A07 | PRIMARY-STUDY PROTOCOL AND REPRODUCIBLE METHODS RESEARCH

Who Receives the Automation Dividend? A Matched-Pair School Trial of Protected Time, Teacher Workload and Verified Quality

Dr. Freedom Cheteni | Human author

ORCID 0009-0006-3507-0094

Prepared 2026-10-15 | Edition 1.0 | 4,421 article words

Primary-study protocol and reproducible methods; separate human editorial record.

Keywords: Movement Thinking, teacher time use, protected planning, school governance, American Time Use Survey 2025, matched-pair cluster trial, AI verification burden

Read online / article companion / citation exports

Contribution

An original protected-time allocation experiment after common AI support, an executed ATUS 2025 source audit, and reproducible design sensitivity with an explicit pre-enrollment activation gate.

Prospective primary study specification plus completed secondary-file audit and artificial-data planning simulations; no local participant outcomes.

Abstract

An AI-assisted draft produced quickly does not establish that a school has reduced work. Verification, adaptation and coordination can consume the apparent saving, while new output expectations can absorb any remainder. This article specifies an original prospective primary study of the institutional allocation of that remainder: a matched-pair school trial in which both conditions receive common AI workflow support, but only one receives an enforceable protected-time charter. The primary outcome is change in teachers' weekday job-related minutes outside an agreed paid timetable; independently blinded quality assessments provide a separate guardrail. The planning design contains 48 schools in 24 district pairs and approximately 576 consenting teachers followed for 12 weeks. No schools or teachers have been recruited for this protocol. Two methodological investigations are complete. A reproducible audit of official American Time Use Survey 2025 files checks 6,146 respondent records and 114,151 activity episodes, finds complete 1,440-minute diary closure and identifies 120 respondents under a narrowly declared teacher occupation definition. A nine-scenario artificial-data analysis shows that 48-school power for a hypothetical 20-minute reduction ranges from 35.6% to 87.7% under different assumed variability. Neither investigation estimates an AI or timetable intervention effect. The contribution is a falsifiable evaluation architecture that separates task acceleration, net human work, the distribution of protected time and acceptable educational quality, with explicit limits on surveillance, national benchmarking and causal interpretation.

1. From an efficient task to a defensible institutional gain

Schools are being offered systems that compose communications, generate lesson materials and summarize administrative information. The visible transaction is often the first output: a plausible document appears sooner. The educational transaction is longer. An educator must decide whether the document is accurate, suitable for the intended learners, accessible, consistent with local obligations and worth using. Someone must correct errors, communicate changes and assume responsibility for consequences. A measure confined to the first transaction cannot establish a reduction in the second.

The preceding conceptual article in this volume proposed that an automation dividend belongs in the timetable. This companion develops a different scholarly contribution: an executable public-data audit and a prospective allocation experiment. Movement Thinking supplies the orientation that educational methods should evolve through inclusive, revisable practice [1]. Here that orientation becomes an institutional proposition: a credible efficiency gain should expand the practical space for educational judgment rather

than automatically increase output quotas. Whether a specific charter accomplishes that remains an empirical question.

The object of study is therefore a governance decision following common technical support. The trial does not compare an enthusiastic AI school with a school denied access to all assistance. It asks whether a precommitted allocation rule changes working conditions when the technical opportunity, training and routine support are comparable. This distinction makes the proposed causal claim narrower and more useful to a superintendent deciding how efficiency gains should be distributed.

2. What existing productivity evidence permits

Noy and Zhang studied generative assistance in bounded professional writing tasks [9]. Their experiment supplies a reason to investigate changes in task completion and quality together. The tasks, participants and observation window do not establish a sustainable reduction in a teacher's working week. The publication is cited under its original authors; the accessible MIT manuscript is identified separately in the source audit rather than described as an independently replicated school trial.

Brynjolfsson, Li and Raymond examined generative assistance in customer support [10]. That organizational setting is relevant because task support operates inside an existing workflow. Nevertheless, customer-support performance is not a direct measure of curriculum judgment, teacher workload or protected planning. Sector differences, managerial expectations and the distribution of correction work are plausible effect modifiers. The published article's identity and publisher information were verified; the accessible NBER working-paper record is not silently treated as the identical final version.

