted for 15%, *jumpstart for me* for 10.5%, *iterate with me* for 15.5%, and *show me what you can do* for 4%.

In phase 2 of data analysis, we used interviews and open-ended survey responses to interpret the pedagogical objectives behind observed prompts and subsequent AI use. We mapped descriptive

Table 2 Prompt Coding Scheme

Code	Description
Make for me	Requests for fully developed artifacts, such as problem sets, quizzes, stories, worksheets, or images, within well-defined parameters; the user engages ChatGPT as a task executor.
Find for me	Informational requests seeking pre-existing, factual information, such as existing facts, quotes, resources, or examples; the user engages ChatGPT as a search engine.
Jumpstart for me	Requests to initiate the development of often lengthy or complex materials, such as projects, activities, or unit plans; the user engages ChatGPT as a catalyst.
Iterate with me	Requests for advice, explanation, refinement, or rethinking of teaching approaches; the user engages ChatGPT as an idea-generator and sounding board.

interview data to the prompts using an inductive-deductive approach, asking not only what teachers requested but why. Similar prompts sometimes served different pedagogical goals. The analysis yielded a two-part categorization of teacher intent: using generative AI for *task execution* or for *task selection*. The prompt code and the execution-selection classification capture different features of an interaction. The first records the form of the request; the second uses the surrounding evidence to determine whether the teacher had already selected the task. Task execution was the goal of all *make for me* prompts, whereas task selection was the goal of all *iterate with me* prompts. *Find for me* and *jumpstart for me* prompts appeared in both categories. The examples below illustrate this distinction and connect the observed uses to the teacher's knapsack in §2.4.

3.4. Generative AI for Task Execution versus Task Selection

All 24 teachers in our sample engage in both task selection and task execution as part of their weekly, and even daily, class preparation. Some teachers have near total discretion about how to reach learning goals. Teacher 2, a high school ELA teacher, is the only person at her school teaching one course. She says, "I create all my own materials." Teacher 4, also a high school ELA teacher, explained that at his school "each of us is like our own little curriculum and assessment shop" that decides what to teach and makes the materials needed to teach it. Likewise, Teacher 19, a middle school social studies teacher, shared that "We are in charge of writing our own curriculum." However, even teachers provided with materials, through a purchased curriculum or a lead teacher, described selecting and executing tasks regularly. So did the International Baccalaureate (IB) teachers in our sample, who are handed learning criteria which are extensions of federal and state learning objectives (see §F for examples) but decide for themselves which activities meet them. Curriculum adoptions and standards revisions trigger years in which that effort rises sharply. Viewed through the baseline knapsack introduced in §2.4, these teachers first construct the pre-AI legacy task bundle, S^L, and later revisit it. Generative AI enters this work at two distinct points, distinguished by what the teacher brings to the interaction. Table 3 summarizes these two forms of use and illustrates each with representative prompts.

Table 3 Task Execution and Task Selection Uses of Generative AI

Use type	Empirical marker	Illustrative prompts from T_1 observation
Task execution	Teacher begins with an artifact in mind and asks AI to create, revise, format, or improve it.	"Make 10 multiple choice questions for chapter 7 for the story *Mi Proprio Auto*." "Create a 5 question quiz for 3rd grade on the topic of forces and motion." "Make slides for the teacher to present this all to the class."
Task selection	Teacher begins with a learning objective, instructional problem, or partially specified unit goal in mind, and asks AI to surface possible explanations, activities, examples, or materials.	"What manipulatives besides a number line can I use?" "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." "Construct a unit plan for teaching the binomial theorem at the Math HL level (use the IB syllabus)."

Observation 1: ***Generative AI can change task time and instructional value.***

The most common use of generative AI in our sample was task execution. Here the teacher initiates the interaction with a pre-specified artifact already in mind, and AI serves as a production tool. This use was pervasive: every teacher in our sample entered at least one prompt aimed at creating, revising, or formatting a specific artifact. Teachers described some highly structured or procedural tasks as particularly easy to complete with AI when verification and adaptation were limited. Teacher 5, a high school Spanish instructor, emphasized how readily she could verify a generated subjunctive quiz: "I already know the answers. I'm going to use this for their quiz on Thursday." Teacher 13, a science instructor, similarly used AI to help author a creative contextual background narrative for an upcoming project. These cases illustrate the model's potential time-saving channel, although our observations do not measure task completion times directly.

Elsewhere the salient change was in learning value. Teacher 3, a high school mathematics teacher, belongs to a team using the Interactive Mathematics Program (IMP), a well-respected curriculum, but must still find materials to motivate students. He prompted AI to generate visual-pattern puzzles and in-out tables, sequentially refining the output by instructing the model to increase, and subsequently decrease, its complexity. He characterized the underlying task as difficult to make engaging, noting that "it's such a fundamental thing to find a pattern and then use a variable to write the rule. But, you know, it's just hard to be inspired when you're talking about in-and-out tables." He copied the generated text into a Google Sheet already containing baseline in-out tables, concluding that while the product was "not word for word," his students were "going to see this tomorrow." While the task was already in his plan, AI raised what he expected it to deliver.

Time can also move in the other direction. Teacher 6, an elementary teacher, could not readily use a generated forces-and-motion quiz: "I would absolutely need to insert illustrations for each of the questions, since a lot of the students at their age that I have are very visual with regards to

their learning." Teacher 12, a high school math teacher, found that mathematical notation made outputs harder to transfer into a usable worksheet. Teacher 24, a high school science teacher, re-prompted repeatedly without obtaining images that satisfied ChatGPT's content policy, and said of the resulting lab write-up, "I couldn't just hand it over right away. I'd want to re-read it and make sure it made sense." Verification, formatting, and adaptation are real costs, and they fall unevenly across tasks. Additional vignettes appear in §A.1.

In knapsack terms, AI use for task execution changes the time or learning value of tasks already in S^L without reopening the bundle choice. Those changes are uneven: AI can speed teachers up or slow them down, and it can raise or lower the learning value of a task.

Observation 2: ***AI helps teachers reconsider which preparation tasks to undertake.*** The second use of generative AI in our sample was task selection. Here the teacher begins with a broad learning objective, an unresolved instructional challenge, or a partially specified unit goal rather than a specific artifact, and generative AI helps her surface, select, and define instructional tasks that were not fully specified before the interaction. Where task execution engages AI as a production tool, task selection engages it as a catalyst and sounding board. Task-selection use was considerably less common than task execution. More broadly, observed prompt-use patterns varied substantially across teachers, as the exploratory analysis in §G illustrates. In some interactions the result was a task the teacher had not previously planned. Teacher 1, a middle school special education teacher, faced the challenge of conveying how adding a negative and a positive number can still yield a negative result. Rather than requesting a specific document, she prompted for general explanations, real-world analogies, alternative manipulatives beyond standard number lines, and varied presentation formats. From the options surfaced, she arrived at something she had not intended to make: "Maybe I do take this word bank or key words and print it off on a half sheet. And they can keep that in their math notebook right there when they're doing their practice problems." The interaction led her to add a task she had not previously planned.

In others, what mattered was how cheaply AI could surface selection ideas. Teacher 4 used AI to search over instructional trajectories. Seeking to transition from short stories to a broader unit on great American novels while explicitly targeting critical analysis over rote plot comprehension, he first asked AI to define a "great American novel," then explored contemporary contextual texts related to Adventures of Huckleberry Finn and The Great Gatsby. The model's mention of Thorstein Veblen prompted him to retrieve the original work on JSTOR: "I saw this Veblen book and I was like, oh, I think I remember reading some passages from that in undergraduate when we were reading The Great Gatsby. I don't think I've ever done that [for my students]." This interaction directly changed the set of tasks he prepared. Generative AI surfaced the text within the interaction, after which he went to JSTOR to retrieve the source. The sequence illustrates how AI

can bring an alternative into consideration without establishing the magnitude of any search-time savings.

Searching is itself not free. Teacher 7, a first- and second-grade teacher, described the evaluation burden that followed an information-gathering exchange: "It's a lot to dig through still. So I still have to like, parse through it." Teacher 22, a high school math teacher, deferred learning to use the tool: "I will figure that out one day, you know. Then next month, or in the summer when there's more time." Surfacing alternatives is only half of selection; the teacher still has to judge them against her students and her time, and she has to be in a position to look in the first place.

The productivity self-reports collected at T_3 sharpen why the asymmetry between these two uses matters. Of the 17 teachers who responded in May 2024, 12 were using generative AI in their work. Among the users, the outcome was evenly split: six reported an improvement in productivity in at least one area of their work, six reported no change, and none reported a decline. Using generative AI was therefore not, on its own, sufficient for a productivity gain. A notable observed difference between the groups was the composition of their use rather than its volume: both groups increased their overall frequency of use over the school year, but the improved-productivity teachers increased their *iterate* and *jumpstart* use, while the no-change teachers reduced *iterate* use, concentrating their increase in *make* use. In knapsack terms, these interactions reopen the question of which tasks belong in the bundle. They can surface tasks beyond the teacher's active consideration set, I^0, and reduce the cost of finding and evaluating them.

Observation 3: ***Generative AI affects task execution and selection outside teacher use.*** So far, we have described the outcomes where teachers reach for ChatGPT. But generative AI also changed the terms of teachers' work through other avenues, chiefly through student use. This is clearest among non-users, as AI does not touch their tasks directly. Five of the 17 teachers who responded in May 2024 were not using generative AI at all by the end of the school year; none of them reported a productivity improvement over the year, and two reported that their productivity had declined. One high school ELA teacher who maintained a strict non-user status throughout the 2023–2024 academic year, Teacher 2, had historically assigned personal poetry. This task had once facilitated critical compositional skills and emotional processing. However, student access to generative AI resulted in uniform, algorithmic outputs, leading her to observe that now "every poem rhymes." The task became worth less. The channel is not confined to non-users, however. Teacher 4, whose own selection work we described above, noted quickly the same shift arrived in his classroom: "We just were alarmed at the incredible pace with which it seemed to take over content within this last school year." Some of his students were no longer using AI for one-off shortcuts but were "basically using it as a substitute for written language."

The time parameter also moves. The total time required to formulate writing assignments can expand with the new overhead of "AI-proofing," and Teachers 2, 22, and 24 all reported spending more time grading as student AI use increased. These shifts affect teachers at both ends of the adoption range. A teacher who never opens a chat window can face higher t_i on some legacy tasks and lower v_i on others. A frequent user faces the same environmental changes in addition to the effects of personal AI use.

These shifts matter beyond execution. A legacy task that has lost enough of its learning value has to be replaced, and replacing it means reconsidering the bundle. The non-using ELA teacher cannot keep assigning personal poetry and meet the same objective; something has to take its place. The qualitative data therefore suggest that generative AI reshapes teacher work even for teachers who never adopt it themselves. Non-use does not exempt a teacher from the selection problem described in Observation 2. It can force her into it, on terms she did not choose and without the search support that AI offered the teachers in that observation. These two margins, changes in task characteristics and changes in the bundle itself, are what §4 formalizes.

4. The Teacher's Knapsack under Generative AI

Teachers in §3 used AI for both task execution and task selection. Accordingly, the model separates changes in task characteristics from the decision to retain or reopen the task bundle.

To isolate changes in task economics, suppose initially that the active task set remains $I = I^0$, with $S^L \subseteq I$, and define $x_i^L = \mathbf{1}\{i \in S^L\}$. Each task $i \in I$ has baseline preparation time $t_i > 0$ and baseline learning value $v_i \geq 0$. Because S^L was optimal over this same set before AI, any strict post-AI gain from changing the bundle must arise from a change in task time or learning value. Holding the candidate set fixed also rules out gains that arise solely because AI surfaces additional tasks. §4.4 relaxes this restriction by allowing AI to expand the candidate set.

4.1. AI Effects on Task Time and Learning Value

Observation 1 in §3 shows that generative AI can change both the preparation time and learning value of a selected task. It may reduce preparation time by helping teachers draft, format, translate, or adapt materials, but add time when they must verify, revise, or redesign the resulting artifact. It may increase learning value through personalization, scaffolding, or engaging examples, yet reduce it when students bypass the intended work or generated material misses the learning objective.