Maas and colleagues' study of perceived time pressure, exhaustion and principal support anchors the education-specific concern [11]. It does not establish an AI effect. Together these sources motivate a question rather than answer it: can a school convert technical support into a measurable improvement in work allocation without degrading the quality for which human educators remain responsible? Evidence from other sectors makes that question worth testing, not already settled.

3. The dividend as a three-part accounting problem

The proposed theory separates gross task acceleration, net work change and institutional allocation. Gross acceleration is a difference in the time needed for an initial output. Net work change includes checking, repair, adaptation, training, coordination and displaced responsibilities. Allocation describes who receives the resulting time and what obligations are removed or added. These components can move in opposite directions. A faster draft and a longer working day are logically compatible.

The protocol does not combine these components into a single proprietary productivity score. Each has a distinct indicator and possible counterexample. Self-reported tool use is an exposure description, not evidence of benefit. A published timetable is evidence of an allocation commitment, not proof that a block occurred. An uninterrupted block is an implementation result, not proof of higher educational quality. A time reduction accompanied by serious errors fails the proposed institutional test.

Time also has structure. Sixty fragmented minutes between alerts may differ from one protected hour. The intervention consequently specifies a continuous block, a removal or consolidation of existing obligations and a route for challenging compensating demands. It leaves educators discretion over the learning work pursued within that block. The study tests this bounded arrangement rather than the general desirability of autonomy or technology.

4. Why ATUS is the external database

The American Time Use Survey provides a documented national architecture for reconstructing a day, organizing activity episodes and distinguishing people from episodes [2] [8]. It is useful here as a secondary measurement and feasibility resource. It is not a database of school AI interventions, and no national respondent is treated as a local participant. Official single-year 2025 respondent and activity ZIPs were downloaded, hashed and analyzed; the release page records June 25, 2026.

The data dictionary identifies TUCASEID, TUACTIVITY_N, TUACTDUR24 and TRCODE for episode reconstruction [3]. The selected occupation subset uses the official 2018 Census classification [5], while the 2025 lexicon identifies the main-job work code [6]. These documented fields make a limited audit reproducible. They cannot reveal whether a teacher used a particular AI service or whether a workplace honored a protected-time promise.

The September 2026 User’s Guide documents the survey design, weighting and single-diary structure [4]. ATUS also illustrates a design choice the local study must

improve upon. One recalled diary can describe a sampled day but cannot establish an individual's stable weekly workload. The prospective study therefore samples repeated, assigned teaching weekdays, preserves missing scheduled days in its target and adds an explicit paid-timetable map. This is a new measurement adaptation requiring pilot work; it is not an assertion that the BLS has validated the local instrument.

5. Executed source audit and occupation boundaries

The executed audit reads 6,146 respondent records with unique TUCASEID values and 114,151 activity episodes with unique respondent-activity keys. Every respondent has activity records, no episode is orphaned, and summing truncated durations yields exactly 1,440 minutes for every diary. These are actual file-integrity findings, not hypothetical counts. ZIP and extracted-data hashes, retrieval URLs and the checking script accompany the article.

A declared teacher subset requires employed status TELFS equal to 1 or 2 and TEIO1OCD equal to 2310, 2320 or 2330. The resulting counts are 85 elementary/middle-school teachers, 24 secondary-school teachers and 11 special-education teachers: 120 respondents. Code 2300 combines preschool and kindergarten and is excluded from this particular definition rather than presented as a clean kindergarten group. Postsecondary teachers, assistants and miscellaneous instruction occupations are also excluded. The resulting subset should not be labeled all education workers or a comprehensive census of K-12 teachers.

Of these 120 diaries, 53 concern weekdays and 67 weekends. Only 62 report positive main-job work; among weekday diaries, 45 do. These modest denominators constrain any benchmarking ambition. The counts justify designing a richer local measurement exercise. They do not justify fine-grained national comparisons by grade, district, teacher identity or AI exposure. Public identifiers provide no authorized link to the future school trial.

6. Executed descriptive estimates and the missing autumn

Using TUFINLWGT in a ratio mean, retaining zero main-job work on nonworking diary days, the selected group averages 294.9 minutes of main-job work per sampled diary day. The weekday-only ratio is 416.0 minutes. Restricting instead to positive-work diaries yields 437.6 and 480.0 minutes, respectively. These calculations answer different denominator questions. A nonworking day is not evidence of low professional commitment, and an all-day average must not be multiplied into a supposed contractual working week.

No replicate-weight standard errors, significance tests or national teacher subgroup rankings were calculated. The estimates are descriptive researcher computations, not an official BLS teacher table. Their small domain counts, occupation exclusions and diary design travel with the numbers. The tools deliberately report no weekend mean or comparative inference; weekend records remain in the sample accounting. ATUS main-job work cannot separate teaching, administration, checking, AI generation or work outside an agreed paid timetable.

A further boundary is substantive rather than merely computational. The 2025 federal shutdown suspended operations, leaving no diaries for September 30 through November 11; the audit finds zero records in that interval [7]. Reweighting does not reveal how teachers actually used time on those unobserved dates. An autumn school-work analysis would be especially vulnerable to assuming that remaining fourth-quarter diaries reproduce the missing season. The present findings are therefore a versioned feasibility demonstration, not a recovered autumn workload history.

7. Primary research questions and estimands

The primary question is whether a protected-time charter reduces job-related work outside an agreed paid timetable relative to ordinary allocation discretion after common AI support. The hypothesis is a reduction, tested two-sided to retain the possibility of harm. The candidate practical contrast of 20 minutes per teaching weekday is used only in design sensitivity. It is not an ATUS-derived estimate, predicted saving or promise to participating teachers.

Let the baseline consenting teacher cohort in school s define its mean weekday outcome under assignment z. Let Delta_s(z) be that school's endline minus baseline mean. The principal estimand is the equal-school average of Delta_s(1) minus Delta_s(0) across recruited schools. Negative values favor the charter. School equality makes a small participating school count as a school rather than allowing larger cohorts to dominate an organizational question. An individual-weighted contrast is secondary and explicitly answers a different question.

The quality question is separate: can the time contrast coexist with a defensible bound on deterioration in independently rated outputs? A non-significant quality difference is not evidence of equivalence. The protocol proposes a noninferiority guardrail, subject to external margin justification before allocation. The sustainable-dividend conclusion is conjunctive: favorable time evidence, acceptable quality evidence and no unresolved material safety problem. Every component is reported even when the conjunction fails.

8. Prospective recruitment and allocation

The candidate design contains 24 district pairs, with two eligible schools per district and approximately twelve consenting teachers per school: 48 schools and 576 planned teachers. These are budgeting and sensitivity assumptions. There are zero recruited participants in this package. School agreements, protected paid research time, an independent ethics determination, an instrument pilot, a finalized analysis implementation and public registration must precede activation. Changed feasibility requires a dated amendment before allocation.

Eligible schools must be able to run both conditions within existing paid schedules, support the common approved workflow and protect individual diary confidentiality from supervisors. Teacher invitation, consent and baseline measurement occur before condition disclosure. Recruitment must not select only staff already enthusiastic about AI. A private screening log records eligibility, invitation and refusal. Part-time educators and those with substantial caring obligations remain eligible; their timetable maps and access needs are documented.

Within each district pair, an independent custodian randomly assigns one school to the charter and one to ordinary discretion. The sequence is concealed until rosters and baseline measures are locked. Its generation method is public, while the seed remains access controlled until allocation cannot be manipulated. Cluster reporting follows the distinction between schools and people emphasized by CONSORT guidance [12]. Postallocation recruitment is reported separately, and a school that withdraws is not quietly replaced.

9. Common support and the active comparator

Both conditions receive access to the same approved AI workflow support, an initial two-hour orientation and a 30-minute paid weekly support clinic. Training covers source checking, accessibility, privacy boundaries and a workable non-AI alternative. The common task families are low-risk preparation and administrative drafting. Individual grading, disability eligibility, discipline and safeguarding decisions remain outside the trial's automation scope. Neither tool availability nor study participation authorizes uploading identifiable student material.