Let $\tilde{t}_i > 0$ and $\tilde{v}_i$ denote the realized post-AI preparation time and learning value of task i. Define the time multiplier $\alpha_i := \tilde{t}_i / t_i > 0$ and the value change $k_i := \tilde{v}_i - v_i \in \mathbb{R}$. Equivalently,

$$\tilde{t}_i = \alpha_i t_i, \qquad \tilde{v}_i = v_i + k_i. \tag{2}$$

The ratio α_i records the proportional change in preparation time, while k_i records the change in learning value in the same units as v_i and L. Values of α_i below one represent time savings; values above one capture added work such as checking, revising, or redesigning an AI-generated artifact. Likewise, k_i may be positive for a task that becomes more effective and negative for a task whose instructional value falls. Because both effects vary by task, AI may make one activity faster while making another slower to verify or less useful for learning.

The core model treats $(\tilde{t}_i, \tilde{v}_i)$ as the task characteristics the teacher faces in the post-AI setting, including changes in the instructional environment, the teacher's own use of AI, or both. Appendix H separates these sources and allows the teacher to decide for which tasks to use AI.

For a bundle $S \subseteq I$, let $\tilde{T}(S) = \sum_{i \in S} \tilde{t}_i$ and $\tilde{V}(S) = \sum_{i \in S} \tilde{v}_i$. If AI changes which bundle meets the learning requirement in the least time, reopening the bundle can produce savings beyond those obtained by executing the original plan. Table 4 illustrates this possibility. The teacher considers three tasks, $I = I^0 = \{1, 2, 3\}$, and must attain learning value $L = 105$.

Table 4 Candidate Tasks in the Numerical Illustration

	Baseline		*Combined AI Effects*		*Post-AI*	
Task i	v_i	t_i	k_i	α_i	$\tilde{v}_i$	$\tilde{t}_i$
1. Worksheet	55	40	5	0.50	60	20.0
2. In-Class Activity	55	50	20	0.90	75	45.0
3. Personalized Project	70	120	10	0.24	80	28.8

Notes: v_i and $\tilde{v}_i$ are baseline and post-AI learning values. t_i and $\tilde{t}_i$ are preparation times in minutes. k_i is the additive change in learning value, α_i is the preparation-time multiplier, and $L = 105$ is the learning requirement.

Before AI, no single task meets the requirement. The worksheet and in-class activity form the least-time feasible bundle, so $S^L = \{1, 2\}$, with $T(S^L) = 90$ and $V(S^L) = 110$. Task execution retains this bundle after AI. Its preparation time falls to $\tilde{T}(S^L) = 65$, while its learning value rises to $\tilde{V}(S^L) = 135$. The legacy-bundle execution effect is therefore $90 - 65 = 25$ minutes.

Task selection reopens the bundle decision. The personalized project was prohibitively time-consuming before AI, but its post-AI preparation time is 28.8 minutes. Replacing the in-class activity with the personalized project produces $S^{AI} = \{1, 3\}$, with $\tilde{T}(S^{AI}) = 48.8$ and $\tilde{V}(S^{AI}) = 140$. Task selection therefore saves an additional $65 - 48.8 = 16.2$ minutes relative to executing the legacy bundle, bringing total post-AI time savings to $90 - 48.8 = 41.2$ minutes. The additional saving arises from substituting tasks, not from executing the original bundle faster.

The sequence $90 \to 65 \to 48.8$ separates the two sources of time savings. AI first reduces the preparation time of the legacy bundle and then creates a further gain when the teacher reopens the task decision. The general problem is to characterize when such substitution gains arise and whether they are large enough to justify finding and evaluating alternatives.

4.2. Task Execution and Task Selection

We first compare the two responses without charging for search. Task execution imposes $x_i = x_i^L$ for every task and evaluates the legacy bundle using the post-AI characteristics. Task selection allows the teacher to choose a new bundle from the same candidate set and solves

$$\begin{aligned} \min_{x} \quad & \sum_{i \in I} \tilde{t}_i x_i \\ \text{s.t.} \quad & \sum_{i \in I} \tilde{v}_i x_i \geq L, \qquad x_i \in \{0,1\} \quad (i \in I). \end{aligned} \tag{3}$$

Assume that at least one bundle is feasible for (3), and let S^{AI} denote the support of an optimal solution. When the legacy bundle remains feasible, $\tilde{V}(S^L) \geq L$, define the task-selection gain

$$G = \tilde{T}(S^L) - \tilde{T}(S^{AI}). \tag{4}$$

The task-selection gain G is the time saved by reopening the bundle after its task characteristics have changed. Both bundles are evaluated using the same post-AI characteristics, so G excludes the baseline-to-post-AI execution effect $T(S^L) - \tilde{T}(S^L)$. The additional 16.2-minute saving in Table 4 is an instance of G. If the legacy bundle is no longer feasible, retaining it is not an available response; §4.5 considers that case.

Frictionless benchmark. When S^L is feasible for (3), it remains one of the available choices. Optimality therefore implies

$$\tilde{T}(S^{AI}) \leq \tilde{T}(S^L), \qquad G \geq 0.$$

The inequality follows directly from the choice set: task selection can always reproduce task execution by retaining S^L. Equality holds exactly when S^L is itself optimal for (3). A strict gain exists exactly when there is a feasible bundle S with $\tilde{T}(S) < \tilde{T}(S^L)$. Such a gain need not be realized once the cost of finding and evaluating the alternative is taken into account.

When S^L remains feasible, the total preparation-time saving relative to the pre-AI legacy bundle has the exact decomposition

$$\underbrace{T(S^L) - \tilde{T}(S^{AI})}_{\text{total post-AI time saving}} = \underbrace{T(S^L) - \tilde{T}(S^L)}_{\text{legacy-bundle execution effect}} + \underbrace{\tilde{T}(S^L) - \tilde{T}(S^{AI})}_{\text{task-selection gain } G}. \tag{5}$$

The left-hand side is a signed measure: a positive value is a net time saving, while a negative value means that preparation takes longer after AI. The first term is the time change from executing the legacy bundle and may be positive or negative. The second is the additional saving from changing the bundle and is nonnegative in the frictionless comparison. The distinction matters when AI raises the learning value of the legacy bundle without reducing its preparation time. With the

bundle fixed, the improvement changes value but not time; it becomes a time saving only if the teacher can remove or replace work.

Equation (5) also clarifies what G does not measure. It is not AI's total productivity effect; it is the value of reopening the bundle when both alternatives are evaluated under the same post-AI conditions. A slowdown in a legacy task can raise G by worsening the retention benchmark even as the legacy-bundle execution effect falls. A larger gap then signals a worse status quo, not a better outcome. The signed time saving on the left side of (5) accounts for both changes.

4.3. Task Substitution and Density Reordering

The numerical illustration in Table 4 considers only three tasks. With n candidate tasks, however, the teacher faces as many as 2^n bundles, and the underlying 0/1 knapsack is NP-hard (Karp 1975). Teachers need not solve the problem explicitly for search and evaluation costs to matter. Even informal comparison becomes demanding when a teacher must anticipate how activities fit together, whether they collectively meet the learning goal, and how much adaptation each requires.

The numerical example replaces one legacy task with one alternative, but task selection need not be a one-for-one exchange. In the general problem, let $A \subseteq S^L$ denote the legacy tasks removed from the bundle and let $B \subseteq I \setminus S^L$ denote the excluded tasks added to it. The resulting bundle is $S(A,B) = (S^L \setminus A) \cup B$.

LEMMA 1 **(Task-substitution representation)**. *Suppose the legacy bundle remains feasible, $\tilde{V}(S^L) \geq L$. Then*

$$G = \max_{\substack{A \subseteq S^L,\, B \subseteq I \setminus S^L \\ \tilde{V}(B) - \tilde{V}(A) \geq L - \tilde{V}(S^L)}} \left\{ \tilde{T}(A) - \tilde{T}(B) \right\}. \tag{6}$$

Lemma 1 expresses the task-selection gain as a substitution problem around the legacy bundle. The constraint requires the tasks in B to replace enough of the learning value removed with A, after accounting for any slack already present in S^L. The objective subtracts the preparation time added through B from the time removed with A. Because choosing $A = B = \varnothing$ reproduces the legacy bundle, the maximum cannot be negative. It is strictly positive exactly when some feasible substitution removes more preparation time than it adds.

Let $s = \tilde{V}(S^L) - L \geq 0$ denote the post-AI learning-value slack in the legacy bundle.

COROLLARY 1 **(Learning-value slack and time productivity)**. *If there is a nonempty set $A \subseteq S^L$ such that $\tilde{V}(A) \leq s$, then the teacher can remove A without adding a replacement, and*

$$G \geq \tilde{T}(A) > 0.$$

With a fixed bundle, an increase in learning value need not save preparation time; it instead creates slack above L. If the slack is large enough, task selection lets the teacher eliminate or replace work and convert the value improvement into a time saving.

Insight 1: Task Selection Converts Learning Value into Time Savings

With a fixed bundle, higher learning value creates slack rather than time savings. Task selection turns that slack into saved preparation time by removing or replacing work.

Generative AI can change relative task attractiveness because its capabilities are uneven across tasks, consistent with the jagged frontier described by Dell'Acqua et al. (2023). The value of reopening the bundle therefore depends on which tasks improve. To make the role of the candidate set explicit, write the task-selection gain as $G(I)$.

PROPOSITION 1 **(Sources of the task-selection gain)**. *Suppose the legacy bundle is feasible over candidate set I. Holding fixed all other task characteristics:*

(i) $G(I)$ *is weakly increasing in* $\tilde{t}_i$ *for every legacy task* $i \in S^L$*;*

(ii) $G(I)$ *is weakly decreasing in* $\tilde{t}_j$ *for every excluded task* $j \in I \setminus S^L$*;*

(iii) $G(I)$ *is weakly increasing in* $\tilde{v}_i$ *for every task* $i \in I$*; and*

(iv) if $I \subseteq I'$ *and task characteristics agree on* I*, then* $G(I') \geq G(I)$*.*

Parts (i) and (ii) show why the location of a time reduction matters. Accelerating a legacy task lowers $\tilde{t}_i$, making the current bundle cheaper to retain and weakly reducing $G(I)$. Accelerating an excluded task lowers $\tilde{t}_j$, making a potential substitute more attractive and weakly increasing $G(I)$. In the numerical example in Table 4, the large time reduction for the personalized project helps make bundle $\{1,3\}$ faster than the legacy bundle.

Learning-value improvements operate differently. Raising $\tilde{v}_i$ never removes a previously feasible bundle from consideration and may make a lower-time bundle feasible, so $G(I)$ weakly increases. The same logic applies to candidate discovery: adding alternatives cannot worsen the best feasible bundle. Neither result guarantees a strict gain. Indivisibilities may prevent a whole-task substitution, and a newly considered task may never enter an optimum. Table 5 pairs these comparative statics with AI impacts that recur in the cases.

The ruined-bundle case marks a boundary of these comparative statics. As long as S^L remains feasible, lowering any task's learning value weakly reduces G, holding the other primitives fixed, because a low-time alternative may cease to meet the requirement. Once $\tilde{V}(S^L) < L$, retaining the legacy bundle is infeasible and G is no longer the relevant discretionary comparison. Task degradation can therefore make revision necessary even as it reduces the time-saving gain measured while the legacy bundle remains feasible.