The comparator retains ordinary school discretion over how any task gain is used. It receives equal scheduled support, outcome measurement and routes to report problems. Existing protected planning cannot be removed to make the comparison artificially weak. New naturally occurring allocation agreements are documented as overlap, not prohibited

or punished. The estimand is consequently the effect of offering a formal charter against real ordinary discretion, including imperfect separation between conditions.

The common platform and clinic reduce a simple attention explanation, but cannot eliminate differences in expectations or leadership commitment. Participants and leaders cannot be blinded to timetable changes. Fidelity interviews and records of changed demands help interpret the contrast without converting an observational mechanism measure into a randomized mediator. The study does not estimate the independent benefit of AI access because both arms receive it.

10. The protected-time charter

Treatment schools identify one continuous 60-minute paid block per week for teacher-selected learning design, feedback or collaborative inquiry. They list the routine obligation that is ceased, consolidated or reassigned within a funded plan. The block is a reallocation of existing work, not an additional meeting or an expectation of unpaid evening completion. Schools record intended and delivered minutes, interruptions, cancellations and any compensating output demands.

The charter names a human quality owner, a teacher representative and an independent appeal contact. It prohibits increasing required output solely because a draft can be generated faster. If verification consumes the anticipated saving, the record shows no realized net gain; leaders cannot manufacture a dividend by excluding correction labor. A teacher may use the protected block without using AI on every task. The assigned policy, rather than compulsory tool compliance, defines treatment.

The 60-minute commitment and hypothetical 20-minute daily contrast are not an accounting identity. Work can be reorganized across days, and checking or coordination can either expand or contract. The analysis therefore reports delivered blocks, total work, outside-timetable work and quality separately. A positive result must identify who received the time, who absorbed displaced tasks and whether the arrangement was sustained rather than infer all three from a calendar entry.

11. A diary that supports research without surveillance

A coordinator assigns two baseline and three endline teaching weekdays, balancing weekdays across the cohort and avoiding scheduled closures. Participants reconstruct the previous 04:00-04:00 day using a brief episode interface. A job-related episode receives an agreed-timetable flag and a task-family tag, including checking or repair. Primary activity durations close to 1,440 minutes; secondary responsibility tags are not added again. Unknown periods and incomplete days remain explicitly unknown.

The primary outcome sums job-related primary episodes outside the paid-timetable map, bounded between zero and 1,440 minutes. This operational category is not a legal determination that salaried work is unpaid. A teacher's map is agreed before baseline and cannot be revised after the outcome is known to make an apparent improvement. Research diary completion itself is separately recorded and supported with paid time so that measurement does not conceal the burden it introduces.

There are no keystroke logs, screen captures, message-content collections or continuous location feeds. Participants may correct a reconstruction, mark uncertainty or decline a day without employment consequences. Supervisors receive only disclosure-controlled school summaries. These protections are part of the intervention's credibility: a study that induces staff to underreport work to satisfy managers cannot provide a trustworthy institutional estimate, even if its timestamp system appears precise.

12. Independent quality measurement and stopping rules

Two parallel synthetic school-task packets are piloted for difficulty and accessibility before the main trial. Baseline and endline forms are counterbalanced. Tasks require a defensible communication, feedback plan or administrative explanation under explicit constraints, without real student records. Two human raters blind to school, condition and occasion score every artifact on factual fidelity, pedagogical appropriateness, accessibility, actionable reasoning and verification, each from zero to four. The primary quality score is the average of their 20-point totals.

This rubric is original and unvalidated. Pilot agreement, ceiling effects, form comparability and stakeholder interpretation must be established before choosing the final noninferiority margin. The candidate margin is a 0.5-point deterioration, evaluated using a lower one-sided 95% confidence limit. It cannot be defended merely because it produces convenient power. The trial's primary time test and this guardrail need separate precision assessment; the executed planning simulation covers only time.

Independent safety flags include invented sources, private-data exposure, discriminatory instruction and materially misleading safeguarding guidance. A third rater may adjudicate a safety concern but does not selectively overwrite the primary mean score. Credible material harm can pause the workflow immediately under the safety lead's documented authority. NIST's voluntary risk framework informs this governance architecture [16]; citation is not certification that a school or tool meets a federal standard.

13. Primary estimation, missingness and interpretation

The primary school outcome is its baseline cohort's mean change over assigned weekdays. Pair differences subtract the control-school change from the treatment-school change; the overall estimate averages the 24 pair differences. The default standard error is their sample standard deviation divided by the square root of the number of pairs, with a t interval using pair count minus one degrees of freedom. Equal weighting and the complete set of recruited schools remain fixed after assignment.

Missing assigned days are not dropped as though voluntary completion were random. The detailed analysis specification requires a validated multilevel multiple-imputation implementation before activation, fifty imputations of bounded daily endpoints, baseline and schedule predictors, and school dependence. Pooled estimates use Rubin's rules with a finite-sample adjustment [17]. No imputation has been run for this article. Its assumptions must be challenged through unfavorable shifts in missing outcomes and worst-case bounds using the outcome's known limits.

A complete-outcome pairwise randomization check addresses the Fisher sharp null, separately from inference about the mean assignment effect. Protocol deviations, staff departures, tool discontinuation and whole-school loss are reported rather than treated as grounds for exclusion. An incomplete or imprecise study can remain inconclusive. A superintendent should not interpret a large p value, a favorable point estimate or a successful process interview as interchangeable evidence of a sustainable dividend.

14. Executed design sensitivity

A reproducible artificial-data calculation generated 10,000 replicates for each combination of 16, 24 or 32 school pairs and an assumed standard deviation of pair differences of 30, 45 or 60 minutes. Independent Gaussian pair differences were generated under a zero-effect null and a hypothetical 20-minute reduction. Every replicate used the stated two-sided paired t test at alpha 0.05. The seed, software versions, scenario outputs and Monte Carlo standard errors are supplied.

For 24 pairs, simulated power is 0.8771, 0.5488 and 0.3561 as assumed variability increases from 30 to 60 minutes. Corresponding Monte Carlo standard errors are approximately 0.0033, 0.0050 and 0.0048. Across the nine scenarios, null rejection rates range from 0.0487 to 0.0550. These checks describe computation under the generating model, not empirical calibration of a school intervention. Approximate 80% minimum detectable effects for 24 pairs are 17.8, 26.7 and 35.6 minutes.

The findings materially constrain planning: 48 schools cannot be described as reliably sufficient without information about aggregate variability and the quality guardrail. ATUS does not provide repeated local school outcomes from which to estimate that parameter. A separate non-efficacy pilot must inform feasibility, missingness, clustering and instrument performance. Rutterford and colleagues explain why cluster designs require explicit sizing assumptions [14]. The present simulation exposes those assumptions rather than borrowing an impressive sample size from an unrelated survey.

14a. A pre-enrollment design activation gate

The 48-school configuration is a provisional planning scenario, not an approved adequately powered confirmatory sample. A separate measurement-only pilot in at least twelve school pairs, excluded from the main cohort, must estimate the variability of paired change scores for time and quality, diary completion and form reliability. The pilot introduces no experimental charter effect to be estimated. An independent statistician uses the upper 80% confidence bound for each paired-score standard deviation, rather than the most optimistic estimate, as the design nuisance parameter.

Before recruiting the main efficacy cohort, candidate designs from 24 to 120 district pairs are evaluated using the finalized estimator and missingness procedure. Allow for loss of 10% of recruited pairs and any greater loss suggested by the pilot. Select the smallest affordable design whose 50,000-replicate planning runs yield a lower 95% Monte Carlo confidence bound of at least 90% power both for a hypothetical 20-minute reduction and for the quality guardrail under zero true deterioration. These are declared policy/design assumptions, not known effects. Meeting both marginal power targets provides an at-least-80% joint-success lower bound by the union bound under the modeled alternative; it does not guarantee real-world power when assumptions fail.

The quality margin requires independent educator justification and successful form calibration before this gate can be evaluated. If no design within the funded cap meets the gate, the confirmatory trial does not launch. It may be redesigned and preregistered, or conducted explicitly as a feasibility study without a confirmatory benefit claim. Final numbers, justified margins, code, attrition assumptions and amendments are frozen in

registration before main-cohort recruitment and assignment. The included nine-scenario calculations demonstrate why this gate is necessary; they have not executed or passed it.