COROLLARY 2 **(Uniform acceleration preserves the bundle ranking)**. *Suppose* S^L *is optimal over the common candidate set* I *under baseline task characteristics. If AI changes no task values and scales every task time by the same factor* $\alpha > 0$*, so that* $\tilde{v}_i = v_i$ *and* $\tilde{t}_i = \alpha t_i$ *for every* $i \in I$*, then* S^L *remains optimal and* $G = 0$*.*

Table 5 How Generative AI Changes the Value of Task Selection

AI impact	Primitive change	Effect on G	Operational interpretation	Illustrative case
Task unlocking	$\tilde{t}_j \downarrow$ or $\tilde{v}_j \uparrow$, $j \notin S^L$	Weakly increases	An excluded alternative becomes more competitive. Strict substitution still depends on whole-task feasibility.	Teacher 1's printable word bank
Legacy-task speed boost	$\tilde{t}_i \downarrow$, $i \in S^L$	Weakly decreases	Familiar work becomes cheaper to retain, which can crowd out task selection.	Teacher 5's subjunctive quiz
Value or quality boost	$\tilde{v}_i \uparrow$	Weakly increases	Higher value can create slack that allows the teacher to remove or replace preparation work.	Teacher 3's in-out tables
Candidate discovery	$I \subseteq I'$	Weakly increases	Additional alternatives expand the feasible choice set.	Teacher 4's Veblen text
Ruined legacy bundle	$\tilde{V}(S^L) < L$	Not applicable	Legacy execution is infeasible. Revision is necessary if another feasible bundle exists, and G is no longer a discretionary gap.	Teacher 2's personal poetry assignment
Legacy-task slowdown	$\tilde{t}_i \uparrow$, $i \in S^L$	Weakly increases	The retention benchmark worsens; this need not represent an overall productivity gain.	Teacher 6's quiz requiring illustrations

Notes: G is the preparation time saved per use by selecting a new bundle rather than retaining S^L, as defined in (4). Each direction holds the other primitives fixed. The candidate-discovery row also holds task characteristics on I fixed. Except in the ruined-bundle row, S^L remains feasible. The directions describe the gain with zero planning cost. The cases in §3 illustrate the corresponding mechanisms; they do not identify task-level primitives or causal changes in G.

Uniform acceleration can create substantial execution gains, but it does not create a task-selection gain when the candidate set and task values are unchanged. In contrast, a strict task-selection gain can arise when AI changes task attractiveness, creates value slack, or expands the candidate set.

Insight 2: The Location of AI Gains Matters

Making legacy tasks faster lowers the value of reopening the bundle, whereas making excluded alternatives faster raises it. Higher task value and a broader candidate set can also increase the return to task selection.

Post-AI density provides local intuition for these changes. For any task with positive post-AI learning value, define

$$\tilde{d}_i = \frac{\tilde{v}_i}{\tilde{t}_i} = \frac{v_i + k_i}{\alpha_i t_i}. \tag{7}$$

Remark 1 (Density in the fractional relaxation). Consider the fractional relaxation of (3), which replaces $x_i \in \{0,1\}$ with $0 \le x_i \le 1$. Suppose a feasible solution uses a positive fraction

of task i and leaves some of task j unused, with $\tilde{v}_i > 0$, $\tilde{v}_j > 0$, and $\tilde{d}_j > \tilde{d}_i > 0$. Shift ℓ units of learning value from i to j, where

$$0 < \ell \leq \min\{x_i \tilde{v}_i,\, (1 - x_j)\tilde{v}_j\}.$$

The bound ensures that the fraction of i remains nonnegative and the fraction of j does not exceed one. Total learning value is unchanged, while preparation time falls by

$$\ell \left(\frac{1}{\tilde{d}_i} - \frac{1}{\tilde{d}_j} \right) > 0.$$

Density therefore gives a useful local comparison: holding learning value fixed, shifting effort toward the higher-density task saves time. The teacher's problem is discrete, however, so this calculation does not imply that two whole tasks can be exchanged. Equation (6) supplies the corresponding whole-task test by requiring the complete replacement bundle to meet L.

4.4. The Decision to Reopen the Bundle

Observation 2 shows that task-selection use was less common than task-execution use, even though AI could surface alternatives quickly. Identifying candidate tasks was only one part of the work: the teacher still had to judge their classroom fit, adapt materials, and verify quality. We represent this one-time effort with a fixed, time-equivalent planning cost incurred when the teacher reopens the bundle-selection problem.

Task selection need not involve AI. A teacher can search manually by drawing on experience, colleagues, curricular repositories, or other familiar sources. AI can support the same decision by helping the teacher generate, evaluate, and adapt alternatives. To isolate this planning-cost channel, both methods search over the candidate set I and therefore target the same frictionless solution S^{AI} in (3). Let $\kappa^M \geq 0$ and $\kappa^{AI} \geq 0$ denote the one-time costs of manual and AI-assisted planning, respectively. The superscript records how the teacher searches over bundles; both methods use the same post-AI task characteristics to evaluate the resulting bundle.

A revised bundle may be reused across sections, repeated lessons, or future course offerings. Let $H \geq 1$ denote the effective reuse count, including the current use. The count can incorporate uncertainty and discounting, so it need not be an integer. Because G is the preparation time saved each time the bundle is used, HG is the cumulative preparation time saved across H effective uses. This quantity has the same units as the one-time planning cost. The net cumulative time savings from manual and AI-assisted planning are therefore

$$N^M = HG - \kappa^M, \qquad N^{AI} = HG - \kappa^{AI}. \tag{8}$$

PROPOSITION 2 **(Costly task selection)**. *Suppose the legacy bundle remains feasible. For* $q \in \{M, AI\}$*, task selection under* q *strictly improves on retaining the legacy bundle if and only if*

$$HG > \kappa^q.$$

Among retaining the legacy bundle, manual task selection, and AI-assisted task selection, the optimal net saving is $\max\{0, N^M, N^{AI}\}$*. Some form of task selection is strictly preferred to retention if and only if* $\max\{N^M, N^{AI}\} > 0$*. If this maximum is positive, any task-selection method attaining it is optimal; if it equals zero, retention and every task-selection method attaining zero are tied.*

Proposition 2 explains why the frictionless gain need not appear in practice. A teacher can know that another approach may be better and still retain S^L when its cumulative time saving is too small to recover the effort required to find, evaluate, and adapt it. Reuse matters because the planning cost is incurred once, whereas the time saving recurs.

COROLLARY 3 **(AI-assisted planning: cost channel)**. *Suppose manual and AI-assisted task selection use the same candidate set* I*, so both produce gap* G*. If* $\kappa^{AI} < \kappa^M$*, AI-assisted planning makes task selection strictly preferable to retention while manual planning does not precisely when*

$$\kappa^{AI} < HG \leq \kappa^M.$$

AI may also contribute through discovery. If it expands the candidate set from I to $I' \supseteq I$, holding the characteristics of tasks in I fixed, Proposition 1(iv) implies $G(I') \geq G(I)$. Planning support can therefore increase the return to task selection by lowering its cost, broadening the set of alternatives, or both.

The result makes only a set-inclusion claim: it does not specify which tasks AI surfaces or how they enter consideration. Observation 2 provides an example. AI's reference to Thorstein Veblen led Teacher 4 to consider a text he had not previously planned to use. The case identifies an expansion of the candidate set, not its causal effect on G. Nor must discovery produce a strict gain. A newly considered task may never enter an optimal bundle, and nonnested candidate sets cannot be ranked without comparing the bundles they contain.

PROPOSITION 3 **(Execution acceleration can crowd out task selection)**. *Fix* $q \in \{M, AI\}$*, and suppose* S^L *remains feasible. Hold* H*,* κ^q*, task values, and the candidate set* I *fixed.*

(i) The net value N^q *is weakly increasing in* $\tilde{t}_i$ *for every* $i \in S^L$ *and weakly decreasing in* $\tilde{t}_j$ *for every* $j \in I \setminus S^L$*. Thus, accelerating a legacy task weakly reduces the incentive to undertake task selection, whereas accelerating an excluded alternative weakly increases it.*

(ii) If all task times in I are multiplied by a common factor $\mu > 0$, with all other primitives fixed,

$$G(\mu) = \mu G(1), \qquad N^q(\mu) = H\mu G(1) - \kappa^q.$$

Within $\mu \in (0,1]$, a smaller μ represents stronger common acceleration and weakly reduces the net value of incurring a fixed planning cost.

(iii) Because $G \leq \tilde{T}(S^L)$, method q is not strictly preferred to retaining the legacy bundle when

$$H\tilde{T}(S^L) \leq \kappa^q.$$

If $H\tilde{T}(S^L) \leq \min\{\kappa^M, \kappa^{AI}\}$, neither task-selection method is strictly preferred to retention.

Part (i) of Proposition 3 shows that crowd-out depends on where acceleration occurs. When acceleration is concentrated on legacy tasks, replacing those tasks recovers less time, and the current bundle becomes harder to replace. Acceleration of excluded alternatives has the opposite effect because it makes substitution more valuable. Holding H and κ^q fixed, accelerating a legacy task cannot turn nonadoption into strict adoption, whereas accelerating an excluded task cannot turn strict adoption into nonadoption. Part (ii) identifies a scale effect. Even when relative task rankings are unchanged, common acceleration leaves less time saving available to recover the planning cost.

Part (iii) provides a conservative sufficient condition for nonadoption. It rules out task selection even under the extreme benchmark in which selecting a new bundle eliminates all recurring preparation time. Task selection may still be unattractive when $H\tilde{T}(S^L) > \kappa^q$ because the best alternative bundle requires preparation time.

AI can therefore change the planning decision through opposing channels. It may lower κ^{AI} by making alternatives easier to generate and evaluate, while legacy-task acceleration lowers G by making familiar tasks less costly to retain. Without additional assumptions, the net effect on $N^{AI} = HG - \kappa^{AI}$ is not signed. Experience using AI for execution may itself lower future planning cost; Proposition 3 holds κ^{AI} fixed to isolate the execution channel.

Insight 3: AI Can Promote or Discourage Task Selection

AI can make task selection more attractive by lowering planning cost or expanding the candidate set. It can make task selection less attractive when acceleration is concentrated on familiar tasks and reduces the value of searching for substitutes.

The comparison also clarifies why teachers can respond differently to the same technology. A teacher who routinely redesigns materials may face a low manual planning cost, while a teacher with a highly refined routine may have little to gain from replacing familiar tasks. AI-assisted planning is valuable when it surfaces viable substitutes and makes them easier to assess, not simply when it produces more suggestions.

4.5. AI Exposure Without Teacher Use

Observation 3 shows that the task-selection problem can change even when the teacher does not use AI. Student use may lower the learning value of a familiar assignment or add work needed to redesign, monitor, and grade it. These changes enter the core model through $\tilde{v}_i$ and $\tilde{t}_i$. For a non-user, they reflect changes in the instructional environment.

Before AI, the legacy bundle may exceed the learning requirement. Let $s_0 := V(S^L) - L \geq 0$ denote this extra learning value. Because AI changes task i's learning value by k_i, the total change in the value of the legacy bundle is $\sum_{i \in S^L} k_i$. For a teacher who does not personally use AI, these changes are the environmental effects k_i^E defined in Appendix H.

PROPOSITION 4 **(Baseline slack and legacy-bundle feasibility after AI)**. *Suppose S^L is feasible before AI. It remains feasible after AI if and only if*

$$\sum_{i \in S^L} k_i \geq -s_0.$$

If this condition fails but another bundle $S' \subseteq I$ satisfies $\tilde{V}(S') \geq L$, the teacher must change the legacy bundle to meet the learning requirement. Making the legacy tasks faster cannot restore feasibility because it does not repair the shortfall in learning value.

Baseline slack therefore separates two cases. When the condition holds, revising the bundle is optional, and the teacher does so only if the cumulative time saving exceeds the planning cost. When it fails, revision is necessary whenever another feasible bundle exists.

Appendix H makes the exposure-use distinction explicit by decomposing each task's time and learning value into environmental and teacher-use effects. It then allows AI assistance to be scarce across tasks. In that setting, planning creates another source of value: choosing the bundle and the allocation of AI assistance jointly can redirect limited capacity to tasks outside S^L, with a strict gain when the added flexibility changes the optimal bundle.

A teacher may therefore respond to generative AI in three ways: retain the legacy bundle and use AI only on selected preparation tasks, reopen the bundle when the expected savings justify the cost of planning, or avoid personal use where verification, policy, or ethical concerns make assistance unattractive. The third response does not restore the pre-AI environment. When exposure reduces the value of S^L below L, bundle revision is necessary even for a teacher who never uses AI.