15. Falsification, distribution and rival explanations

The theory fails its institutional purpose if protected blocks appear on calendars while outside-timetable work, checking burden or output expectations rise. It also fails a sustainable-dividend claim if time improves but independently assessed quality crosses a justified harm margin. Increased AI use, more generated documents or satisfied leadership alone cannot rescue those counterexamples. Conversely, a null time contrast with wide uncertainty is inconclusive about the broader theory; the protocol must distinguish lack of precision from evidence of a practically negligible effect.

Displacement is a central rival explanation. The teacher benefit may be purchased by added work for assistants, principals or technical staff. Process records therefore identify reassigned obligations and obtain voluntary role-level accounts without exposing individual staff performance. Another rival is reporting pressure: treatment teachers may feel expected to demonstrate a benefit. Confidential collection, independent raters and explicit permission to report no gain reduce this threat but cannot remove it completely.

Exploratory distributional analyses examine contracted fraction, caregiving responsibilities and baseline time burden with consent and disclosure safeguards. Their interaction intervals are reported without ranking small demographic groups. A mean improvement can conceal worsening for those who already have the least discretionary time. The proposed interpretation consequently combines the assignment contrast with a distributional account and negative cases, while reserving confirmatory status for the declared primary endpoint.

Method and data statement

The completed empirical components are a secondary-file integrity and descriptive audit of pinned ATUS 2025 files and synthetic matched-pair planning simulations. Primary participant outcomes, recruited sample counts, registration and ethics approval do not exist in this package. The downloadable Python program executes those two completed components; it does not implement or claim to have fitted the future imputation model, collected diaries, randomized actual schools or evaluated the primary quality endpoint.

The public materials include the source manifest, exact hashes, result JSON, instruments, study specification, statistical plan, requirements and tests. Original respondent files remain at the BLS source rather than being repackaged into the journal. No local personal

records are included. A future registered protocol must identify sponsor responsibilities, compensation, data-controller arrangements, access limits, retention and independent analysis roles. SPIRIT provides a reporting scaffold [13], not an assertion that this school study is a registered clinical trial.

The ATUS source is contextual and methodological. No respondent is reidentified, contacted or linked to a school, ORCID record or journal account. A paper agent may retrieve these statements and source citations, but its response is not a new analysis. Reproduction requires the pinned files and reported software; a changed source hash triggers explicit refusal rather than silently accepting a different release as the same evidence.

17. Limitations

The national teacher domain is small and excludes occupations that cannot be separated cleanly. A single day's main-job category lacks the local constructs needed to estimate net automation effects. The 2025 shutdown creates a specific seasonal observation gap, while descriptive ratio means lack replicate-weight uncertainty in this edition. Nothing in these audit results establishes national after-hours work or a baseline AI effect.

The proposed trial also faces unblinded behavior, diary recall error, outcome missingness, school contamination and limited transportability from consenting districts. Synthetic quality tasks make independent blinding more feasible but may not capture all consequential classroom decisions. The charter combines a block, a demand-removal agreement and an appeal route; assignment identifies the package, not each component. Twelve weeks cannot establish long-term retention, labor-market preparation or institutional transformation.

The simulations omit missingness, unequal size, interference and joint quality power. Normal pair differences and assumed variability may be unrealistic. Before recruitment, the local pilot and external review may require a larger design, different task packets or revised margins. Publishing the protocol makes these dependencies inspectable; it does not solve them by declaration. These limits remain part of the scholarly contribution because they specify where a future claim would exceed the evidence.

AI, authorship and editorial disclosure

The named human author is Dr. Freedom Cheteni. This is a new primary-study protocol corresponding to conceptual article A02, not a revised version that inherits that article's review. No human peer review, acceptance, registration, institutional collaboration or participant consent is asserted for this new text. The publisher is The Research University, and the venue is the Journal for School Superintendents. Citations retain original authors, dataset agencies and source versions.

OpenAI's ChatGPT supported literature discovery, drafting, code preparation and computational checking [15]. No specific model version is asserted. AI is neither an author nor a human peer reviewer. The human author and independent editors remain responsible for approving claims, executing the proposed study and documenting review. Source-derived reasoning, actual audit results and artificial simulations are labeled separately, allowing subsequent readers to challenge each on its own terms.

19. Implications for superintendent decisions

The study turns a familiar technology question into a concrete governance test. Instead of asking only whether staff can produce drafts sooner, it asks whether a school can defend the allocation of the resulting human effort. That requires naming the tasks removed, recognizing verification labor, protecting discretion and permitting an unfavorable finding. The practical output would be an estimate with uncertainty, a quality boundary and a visible account of displaced work.

If a future study supports a sustainable dividend, replication should vary school contexts, task families and leadership constraints before broad scaling. If it does not, the record should distinguish inadequate precision, failed implementation, shifted work and actual quality harm. Each implies a different next decision. Movement Thinking is operational here when a school can revise its own arrangements in response to such evidence, including the decision to stop a technically impressive system that does not improve the conditions of educational work.

Research materials and reproducibility

Inspect the prospective study specification, original draft instruments, executed public-file audit, design simulations and documented tools. No local participant trial outcomes are reported.

Complete research package · ZIP

Research guide

Study specification

Source and variable audit

Diary, charter and quality instruments

Statistical analysis specification

Executable source-audit and planning tools

Executed ATUS source audit

Artificial-data design sensitivity

Methods tests

Read the corresponding conceptual foundation

Evidence and limits

Claim	Evidence	Limit
Task acceleration does not establish net institutional time release.	Distinct production, checking, total-work and allocation measures in an active-comparator protocol. [1, 9, 10, 11]	The intervention has not been tested in recruited schools.
The audited files support diary reconstruction but do not measure school AI use.	6,146 unique respondents,114,151 unique episodes, all diaries close to 1,440 minutes; 120 selected teacher occupations. [2, 3, 4, 5, 6, 8]	Small occupation domain, one diary per person, no school linkage, no AI exposure or contract-hour variable.
Autumn 2025 ATUS observations have a specific shutdown gap.	Official notice and zero audited diaries for September 30-November 11. [7]	Weighting cannot recover unobserved seasonal work behavior.
The 48-school planning design is not established as sufficiently powered.	10,000 replicates per nine scenarios yield 35.6-87.7% power at 24 pairs. [12, 13, 14, 17]	Assumed Gaussian pair differences, no missingness, quality or joint-success power; mandatory independent pilot and activation gate before main recruitment.

References and rights

[1] Freedom Cheteni. Embracing the AI Revolution in Education: A Movement Thinking Approach. K12 Digest. 2023-11-09.

https://www.k12digest.com/embracing-the-ai-revolution-in-education-a-movement-thinking-approach/

Citation only. Rights retained by the original rights holder; no text or figures reproduced.

Authoritative source / evidence | checked 2026-10-15

[2] U.S. Bureau of Labor Statistics. American Time Use Survey - 2025 Microdata Files. American Time Use Survey, 2025 single-year release. 2026-06-25.

https://www.bls.gov/tus/data/datafiles-2025.htm

Citation only; no Creative Commons reuse permission asserted. Original authors, agencies and rights holders retain applicable rights; source files are not redistributed.

Authoritative source / evidence | checked 2026-10-15

[3] U.S. Bureau of Labor Statistics. American Time Use Survey Data Dictionary: 2025 Interview Data. ATUS data documentation. 2026-06.

https://www.bls.gov/tus/dictionaries/atusintcodebk25.pdf

Citation only; no Creative Commons reuse permission asserted. Original authors, agencies and rights holders retain applicable rights; source files are not redistributed.

Authoritative source / evidence | checked 2026-10-15

[4] U.S. Bureau of Labor Statistics. American Time Use Survey User’s Guide: Understanding ATUS 2003 to 2025. ATUS User’s Guide, September 2026 version. 2026-09.

https://www.bls.gov/tus/atususersguide.pdf

Citation only; no Creative Commons reuse permission asserted. Original authors, agencies and rights holders retain applicable rights; source files are not redistributed.