5. From AI Use to Teacher Productivity

The model directs attention to where AI enters the teacher's workflow, which growth in total use does not reveal. We next examine three empirical traces: how teachers' modes of use changed over time, how their prompts reflected goal-directed interaction, and why some teachers limited

or avoided AI. Each speaks to whether AI was applied to tasks already chosen or to the choice itself, and to the constraints that shaped that decision. Because we observe neither a teacher's complete task bundle nor the bundle she would have chosen without AI, we treat these patterns as descriptive rather than as estimates of G, κ^{AI}, or optimality. Longer teacher–AI interaction vignettes appear in §A.

5.1. Use Trajectories and Reported Productivity

Of the twelve teachers using generative AI in May 2024, six reported a productivity improvement in at least one area of work and six reported no change. We code a teacher as reporting improved productivity if the teacher indicated doing more tasks in less time, the same amount of tasks in less time, fewer tasks in less time, or more tasks in the same amount of time in one or more areas of work: teaching, preparation, grading, and emailing.[3] We find no clear differences in the groups by grade level, subject area, or initial prompt variety. Both groups also increased their overall frequency of use over the school year by similar amounts, 0.6 and 0.4 points on the five-point scale, respectively. However, the two groups do differ in the composition of their AI use.

The decomposition in (5) shows why composition may matter even when total use grows by similar amounts. Both groups could save time by applying AI to work they had already planned. Additional savings from task selection require the teacher to reopen the bundle before execution. Table 6 shows that improved-productivity teachers increased *iterate with me* use by 1.2 points, while no-change teachers reduced it by 0.2 points and concentrated their increase in *make for me* use. Improved-productivity teachers also increased *jumpstart for me* use by 1.0 point. We interpret these modes cautiously. *Make for me* generally involves producing a known artifact, whereas *iterate with me* is more consistent with AI entering before the instructional direction is fully settled. *Jumpstart for me* can reflect either task execution or task selection, depending on whether the teacher has already chosen the task. *Find for me* is not counted on either side. A search-like prompt is task execution when the teacher is retrieving information for a task she has already chosen, and is closer to task selection only when it helps her compare or substitute among candidates.

These trajectories are consistent with the model's distinction between execution and selection, but they do not identify the crowd-out comparative static. Proposition 3 concerns where time reductions occur relative to S^L, whereas the survey records modes of use rather than task-specific time changes or bundle choices. The pattern therefore reflects different ways AI entered teachers' workflows; it does not show that execution gains reduced search, and reverse selection is possible.

[3] We did not specifically ask whether the tasks they were doing had changed.

Table 6 Change in Teacher Generative AI Use from 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. +1.0)	Increase (Avg. +0.4)
Iterate With Me	Increase (Avg. +1.2)	Decrease (Avg. -0.2)
Overall[a]	Increase (Avg. +0.6)	Increase (Avg. +0.4)

[a] 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.

5.2. Goal Clarity and Prompt-Level Traces

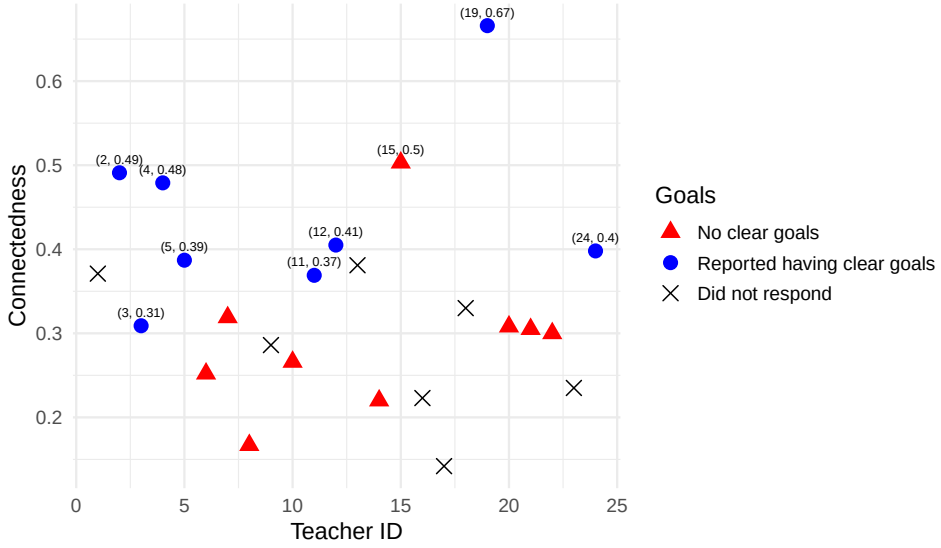
Task selection requires a teacher to generate or locate alternatives and then judge whether they serve the instructional objective. A clear objective can lower the cost of this process by providing a sharper criterion for evaluating candidates. It can increase the frictionless gain G only if it also changes which candidates are found or how their task characteristics are assessed. We examine two traces of this process in teachers' prompts, connectedness within a conversation and dispersion across repeated outputs. Neither measure identifies a bundle choice or an optimal plan.

Clear Goals. In the January survey, we asked teachers whether they had clear goals or visions for how they would use ChatGPT during the open-ended material creation stage. Among teachers who both reported clear goals and used ChatGPT more frequently, four of five later reported productivity increases. In the broader comparison, 72% of teachers reporting clear planning goals reported improvements, against 23% of those with ambiguous objectives. These comparisons are descriptive and involve small denominators. Teachers with clear goals often used ChatGPT across an extended planning sequence rather than for a single artifact. Teacher 18 began by asking for a unit plan for the binomial theorem on the IB syllabus, then moved through applications, problem types, and a solution key (§A). The interaction ended in execution but began with a learning objective and used AI to search over possibilities before settling on materials. This association is not evidence that goal clarity causes productivity improvement. It suggests that clear goals can make AI-supported task selection more directed and the resulting alternatives easier to evaluate.

Prompt Connectedness. We measure connectedness as the average pairwise cosine similarity of SBERT prompt embeddings within each teacher conversation, with implementation details in §J.1. A conversation organized around an instructional goal may remain semantically centered as the teacher considers activities, examples, and materials. Connectedness is not unique to task selection. Repeated refinement of an artifact that has already been chosen can also produce a highly connected sequence. Figure 1 groups teachers by whether they later reported clear goals. Teachers reporting clear and specific goals generally exhibited higher connectedness, whereas those reporting less clear goals generally exhibited lower connectedness. This pattern is consistent with

sustained, goal-directed interaction, but connectedness is neither a measure of productivity nor evidence that the teacher selected a different bundle.

Figure 1 Connectedness Scores of Prompts Within Conversations, by Self-Reported Goal Clarity



Output Dispersion. Finally, we examine how narrowly teacher prompts specify the desired response, re-running each prompt multiple times and measuring similarity across the resulting outputs (§J.2). Mean pairwise similarity is 78.09%, ranging from 30.11% to 94.75%.

Neither concentrated nor dispersed output is inherently preferable. For task execution, concentrated outputs may reduce post-processing and verification. For task selection, dispersion may expose alternatives the teacher had not considered, but it can also increase the effort required to compare and adapt them. As Remark 2 states, moving to a more dispersed prompting strategy is worthwhile only when the additional cumulative task-selection gain exceeds the additional evaluation and adaptation cost. The useful degree of dispersion therefore depends on what the teacher is trying to accomplish in that interaction.

5.3. Limits on AI Use

Five of the 17 teachers who completed the May 2024 survey were not using generative AI by the end of the school year, and none reported a productivity improvement. Their accounts show why use may remain limited even when AI can generate relevant materials. Planning competes with other demands on teachers' time; expected reuse may be too limited to recover the initial effort; and some tasks may be unsuitable for AI assistance because of verification, policy, or ethical concerns.

First, planning costs compete with everything else. Integrating generative AI requires adapting routines and work practices (Leonardi 2011), which imposes a short-run cost even when the

adaptation is valuable later. Teacher 22, facing possible layoffs announced in January, explained why she stopped: "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." Her account illustrates the opportunity cost represented by κ^{AI}: even potentially useful AI-assisted planning can be unattractive when the expected cumulative saving does not justify the effort required to learn and use it.

Second, the technology itself was changing. All six improved-productivity users were exposed earlier in the study period, before the November 2023 version change marked in Table 1. Among T_3 respondents, no teacher first exposed after that point reported a productivity gain by the end of the school year. Because exposure timing was not randomized and coincided with calendar time and other teacher differences, we treat this pattern as descriptive rather than as evidence of a version-update effect. The model nevertheless clarifies why rapid technological change can matter for task selection. If teachers expect plans built around the current technology to become obsolete more quickly, the effective reuse count H falls. If a changing tool requires additional learning, experimentation, or evaluation, the planning cost κ^{AI} rises. Either change reduces the net value of AI-assisted task selection, holding the task-selection gain G fixed. Our data do not identify whether either mechanism operated here, but they illustrate why technological instability can affect the planning decision through margins that are distinct from task-level performance.

Third, ethical aversion limits where AI use is acceptable at all. Teacher 2 summarized her stance as "it feels like cheating," adding that AI companies profit from uncompensated writing by professional authors. Teacher 21 raised similar concerns. Even users hesitated: Teacher 12 wrote, "I use it to list sometimes (5 main reasons for revolutionary war) but I question sources." The extension in §H.2 distinguishes three constraints behind these accounts. Ethical, legal, or policy restrictions can make AI unacceptable for a particular task. Prompting, checking, and revising can increase the time required to complete a task. Attention, access, or review capacity may also be scarce across tasks. Each constraint can limit use, but each has a different operational implication.

Choosing not to use AI does not restore the pre-AI instructional environment. Proposition 4 formalizes the implication of Observation 3: if student AI use or another environmental change pushes the learning value of S^L below L, faster execution of the same tasks cannot restore feasibility. The teacher must revise the bundle, whether the revision is conducted with AI or without it.

6. Discussion and Conclusion

Generative AI is often evaluated as a tool for executing assigned tasks faster or better. That view fits settings in which work arrives as a queue of externally specified tasks, such as customer service conversations, writing assignments in an experiment, or other standardized workflows (Noy and Zhang 2023, Dell'Acqua et al. 2023, Brynjolfsson et al. 2023). Teacher work is different. Teachers

operate under externally defined learning requirements, but they retain substantial discretion over the activities, examples, assessments, and materials used to meet those requirements. In such settings, the productivity question is not only whether AI improves a given task. It is whether AI changes the task bundle that a worker chooses to execute.

The teacher's knapsack formalizes this distinction. In the baseline model, the teacher selects the least-time bundle that achieves a learning requirement. The case observations motivate two uses of AI. In task execution, the teacher applies the post-AI task characteristics to the legacy bundle S^L. In task selection, the teacher reopens the bundle before execution and searches for the post-AI bundle S^{AI}. The gap $G = \tilde{T}(S^L) - \tilde{T}(S^{AI})$ is the additional time saved per use by replacing the legacy bundle. It increases when excluded alternatives speed up, when learning-value improvements create useful slack, or when discovery expands the candidate set. It decreases when acceleration is concentrated on tasks already in S^L.

The same model explains why many teachers do not use generative AI for task selection. Searching over candidate tasks, evaluating suggestions, adapting materials, and choosing to replace familiar routines impose a fixed cost. AI may lower this cost or expand the candidate set. Or, holding planning costs fixed, AI might accelerate legacy tasks, thereby reducing G. This opposing effect reconciles two patterns that otherwise appear in tension: teachers can benefit from AI-generated worksheets, quizzes, slides, or examples while still leaving task-substitution gains unrealized.

Returning to the case study, the evidence is consistent with this interpretation without establishing it. Teachers commonly used AI to produce familiar artifacts, while teachers reporting productivity gains increased their *iterate with me* use more than users reporting no productivity change. Other cases illustrate the planning, verification, and ethical frictions that can limit task selection or AI use. The data do not reveal whether teachers' task bundles changed or whether legacy-task acceleration crowded out search.

6.1. Implications for AI Support

The distinction between execution and selection changes what it means to support teachers well. A tool designed for execution can reduce the burden of producing a quiz, worksheet, or slide deck that the teacher has already chosen. A tool designed for selection must instead help the teacher work backward from a learning objective, surface credible alternatives, compare their classroom fit and preparation demands, and adapt a promising alternative into usable materials. The latter process depends on teacher judgment. Learning value, student fit, and ethical acceptability are contextual judgments that remain with the teacher.

Selection support does not mean presenting the largest possible menu. Expanding the candidate set can increase the gain from task selection, but the teacher must still screen, compare, and adapt

the alternatives. A system may therefore be more useful when it connects a manageable set of options to an explicit learning goal and makes their preparation demands, verification burden, and classroom fit legible. Generating alternatives and supporting their evaluation are complements: more options create value only when the teacher can assess them at reasonable cost.

For schools and districts, the same distinction suggests that training should go beyond artifact generation. Teachers may need support articulating learning objectives, evaluating AI-suggested substitutes, recognizing when student AI use has reduced the value of a legacy task, and deciding whether a redesigned bundle will be reused often enough to justify the planning effort. This does not diminish the value of execution support. Teachers face severe time constraints, and savings on routine preparation can be consequential. It does imply that adoption and prompt counts alone will miss whether AI changed the allocation of work or only made the existing allocation cheaper.

The fixed planning cost also indicates where selection support is likely to pay off. A unit plan, recurring project, rubric, or assessment used across classes or school years can spread the upfront effort over many uses. Shared planning can distribute that effort across teachers. Reuse is not guaranteed, however, because student needs, curricula, and school policies change. Schools can therefore prioritize selection support for recurring instructional decisions while preserving the teacher's ability to adapt a bundle to local conditions.

The model also changes what organizations should measure when evaluating teacher-facing AI. Artifact completion time captures execution but misses upfront planning, changes in task composition, and reuse. An evaluation should record which tasks are added or removed, the effort required to compare and adapt alternatives, and whether benefits recur across classes or school years. These measures distinguish local acceleration from a durable change in the work bundle.

6.2. Limitations and Future Research

The study has important empirical limits. The 24 cases were selected to capture variation in grade, subject, experience, and AI use, not to estimate population frequencies. Productivity outcomes are self-reported, and we do not observe each teacher's complete task bundle, task-specific preparation times, or the learning value students obtain from individual tasks. Prompt connectedness and output dispersion describe features of teacher interactions with AI; neither identifies the optimal bundle, realized time savings, or student learning. The evidence therefore motivates and illustrates the model's mechanisms rather than testing its comparative statics.

The model also represents the learning requirement as a scalar. With several standards, let the requirement be a vector $\mathbf{L}$ and task i's learning contribution be $\mathbf{v}_i$, and impose $\sum_i \mathbf{v}_i x_i \geq \mathbf{L}$ componentwise. This produces a multidimensional covering knapsack. The distinction between executing a legacy bundle and reopening the bundle remains, but AI can now change which requirements a task covers as well as the time required to prepare it.

These limits point to several empirical tests. Field or laboratory experiments could compare execution-oriented tools with tools that explicitly support task selection to see whether the selected bundle changes. Training interventions could vary whether teachers receive help defining objectives, comparing substitutes, or planning for reuse. A direct test of crowd-out would also need to show where execution gains occur: improvements to tasks already in the legacy bundle should reduce the return to search, holding planning costs fixed, whereas improvements to excluded alternatives should increase it. Finally, studies linking bundle changes to student outcomes could determine whether time savings preserve or improve learning rather than merely shifting preparation effort.

The mechanism may also apply beyond teaching. Managers, consultants, designers, and researchers often work toward specified goals while choosing the analyses, artifacts, or workstreams used to reach them. In those settings, AI can change productivity through the cost and value of current tasks and through the composition of the task bundle. The knapsack structure is most useful when three features coincide: the worker faces a high-level requirement, can choose among discrete substitutes or complements, and retains discretion over the bundle. It is less suitable when work arrives as an exogenous queue, must follow a fixed sequence, or depends primarily on congestion and timing. Queueing, routing, or scheduling models may better describe those workflows. The broader execution-selection distinction can still apply, but its operational representation should follow the structure of the work.

6.3. Conclusion

This paper studies generative AI productivity in a setting where workers have discretion over what work to do. In K12 teaching, teachers must meet learning requirements, but they choose the bundle of activities, examples, assessments, and materials through which students work toward those requirements. Generative AI therefore affects productivity through two margins. It can help teachers execute known tasks, and it can help teachers reconsider which tasks belong in the bundle.

The teacher's knapsack shows why the second margin matters. When AI changes task time and learning value unevenly, productivity gains can arise from substituting toward newly attractive tasks. If AI exposure makes the legacy bundle infeasible, such revision may be necessary even without personal teacher use. These responses are not automatic. They depend on planning cost, goal clarity, reuse, evaluation capacity, and ethical or institutional limits on AI use. The same technology can therefore save time locally while leaving larger bundle-level gains unrealized.

Teacher-facing AI should therefore be judged by more than how quickly it produces an artifact. It should also be judged by whether it helps teachers identify, evaluate, and implement a more productive bundle of instructional work. More broadly, the productivity effect of a technology depends on the discretion workers retain over the tasks it changes.

References

Baek, C., Tate, T., and Warschauer, M. (2024). "chatgpt seems too good to be true": College students' use and perceptions of generative ai. *Computers and Education: Artificial Intelligence*, 7:100294.

Balakrishnan, M., Ferreira, K. J., and Tong, J. (2026). Human-algorithm collaboration with private information: Naïve advice-weighting behavior and mitigation. *Management Science*, 72(1):265–284.

Bastani, H., Bastani, O., and Sinchaisri, W. P. (2026). Improving human sequential decision making with reinforcement learning. *Management Science*, 72(1):733–755.

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., and Mariman, R. (2025). Generative ai without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences*, 122(26):e2422633122.

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

Caro, F. and de Tejada Cuenca, A. S. (2023). Believing in analytics: Managers' adherence to price recommendations from a dss. *Manufacturing & Service Operations Management*, 25(2):524–542.

Chen, Z. and Chan, J. (2024). Large language model in creative work: The role of collaboration modality and user expertise. *Management Science*, 70(12):9101–9117.

Chung, A. T.-H., Zhang, B., Kung, L.-C., Bastani, H., and Bastani, O. (2026). Effective personalized ai tutors via llm-guided reinforcement learning. *Available at SSRN*.

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).

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.

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

Gates, B. (2024). Unconfuse Me with Bill Gates. https://www.gatesnotes.com/Unconfuse-Me-podcast-with-guest-Sam-Altman.

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.

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

Karp, R. M. (1975). On the computational complexity of combinatorial problems. *Networks*, 5(1):45–68.

Kc, D. S., Staats, B. R., Kouchaki, M., and Gino, F. (2020). Task selection and workload: A focus on completing easy tasks hurts performance. *Management Science*, 66(10):4397–4416.

Kellerer, H., Pferschy, U., and Pisinger, D. (2004). Multidimensional knapsack problems. In *Knapsack problems*, pages 235–283. Springer.

Keppler, S., Sinchaisri, W. P., and Snyder, C. (2025a). Making chatgpt work for me. *Proceedings of the ACM on Human-Computer Interaction*, 9(2):1–23.

Keppler, S. M., Li, J., and Wu, D. (2025b). Stopping the revolving door: A causal and textual analysis of crowdfunding and teacher turnover. *Manufacturing & Service Operations Management*, 27(6):1795–1813.

Klopfer, E., Reich, J., Abelson, H., and Breazeal, C. (2024). Generative ai and k-12 education: An mit perspective.

Kulesa, A. C., Mission, M., Croft, M., and Wells, M. K. (2025). Productive struggle: How artificial intelligence is changing learning, effort, and youth development in education. *Bellwether*.

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.

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

Olster, S. (2023). 34 big ideas that will shape 2024. *LinkedIn News*.

Pinedo, M. L. (2012). *Scheduling*, volume 29. Springer.

Poulidis, S., Bastani, H., and Bastani, O. (2025). Self-regulated ai use hinders long-term learning. *Available at SSRN 5604932*.

Reimers, N. and Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.

Savva, N. (2026). Process architecture in ai-augmented service operations: From horizontal adoption to vertical redesign. *Available at SSRN 6506624*.

Schaeffer, K. (2024). Key facts about public school teachers in the U.S.

Snyder, C., Keppler, S., and Leider, S. (2026). Algorithm reliance: Fast and slow. *Management Science*, 72(1):368–385.

Spillane, J. P., Seelig, J. L., Blaushild, N. L., Cohen, D. K., and Peurach, D. J. (2019). Educational system building in a changing educational sector: Environment, organization, and the technical core. *Educational Policy*, 33(6):846–881.

Steiner, E. D., Woo, A., and Doan, S. (2023). All work and no pay—teachers' perceptions of their pay and hours worked. *Rand Corporation*.

Sungu, A., Lira, B., and Duckworth, A. (2026). Generative ai can harm teaching. *Available at SSRN*.

Van den Berg, G. and Du Plessis, E. (2023). Chatgpt and generative ai: Possibilities for its contribution to lesson planning, critical thinking and openness in teacher education. *Education Sciences*, 13(10):998.

Virudachalam, V., Savin, S., and Steinberg, M. P. (2024). Too much information: When does additional testing benefit schools? *Management Science*, 70(9):6220–6233.

Appendix A: §3 Additional Empirical Vignettes

This appendix preserves longer empirical vignettes from the teacher observations and interviews. The main text uses a smaller number of examples to support the model-guided findings. The vignettes below provide additional evidence on the same distinction: task execution applies generative AI to a known artifact, whereas task selection helps the teacher decide what task or activity should enter the bundle.

A.1. Task Execution: Executing Known Artifacts

Teacher 3, a high school math teacher, used ChatGPT to create "in-out tables," or tables with two columns where the left-side column is the input and the right-side column is the output. The objective is to teach students about patterns and introduce the idea of functions as a "rule" through which an input is transformed into an output. The teacher explained he was hoping ChatGPT could help him come up with more creative, inspired in-out problems: *"It's such a fundamental thing to find a pattern establish and then use a variable to write the rule. It's such a critical thing. But, you know, it's just hard to be inspired when you're talking about in and out tables. And I mean, maybe there's other teachers out there that are more creative than me on something like this."*

Table A1 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.

As shown in Table A1, the teacher prompted ChatGPT to come up with visual, non-numerical in-out tables, "some goofy patterns, you know, with words." After prompting for a more complicated example, ChatGPT produced a table where the input was a word and the output was the number of vowels minus the number of consonants, and another where the input was a word and the output was the second vowel of the

word. The teacher said, *"I think this one, there's a pretty good one here. This one here, it's like the second vowel in the word, which is kind of hard to pick out."* He then copied the examples into a Google Sheet that already included other, more basic in-out tables he had prepared. He said, *"Not word for word, but my students are going to see this tomorrow. They'll fill in a table and create a rule or create a pattern and a table, and then they'll give it to the person next to them and they'll swap."* ChatGPT helped him produce a more creative version of a task he already intended to use.

Teacher 5, a high school Spanish teacher, similarly used ChatGPT to create immediate materials. She explained that her curriculum comes with a textbook and quizzes, and the school wants her to follow those materials closely. However, she said, *"I wanted to create a quiz on the subjunctive with the story that I'm teaching (*Mi Propio Auto*), rather than giving them a subjunctive quiz [from the textbook]."* She described the textbook quiz as a "drill sheet," but said, *"when it comes to engaging in classroom and direct instruction, I try my best to not implement that."* She felt it was more engaging and effective for students to learn grammar through stories and context. About the quizzes ChatGPT created, she said, *"It's reminding you of what happened in the story, while also hitting reading comprehension. Yep. And I'm also hitting the grammar concepts. So I'm hitting all these other components in a story rather than isolating each."* She concluded, *"I'm going to use this for their quiz on Thursday."*

Teacher 6, an elementary teacher, was not able to readily use the ChatGPT output. She teaches third graders, and for the forces-and-motion quiz she asked ChatGPT to create, she said, *"I would absolutely need to insert illustrations for each of the questions, since a lot of the students at their age that I have are very visual with regards to their learning. I'd have to be very specific about direction too, as a matter of fact, and push their eyes in the right direction so they can correctly identify what force is actually being demonstrated in the picture."* In this case, ChatGPT generated a starting point, but the task still required substantial teacher adaptation before it could be used.

Teacher 12, a high school calculus teacher, similarly liked the idea of generated math questions but found that using the output would require additional formatting work. After prompting ChatGPT for a question about absolute maxima, the teacher noted that math notation made the output harder to transfer into a usable worksheet: *"it's a bummer because some of these cool things that other teachers can use, makes it a little harder with the math notation."*

Teacher 24, a high school science teacher, had problems when ChatGPT would not create the materials he asked it to create. He asked ChatGPT to create a presentation with information about a science lab that he had attached, and he asked it to put images on the slides as well. However, ChatGPT would not create the images because it treated the request as violating content policy, presumably because the image of a child launching a marshmallow was flagged as potentially violent. He re-prompted several times in different ways, but could not get ChatGPT to make the image. About the lab write-up, he said, *"I would probably prompt it to put it into a worksheet or packet format so that students would have somewhere to write it. I would still have to go through and edit it to make sure that there was spacing, appropriate spacing, but I wonder if I could ask it to put appropriate spacing in there so students can hand-write in answers. I've just never tried that before. I usually just I end up formatting on my own afterwards. Also, I couldn't just hand it over right away. I'd want to re-read it and make sure it made sense."*

Together, these task-execution vignettes show why execution support can be valuable without fully resolving the productivity problem. ChatGPT helped teachers create or improve artifacts they already had in mind, but the residual work varied across tasks. In model terms, teacher-use AI can lower $\tilde{t}_i$ or raise $\tilde{v}_i$ for some tasks, while increasing verification, formatting, or adaptation burden for others.

A.2. Task Selection: Searching Over Candidate Tasks

Teachers did not always approach ChatGPT with a particular task in mind. Instead, some teachers had a broader workflow or learning objective and asked ChatGPT for help figuring out what to do. Teacher 1, a middle-school special education teacher, explained, *"I know our seventh graders are working with positive and negative integers."* Given her special education support role, she emphasized the importance of having many different ways to explain the same idea. She said, *"I explain it once and sometimes, in that moment, I can explain it a different way, but it's sometimes hard. We all get stuck in our own ways of explaining things, and sometimes it's hard to get our brains out of that."* She prompted ChatGPT to help her think through alternative explanations and examples, as shown in Table A2. Considering ChatGPT's output, she said, *"Maybe I do take this word bank or key words and print it off on a half sheet. And they can keep that in their math notebook right there when they're doing their practice problems. Rather than the small bank of the common ones that are always coming up."* In this case, the teacher did not begin with a specific artifact. ChatGPT helped her identify a task she might do next.

Teacher 4, a high school ELA teacher, also used ChatGPT in this way. He explained, *"We are reading short stories right now and it's leading to students reading in small book clubs, different short story collections. But what I'd like to do in American Lit is then move into a different book club where they read a great American novel. And I am wondering how to frame their reading such that they're doing more than just reading for basic comprehension, plot comprehension, all of that."* He was planning a pivot to "great American novels" and wanted help deciding how to structure the unit. As shown in Table A2, he originally asked about *Adventures of Huckleberry Finn* but then shifted to *The Great Gatsby*. He explained, *"At first, I didn't think about* The Great Gatsby*. First, I was just thinking about Great American novel, and then I, you know, asked first about Huck Finn."*

Reading ChatGPT's output describing the Jazz Age, materialism, and class made him want to explore nonfiction texts that could accompany the novel. ChatGPT's mention of Thorstein Veblen's work was useful because it reminded him of a text he had not previously planned to use: *"I saw this Veblen book and I was like, oh, I think I remember reading some passages from that in undergraduate when we were reading* The Great Gatsby*. I don't think I've ever done that [for my students]."* With this idea, he said, *"ChatGPT, you kind of gave me all I can use,"* and then went to JSTOR to find the actual text and passages for his students. He pulled up JSTOR and a Google Sheet and explained, *"I'm going to be working between these two. To try to build up this activity. With a work that's contemporary, like within a 20 year period to that novel that shows us something about the culture, so that people who are unfamiliar with the time period can learn a little bit about it through a different source, through a secondary source."* Again, the use of ChatGPT was not only to complete a task, but to identify what task should be done.

Table A2 Examples of Generative AI Prompts For Task Planning

Teacher ID	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

A final illustrative example comes from Teacher 7, a first- and second-grade teacher. She explained, *"I work really hard to talk about the fact that although all cultures are very different, there's so many similarities between them, like the harvest and the time of fall. And the transition of the seasons is like a universally appreciated experience...but at my grade level, they're very like me-centric."* She was looking for ChatGPT to help her understand how to help students see these commonalities across cultures. She said that her requests to ChatGPT were *"more like information gathering for me, than creation, which is what I honestly spend most of my time doing when I'm like trying to tackle a topic."* Considering ChatGPT's output about celebrations, books, and potential activities, she explained, *"It's a lot to dig through still. So I still have to like, parse through it. I really appreciated that I can say 'make it into a bulleted list' instead of a paragraph."* She was still deciding where to go next: *"Very excited about the books. I'm going to spend a little bit of the time here seeing if we have any of these books."* She summarized the value of ChatGPT as follows: *"I've used [ChatGPT] in the past to start hard tasks, to kind of get me past that. Do some of the grunt work tasks. But it can help compile resources or find book recommendations instead of me having to do all of that all the time."*

These task-selection vignettes illustrate the search channel in the model. Teachers used ChatGPT not only to generate an artifact, but to surface candidate explanations, readings, activities, and materials. Such use

can lower the cost of searching over the feasible set and can reveal substitutions that would not arise from applying AI only to the legacy bundle.

A.3. Selective Use, Aversion, and Mixed Responses

All teachers in our sample, even the non-users, were attuned to generative AI's ability to create output, even as some were averse to this capability and others were attracted to it. This was less true of the technology's ability to provide input into planning: that capability was highly salient to some teachers but did not appear equally salient to others. Only some teachers described using generative AI to help think through what work to do.

Several factors may help explain these differences in attention. First, the features of the technology itself can shape which capabilities are most visible to users. As ChatGPT evolved during the study period to support a broader range of generated outputs, teachers had increasingly salient opportunities to use it for production. Second, the way generative AI was discussed and encountered by teachers may also have foregrounded its generation capabilities relative to its use as a planning or thinking aid. Finally, competing demands on teachers' time limited their opportunities to experiment with the technology and learn less obvious ways of using it. These mechanisms are difficult to separate in our data, but together they help explain why output generation was salient across the sample while planning support was much more unevenly recognized.

The vignettes below show how task-specific gains, verification burdens, policy or ethical restrictions, and limited attention can produce different combinations of attraction and aversion. Improved-productivity teacher users often combined input-seeking and output-generation uses. Teacher 11, for example, explained her appreciation for generative AI as a source of input: *"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."* She also valued the generation capability, explaining, *"throughout the year, I've used it to make study guides, tests, quizzes."*

Meanwhile, non-users like Teachers 2 and 21 expressed aversion to what generative AI could provide, both as output and as input. Users of generative AI who did not report productivity gains often appeared to experience a combination of attraction and aversion to the technology's ability to create outputs. Teacher 4 explained this struggle in his 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."*

The task and AI-use discretion extension in §H.2 provides a compact interpretation of these mixed responses. Selective use can arise because teachers face task-specific verification burdens, ethical constraints, policy constraints, or limited cognitive bandwidth. In model terms, verification burdens can increase α_i^T, ethical, legal, or policy restrictions can impose $y_i = 0$, and limited attention, access, or review capacity can reduce C_A. Concerns that AI use lowers a task's learning value can also appear as lower k_i^T.

A.4. Adaptive Task-Bundle Revision in Teacher Material Creation

Material creation is a sequential and adaptive process in which teachers can choose to skip, implement, revise, or return to different types of tasks. One way to visualize this process is to consider four possible

tasks: (1) the lesson itself, such as slides; (2) student practice, such as a worksheet; (3) a student activity, such as a game; and (4) student evaluation, such as a quiz. Success in this example is student proficiency on the learning standard covered by the lesson, practice, activity, and evaluation.

In one workflow, the teacher does all four tasks in sequence and achieves success. In another workflow, the teacher skips practice, uses only an activity before evaluation, and also achieves success. In a third workflow, the teacher skips practice, but the evaluation reveals insufficient proficiency, so the teacher returns to create materials for additional practice. The teacher then re-evaluates and achieves success. As Teacher 24 explained, *"If I feel like students are struggling on a topic, I double back on stuff, especially for ninth grade."*

This workflow endogeneity is the practical reason the knapsack framing is useful. Teachers are not merely assigned a fixed sequence of artifacts to produce. They choose, revise, and sometimes revisit the task bundle as they learn about student needs and as the operating environment changes. Generative AI can therefore affect productivity by changing execution costs, but also by changing which tasks teachers consider worth doing.

Appendix B: T_1 Semi-Structured Protocol

Part 1 (10–15 minutes): 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 minutes): Observation

ChatGPT practice (10 minutes)

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 minutes)

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 minutes): Debrief

- Tell me about what you were 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 C: 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 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 to 5, where 1 = Really bad/not useful and 5 = Really good/useful.*
5. Did that (the quality of the outputs) match your expectation?
   - *Likert scale from 1 to 5, where 1 = Much worse and 5 = Much better.*
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 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
   - *Likert scale from 1 to 5, where 1 = Not useful and 5 = 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
   - *Likert scale from 1 to 5, where 1 = Never and 5 = 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 D: T_3 (May 2024) Additional Survey Questions

1. 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.

   - *Likert scale from 1 to 5, where 1 = Strongly Disagree and 5 = Strongly Agree.*

2. 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.

   - *Likert scale from 1 to 5, where 1 = Strongly Disagree and 5 = Stongly Agree.*

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?

   a) Prepping b) Teaching c) Grading d) Emailing

   - *Do more, Do about the same/no change, Do less*

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?

   a) Prepping b) Teaching c) Grading d) Emailing

   - *Do more, Do about the same/no change, Do less*

5. For each aspect of your job, what is currently true about the effect of generative AI on the quality of your work?

   a) Prepping b) Teaching c) Grading d) Emailing

   - *Higher quality, About the same quality/No change, Lower quality*

6. For each aspect of student work, what is currently true about the effect of generative AI on the quality of your students' work?

   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

   - *Higher quality, About the same quality/no change, Lower quality, NA*

Appendix E: T_4 (June 2024) Follow-Up Interview Questions

Questions for Improved-productivity Teachers

1. You indicated that you are seeing some productivity boosts from using ChatGPT. [Verify agreement.] Can you explain 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 agreement.] 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 agreement.] Can you talk about why?
   (a) Whether/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 F: Sample State Learning Standards

Table A3 Examples of 5th Grade Common Core State Standards in ELA and Mathematics

CCSS.	Standard
ELA-LITERACY.L.5.1.A	Explain the function of conjunctions, prepositions, and interjections in general and their function in particular sentences.
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).
ELA-LITERACY.L.5.3.B	Compare and contrast the varieties of English (e.g., dialects, registers) used in stories, dramas, or poems.
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.
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.
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.

Appendix G: Exploratory Patterns in Teachers' ChatGPT Use

The analyses in this appendix provide additional descriptive context for heterogeneity in teachers' observed ChatGPT use at T_1. Table A4 summarizes groups of teachers with similar prompt-use patterns, while Table A5 reports simple associations between teaching experience and observed use. Given the small, non-representative case-study sample, we treat these patterns as exploratory. The clusters are not behavioral types in the model, and the associations are not used to test its mechanisms.

Table A4 Descriptive Prompt-Use Groups

Cluster	Description	Average Experience *(years)*	Median Years of Experience	Teacher IDs
1	Actions consist only of *Iterate With Me* and *Make For Me*; *Iterate With Me* accounts for at least 42.9% of actions.	14.25	16.5	1, 4, 15, 16
2	Actions are distributed across *Jumpstart For Me*, *Make For Me*, and *Find For Me*.	15.5	13	8, 13, 17, 21, 22, 23
3	*Make For Me* accounts for at least 61.5% of actions.	11.55	11	3, 5, 6, 9, 11, 12, 14, 18, 19, 20, 24
4	*Find For Me* accounts for at least 53.3% of actions, and teachers in this cluster entered relatively few actions.	17.67	22	2, 7, 10

Table A5 Summary Statistics and Exploratory Associations with Teaching Experience

Variable	Mean	Median	SD	Min	Max	Q1	Q3	Correlation with Experience
Conversations	3.24	3	2.036	1	7	1	5	More years of experience, more conversations ($\rho = 0.3589, p = 0.085$)
Prompts	8.375	8	4.362	2	17	5	11	More years of experience, fewer total prompts ($\rho = -0.4631, p = 0.0227$)
Average prompts per conversation	3.559	2.3	2.835	1	11	1.482	4.438	More years of experience, fewer prompts per conversation on average ($p = 0.003581$)
Median prompts per conversation	3.396	2	2.978	1	11	1	3.396	More years of experience, fewer median number of prompts per conversation ($p = 0.002$)

Appendix H: Extensions to the Teacher's Knapsack Model

H.1. AI Exposure and Personal Teacher Use

Generative AI can change an instructional task even when the teacher does not use the technology. Student AI use may reduce the value of an at-home essay, for example, or create additional verification and redesign work. Personal teacher use can then produce a second change by helping the teacher prepare or adapt that task. To distinguish these two channels, we decompose the post-AI task characteristics into environmental and teacher-use effects.

Let $t_i^E > 0$ and v_i^E denote the preparation time and learning value of task i after environmental exposure alone. Define

$$\alpha_i^E := \frac{t_i^E}{t_i} > 0, \qquad k_i^E := v_i^E - v_i \in \mathbb{R},$$

so that

$$t_i^E = \alpha_i^E t_i, \qquad v_i^E = v_i + k_i^E.$$

Environmental effects include changes induced by student AI use, AI-related verification, and the need to redesign assignments, regardless of whether the teacher personally uses AI.

Teacher-use effects are defined relative to the environment-only state. Let $\tilde{t}_i > 0$ and $\tilde{v}_i$ denote the realized characteristics when the teacher also uses AI for task i, and define

$$\alpha_i^T := \frac{\tilde{t}_i}{t_i^E} > 0, \qquad k_i^T := \tilde{v}_i - v_i^E \in \mathbb{R}.$$

Then

$$\tilde{t}_i = \alpha_i^T \alpha_i^E t_i, \qquad \tilde{v}_i = v_i + k_i^E + k_i^T. \tag{9}$$

Accordingly, the combined effects in (2) satisfy $\alpha_i = \alpha_i^E \alpha_i^T$ and $k_i = k_i^E + k_i^T$. Teacher-use effects are measured relative to the environment-only state. They may therefore depend on the environmental effect; the decomposition does not assume that exposure and personal use operate independently.

For a bundle S, define $T^E(S) = \sum_{i \in S} t_i^E$ and $V^E(S) = \sum_{i \in S} v_i^E$.

Proposition 4 already covers the environment-only case. Applied to (t_i^E, v_i^E), it implies that if $V^E(S^L) < L$ and some $S' \subseteq I$ satisfies $V^E(S') \geq L$, every feasible plan must differ from S^L, even if the teacher does not personally use AI. AI adoption and AI exposure are therefore distinct. Declining personal use does not preserve the pre-AI task-selection problem when AI has already changed the instructional environment.

H.2. Task-Specific AI Use under Limited Capacity

Teachers may use AI for some selected tasks but not for others. Verification attention, access, or willingness to delegate may limit how broadly AI can be used even when the technology is available. Let $y_i \in \{0,1\}$ indicate whether AI is used for task i, and let $c_i > 0$ denote the amount of a separately constrained AI-use resource required by that task. Elapsed prompting, checking, and revision time remains part of α_i^T; c_i captures a distinct scarce capacity. A hard ethical, legal, or policy prohibition is represented by the task-specific restriction $y_i = 0$.

Let $C_A \geq 0$ denote available AI-use capacity. Task execution fixes the bundle at S^L and allocates AI use only within that bundle:

$$\begin{aligned} T^{TE}(C_A) = \min_{y} \quad & \sum_{i \in S^L} \left[t_i^E + (\alpha_i^T - 1) t_i^E y_i \right] \\ \text{s.t.} \quad & \sum_{i \in S^L} \left[v_i^E + k_i^T y_i \right] \geq L, \\ & \sum_{i \in S^L} c_i y_i \leq C_A, \\ & y_i \in \{0,1\}, \quad i \in S^L. \end{aligned} \tag{10}$$

Task selection instead jointly chooses the bundle and the allocation of AI use:

$$\begin{aligned} T^{TS}(C_A) = \min_{x,y} \quad & \sum_{i \in I} \left[t_i^E x_i + (\alpha_i^T - 1) t_i^E y_i \right] \\ \text{s.t.} \quad & \sum_{i \in I} \left[v_i^E x_i + k_i^T y_i \right] \geq L, \\ & \sum_{i \in I} c_i y_i \leq C_A, \\ & y_i \leq x_i, \quad i \in I, \\ & x_i, y_i \in \{0,1\}, \quad i \in I. \end{aligned} \tag{11}$$

Equation (10) allocates the available capacity after the legacy bundle has been selected. Equation (11) allows the teacher to reconsider the bundle and the AI allocation together. Their difference is

$$G_A(C_A) = T^{TE}(C_A) - T^{TS}(C_A).$$

This is the time saved per effective use when the task and AI-use decisions are made jointly rather than allocating AI use within the legacy bundle. Across H effective uses, the cumulative saving is $HG_A(C_A)$.

PROPOSITION 5 **(Joint task and AI-use discretion)**. *Suppose* (10) *is feasible. Then*

$$T^{TS}(C_A) \leq T^{TE}(C_A).$$

Equality holds if and only if (11) *has an optimal solution that retains* S^L*. If joint task selection requires the one-time planning cost* κ^{AI}*, it strictly improves on fixed-bundle AI allocation if and only if*

$$HG_A(C_A) > \kappa^{AI}.$$

If (10) *is infeasible but* (11) *is feasible, bundle revision is necessary to meet the learning requirement.*

COROLLARY 4 **(Abundant AI-use capacity)**. *Suppose the fixed-bundle problem is feasible, AI use is permitted for every task,* $C_A \geq \sum_{i \in I} c_i$, $\alpha_i^T \leq 1$, *and* $k_i^T \geq 0$ *for all* $i \in I$. *Then both problems admit an optimal solution in which AI is used for every selected task. The task times and values reduce to the combined characteristics in* (9).

Scarce AI-use capacity changes the ranking problem. An AI-supported task consumes preparation time and capacity, so learning value per minute is no longer sufficient for deciding where AI should be used. In a fractional relaxation with capacity shadow price $\lambda \geq 0$, the time-equivalent resource cost of AI mode $(i,1)$ includes λc_i. A high-density task may therefore be passed over when it consumes capacity that produces a larger benefit elsewhere. Because λ depends on the full allocation problem, task modes generally cannot be ranked by one fixed density.

Appendix I: Proofs for §4

I.1. Proof of Lemma 1

Every bundle $S \subseteq I$ has a unique representation

$$S = (S^L \setminus A) \cup B, \qquad A = S^L \setminus S, \qquad B = S \setminus S^L,$$

with $A \subseteq S^L$ and $B \subseteq I \setminus S^L$. For this representation,

$$\tilde{V}(S) = \tilde{V}(S^L) - \tilde{V}(A) + \tilde{V}(B).$$

Thus, $\tilde{V}(S) \geq L$ if and only if

$$\tilde{V}(B) - \tilde{V}(A) \geq L - \tilde{V}(S^L).$$

Likewise,

$$\tilde{T}(S^L) - \tilde{T}(S) = \tilde{T}(A) - \tilde{T}(B).$$

Maximizing this time difference over all feasible bundles is equivalent to minimizing $\tilde{T}(S)$ over the post-AI feasible set. Therefore, the maximum equals $\tilde{T}(S^L) - \tilde{T}(S^{AI}) = G$. □

I.2. Proof of Corollary 1

Set $B = \varnothing$. Since $\tilde{V}(A) \leq s = \tilde{V}(S^L) - L$,

$$\tilde{V}(S^L \setminus A) = \tilde{V}(S^L) - \tilde{V}(A) \geq L.$$

Thus, removing A is feasible. Lemma 1 gives

$$G \geq \tilde{T}(A) - \tilde{T}(\varnothing) = \tilde{T}(A).$$

Because A is nonempty and every task has positive preparation time, $\tilde{T}(A) > 0$. □

I.3. Proof of Proposition 1

Parts (i) and (ii) follow from the representation in Lemma 1. Feasibility of a removal-and-replacement pair (A, B) does not depend on task times. Increasing the time of a legacy task $i \in S^L$ weakly increases $\tilde{T}(A) - \tilde{T}(B)$ for every pair with $i \in A$ and leaves it unchanged for all other pairs. The maximum is therefore weakly increasing in $\tilde{t}_i$. Increasing the time of an excluded task $j \in I \setminus S^L$ weakly decreases the objective for every pair with $j \in B$ and leaves it unchanged otherwise. The maximum is therefore weakly decreasing in $\tilde{t}_j$.

For part (iii), increasing any $\tilde{v}_i$ weakly expands the set of bundles satisfying $\tilde{V}(S) \geq L$ and does not change the preparation time of any bundle. The optimal post-AI preparation time is therefore weakly decreasing, while $\tilde{T}(S^L)$ is unchanged. Hence $G(I)$ is weakly increasing.

For part (iv), every bundle feasible over I remains available over I'. The minimum feasible preparation time over I' is therefore weakly lower than the minimum over I, while $\tilde{T}(S^L)$ is unchanged. Hence $G(I') \geq G(I)$. □

I.4. Proof of Corollary 2

Under the stated assumptions, the post-AI problem is

$$\min_{S \subseteq I} \alpha T(S) \qquad \text{s.t.} \qquad V(S) \geq L.$$

Multiplication of the objective by the positive constant α does not change the ordering of feasible bundles. Since S^L is optimal under the baseline characteristics over I, it remains optimal after the common scaling. Thus $\tilde{T}(S^{AI}) = \tilde{T}(S^L)$ and $G = 0$. □

I.5. Derivation for Remark 1

Let x be the feasible fractional solution in Remark 1. Reduce x_i by $\ell/\tilde{v}_i$ and increase x_j by $\ell/\tilde{v}_j$. The bound

$$0 < \ell \leq \min\{x_i \tilde{v}_i,\, (1 - x_j)\tilde{v}_j\}$$

ensures that the adjusted fractions remain in $[0, 1]$. The change in total learning value is

$$-\frac{\ell}{\tilde{v}_i}\tilde{v}_i + \frac{\ell}{\tilde{v}_j}\tilde{v}_j = 0.$$

The change in preparation time is

$$-\frac{\ell}{\tilde{v}_i}\tilde{t}_i + \frac{\ell}{\tilde{v}_j}\tilde{t}_j = \ell\left(\frac{1}{\tilde{d}_j} - \frac{1}{\tilde{d}_i}\right) < 0,$$

where the inequality follows from $\tilde{d}_j > \tilde{d}_i > 0$. The corresponding reduction in preparation time is therefore $\ell(1/\tilde{d}_i - 1/\tilde{d}_j) > 0$. □

I.6. Proof of Proposition 2

Fix $q \in \{M, AI\}$. Retaining the legacy bundle for H effective uses requires total preparation time $H\tilde{T}(S^L)$. Task selection produces the common-set optimum S^{AI}, with per-use preparation time $\tilde{T}(S^L) - G$, and incurs the one-time cost κ^q. Its total time-equivalent cost is

$$\kappa^q + H\left[\tilde{T}(S^L) - G\right].$$

Task selection strictly improves on retaining the legacy bundle if and only if

$$\kappa^q + H\left[\tilde{T}(S^L) - G\right] < H\tilde{T}(S^L),$$

which is equivalent to $HG > \kappa^q$, or $N^q > 0$. Retaining S^L has net saving zero. The three alternatives therefore have net savings 0, N^M, and N^{AI}, and the optimal net saving is their maximum. Task selection is strictly preferred to retention exactly when $\max\{N^M, N^{AI}\} > 0$. If this maximum is positive, every method attaining it minimizes total time-equivalent cost; if it equals zero, retention and each task-selection method attaining zero have the same cost. □

I.7. Proof of Corollary 3

Because both methods search over I, they have the common frictionless gap G. By Proposition 2, manual task selection is strictly preferable to retention if and only if $HG > \kappa^M$, whereas AI-assisted task selection is strictly preferable if and only if $HG > \kappa^{AI}$. When $\kappa^{AI} < \kappa^M$, AI-assisted planning makes task selection strictly preferable while manual planning does not if and only if

$$HG > \kappa^{AI} \qquad \text{and} \qquad HG \le \kappa^M.$$

Combining the inequalities yields $\kappa^{AI} < HG \le \kappa^M$. □

I.8. Proof of Proposition 3

For part (i), Proposition 1 shows that G is weakly increasing in the time of each legacy task and weakly decreasing in the time of each excluded task. Because $H > 0$ and κ^q is fixed, $N^q = HG - \kappa^q$ has the same coordinatewise monotonicity.

For part (ii), let $\bar{t}_i$ denote task times at $\mu = 1$ and let $\tilde{T}_1(S) = \sum_{i \in S} \bar{t}_i$. Multiplying all task times in I by $\mu > 0$ multiplies the preparation time of every bundle by μ without changing the feasible set or the ordering of bundle times. Therefore,

$$\begin{aligned} G(\mu) &= \mu\tilde{T}_1(S^L) - \min_{S \subseteq I:\, \tilde{V}(S) \ge L} \mu\tilde{T}_1(S) \\ &= \mu\left[\tilde{T}_1(S^L) - \min_{S \subseteq I:\, \tilde{V}(S) \ge L} \tilde{T}_1(S)\right] = \mu G(1). \end{aligned}$$

Substitution into (8) gives $N^q(\mu) = H\mu G(1) - \kappa^q$.

For part (iii), preparation times are nonnegative, so

$$G = \tilde{T}(S^L) - \min_{S \subseteq I:\, \tilde{V}(S) \ge L} \tilde{T}(S) \le \tilde{T}(S^L).$$

If $H\tilde{T}(S^L) \le \kappa^q$, then $HG \le \kappa^q$, so Proposition 2 implies that method q is not strictly preferred to retention. If the bound holds for $\min\{\kappa^M, \kappa^{AI}\}$, it holds for both methods. □

I.9. Proof of Proposition 4

Feasibility before AI implies $s_0 = V(S^L) - L \ge 0$. Summing (2) over S^L gives

$$\tilde{V}(S^L) = V(S^L) + \sum_{i \in S^L} k_i.$$

Therefore,

$$\tilde{V}(S^L) \ge L \iff \sum_{i \in S^L} k_i \ge L - V(S^L) = -s_0.$$

If the condition fails, S^L violates the learning constraint in (3). If another bundle S' is feasible, every feasible solution differs from S^L. Because preparation times do not enter the learning constraint, changing only $\tilde{t}_i$ for tasks in S^L cannot restore its feasibility. □

I.10. Proof of Proposition 5

Let y^{TE} be an optimal solution to (10). Set $x_i = x_i^L$ and $y_i = y_i^{TE}$ for $i \in S^L$, and set $x_i = y_i = 0$ for $i \in I \setminus S^L$. This solution satisfies the learning, capacity, coupling, and binary constraints of (11). The joint problem therefore optimizes over a feasible set containing the fixed-bundle solution, so $T^{TS}(C_A) \le T^{TE}(C_A)$.

If the joint problem has an optimal solution retaining S^L, its AI-use vector is feasible for the fixed-bundle problem, so $T^{TE}(C_A) \le T^{TS}(C_A)$. Together with the reverse inequality, equality follows. Conversely, if equality holds, the embedded optimal fixed-bundle solution is also optimal for the joint problem and retains S^L.

Across H effective uses, fixed-bundle AI allocation requires $HT^{TE}(C_A)$ units of time. Joint task and AI-use selection requires $HT^{TS}(C_A) + \kappa^{AI}$. The joint problem strictly improves on the fixed-bundle problem if and only if

$$HT^{TS}(C_A) + \kappa^{AI} < HT^{TE}(C_A),$$

which is equivalent to $HG_A(C_A) > \kappa^{AI}$. If the fixed-bundle problem is infeasible and the joint problem is feasible, retaining S^L cannot meet the learning requirement, so bundle revision is necessary. □

I.11. Proof of Corollary 4

The fixed-bundle problem is feasible by assumption, and any of its feasible solutions can be embedded in the joint problem, so both problems are feasible. Consider any feasible solution to either problem. Set AI use to one for every selected task. The capacity constraint remains satisfied because $C_A \ge \sum_{i \in I} c_i$. Because $\alpha_i^T \le 1$, this change weakly reduces preparation time. Because $k_i^T \ge 0$, it weakly increases learning value. Hence each problem admits an optimal solution using AI for every selected task. Under such a solution, selected task i has time $\alpha_i^T t_i^E = \alpha_i^T \alpha_i^E t_i$ and value $v_i^E + k_i^T = v_i + k_i^E + k_i^T$, which are the combined characteristics in (9). □

Appendix J: Prompt Traces

We use prompt connectedness and output dispersion in §5 to describe how teachers interacted with AI. This appendix explains how we construct the measures and how they relate to the model. Neither measure identifies the chosen task bundle, estimates the model primitives, or establishes that a teacher used AI optimally.

J.1. Prompt Connectedness

Prompt connectedness captures whether a teacher's prompts remain organized around a common learning objective. We interpret it as one feature of the interaction, rather than as a measure of the teacher's task bundle or the optimality of AI use.

We construct the empirical connectedness measure from the prompts teachers entered during the open-ended observation. For each teacher conversation with at least two prompts, we embed each prompt using *SBERT* (Sentence Bidirectional Encoder Representations from Transformers) (Reimers and Gurevych 2019), implemented with the pre-trained `all-MiniLM-L6-v2` model from the `SentenceTransformer` library. We then compute pairwise cosine similarity scores across prompts within the same teacher conversation and average those scores. Higher values indicate that the prompt sequence remains more semantically centered.

Because connectedness is mechanically harder to assess in short conversations, we interpret scores based on four or fewer prompts with caution.

Connectedness is consistent with prompts remaining organized around a common objective, as may occur during structured task selection. It is not unique to that process. A teacher can also generate a connected sequence while repeatedly refining a worksheet, quiz, or other task that has already been selected. We therefore interpret connectedness alongside prompt content and use trajectories. It is a descriptive trace of how the interaction unfolds, not a measure of productivity, bundle optimality, or the task-selection gain G.

J.2. Output Dispersion

The output-dispersion analysis in §5.2 captures how broadly a prompt opens the space of candidate responses. We use repeated model runs to characterize this property of prompts, not to recover the exact outputs teachers saw during the study or to define a universal measure of prompt quality.

For each of the 201 teacher prompts from the observation period, we generated 10 independent outputs using the GPT-4 model available through ChatGPT at the time of this analysis in fall 2024. Each output was generated in a fresh session so that previous responses did not provide conversational context. We embedded the 10 resulting outputs using SBERT and computed all pairwise cosine similarities among outputs generated from the same prompt. Higher average similarity indicates that repeated runs produce more semantically concentrated outputs; lower average similarity indicates greater output dispersion. As a complementary measure, we calculated the variance of the output embeddings across dimensions and averaged across dimensions to obtain a single dispersion measure for each prompt.

Dispersion is not inherently beneficial. As stated in Remark 2, moving from dispersion level q to q' is beneficial when the added reuse-adjusted task-selection gain exceeds the added evaluation and adaptation cost:

$$H\left[G(q') - G(q)\right] > \kappa^{AI}(q') - \kappa^{AI}(q).$$

Thus, output dispersion is aligned with productivity only when it changes the economics of search in the right direction. Dispersion can be valuable in task selection when it surfaces higher-value candidate tasks. It can be costly when it expands the set of outputs without providing enough structure to evaluate them.

J.3. Goal Clarity and Search

Let θ denote goal clarity. A clearer objective may reduce $\kappa^{AI}(\theta)$ by giving the teacher a sharper criterion for evaluating candidate tasks. If clarity also changes which candidates are found or how their task characteristics are assessed, write the resulting task-selection gain as $G(\theta)$. Goal clarity alone does not change the frictionless gain when the candidate set and post-AI task characteristics are fixed.

Corollary 5 **(Goal clarity expands the adoption region)**. *Suppose $G(\theta)$ is weakly increasing in θ and $\kappa^{AI}(\theta)$ is weakly decreasing in θ. Then the net value of AI-assisted task selection, $HG(\theta) - \kappa^{AI}(\theta)$, is weakly increasing in θ. If $G(\theta) > 0$, the threshold effective reuse count $H^*(\theta) = \kappa^{AI}(\theta)/G(\theta)$ is weakly decreasing in θ.*

Proof. The net value is $N^{AI}(\theta) = HG(\theta) - \kappa^{AI}(\theta)$. Because $H > 0$, a weak increase in $G(\theta)$ and a weak decrease in $\kappa^{AI}(\theta)$ make $N^{AI}(\theta)$ weakly increasing in θ. When $G(\theta) > 0$, the threshold is $H^*(\theta) = \kappa^{AI}(\theta)/G(\theta)$. An increase in θ weakly decreases the nonnegative numerator and weakly increases the positive denominator, so $H^*(\theta)$ is weakly decreasing. □

Goal clarity matters because it changes the economics of search. Prompt connectedness is best interpreted as a possible trace of a goal-directed interaction, not as a direct measure of productivity or optimality. As discussed in §J.1, connected prompts may also arise when a teacher repeatedly refines a task that has already been selected.

Remark 2 (Objective-aligned output dispersion). Let q index the dispersion of outputs generated by a prompt or prompting strategy. Moving from dispersion level q to q' improves the net value of AI-assisted task selection exactly when

$$H\left[G(q') - G(q)\right] > \kappa^{AI}(q') - \kappa^{AI}(q).$$

Thus, dispersion is useful only when the additional value of the alternatives it surfaces exceeds the additional cost of evaluating and adapting those alternatives.

The implication is alignment. For task execution, concentrated outputs may be valuable because they reduce verification and post-processing. For task selection, dispersed outputs may be valuable because they reveal alternatives, but only when those alternatives are worth evaluating.

J.4. Robustness to a Value-Maximization Objective

The inverse-knapsack formulation in §2.4 and §4 fits our setting because teachers face learning requirements and seek to limit preparation time. The execution-selection distinction does not depend on this orientation of the objective. To see this, suppose instead that the teacher has a preparation-time budget B and chooses a bundle to maximize learning value. Let the post-AI candidate set I contain the baseline set I^0. The baseline problem is

$$S_B^L \in \arg\max_{S \subseteq I^0} \{V(S) : T(S) \leq B\},$$

and the corresponding post-AI task-selection problem is

$$S_B^{AI} \in \arg\max_{S \subseteq I} \left\{\tilde{V}(S) : \tilde{T}(S) \leq B\right\}.$$

Task execution retains S_B^L and evaluates it using the post-AI task characteristics. If $\tilde{T}(S_B^L) \leq B$, the legacy bundle remains available to the post-AI task-selection problem, so optimality implies $\tilde{V}(S_B^{AI}) \geq \tilde{V}(S_B^L)$. If $\tilde{T}(S_B^L) > B$ but another post-AI bundle satisfies the budget, retaining the legacy bundle is infeasible and bundle revision is necessary. Thus, the distinction between executing a legacy bundle and reopening the bundle is not specific to minimizing preparation time subject to a learning requirement. A parallel adoption threshold would require an explicit utility or shadow-price conversion because planning time and learning value are not directly commensurable. That additional structure is not needed for this robustness comparison.