Authoritative source / evidence | checked 2026-10-15

[5] U.S. Census Bureau; U.S. Bureau of Labor Statistics. 2018 Census Occupation Classification Codes. 2018 Census occupation classification, official BLS-hosted codes. 2018 classification; accessed 2026-10-15.

https://www.bls.gov/tus/iocodes/census18ocodes.pdf

Citation only; no Creative Commons reuse permission asserted. Original authors, agencies and rights holders retain applicable rights; source files are not redistributed.

Authoritative source / evidence | checked 2026-10-15

[6] U.S. Bureau of Labor Statistics. American Time Use Survey Activity Lexicon 2025: no activity examples. ATUS 2025 activity coding lexicon. 2025 lexicon; accessed 2026-10-15.

https://www.bls.gov/tus/lexicons/lexiconnoex2025.pdf

Citation only; no Creative Commons reuse permission asserted. Original authors, agencies and rights holders retain applicable rights; source files are not redistributed.

Authoritative source / evidence | checked 2026-10-15

[7] U.S. Bureau of Labor Statistics. Federal Government Shutdown: 2025 American Time Use Survey. ATUS 2026 methodological notice. 2026; accessed 2026-10-15.

https://www.bls.gov/tus/notices/2026/october2025shutdown.htm

Citation only; no Creative Commons reuse permission asserted. Original authors, agencies and rights holders retain applicable rights; source files are not redistributed.

Authoritative source / evidence | checked 2026-10-15

[8] U.S. Bureau of Labor Statistics. How to use ATUS microdata files. ATUS official microdata instructions. Accessed 2026-10-15.

https://www.bls.gov/tus/other-documentation/howto.htm

Citation only; no Creative Commons reuse permission asserted. Original authors, agencies and rights holders retain applicable rights; source files are not redistributed.

Authoritative source / evidence | checked 2026-10-15

[9] Shakked Noy; Whitney Zhang. Experimental evidence on the productivity effects of generative artificial intelligence. Science, 381(6654), 187-192. 2023-07-14.

https://doi.org/10.1126/science.adh2586

Citation only; no Creative Commons reuse permission asserted. Original authors, agencies and rights holders retain applicable rights; source files are not redistributed.

Authoritative source / evidence | checked 2026-10-15

[10] Erik Brynjolfsson; Danielle Li; Lindsey Raymond. Generative AI at Work. The Quarterly Journal of Economics, 140(2), 889-942. 2025-02-04.

https://doi.org/10.1093/qje/qjae044

Citation only; no Creative Commons reuse permission asserted. Original authors, agencies and rights holders retain applicable rights; source files are not redistributed.

Authoritative source / evidence | checked 2026-10-15

[11] Jasper Maas; Simone Schoch; Urte Scholz; Pamela Rackow; Julia Schüler; Mirko Wegner; Roger Keller. Teachers' perceived time pressure, emotional exhaustion and the role of social support from the school principal. Social Psychology of Education. 2021-03-11.

https://doi.org/10.1007/s11218-020-09605-8

CC BY 4.0. © The Author(s) 2021

Authoritative source / evidence | checked 2026-10-15

[12] Marion K. Campbell; Gilda Piaggio; Diana R. Elbourne; Douglas G. Altman; for the CONSORT Group. Consort 2010 statement: extension to cluster randomised trials. BMJ, 345, e5661. 2012-09-04.

https://doi.org/10.1136/bmj.e5661

Citation only; no third-party text or instrument republished. Original authors and institutions retain their rights. Data and instrument use must follow the source terms.

Authoritative source / evidence | checked 2026-10-15

[13] An-Wen Chan; Isabelle Boutron; Sally Hopewell; David Moher; Kenneth F Schulz; Gary S Collins; Ruth Tunn; Rakesh Aggarwal; Michael Berkwits; Jesse A Berlin; Nita Bhandari; Nancy J Butcher; Marion K Campbell; Runcie C W Chidebe; Diana R Elbourne; Andrew J Farmer; Dean A Fergusson; Robert M Golub; Steven N Goodman; Tammy C Hoffmann; John P A Ioannidis; Brennan C Kahan; Rachel L Knowles; Sarah E Lamb; Steff Lewis; Elizabeth Loder; Martin Offringa; Philippe Ravaud; Dawn P Richards; Frank W Rockhold; David L Schriger; Nandi L Siegfried; Sophie Staniszewska; Rod S Taylor; Lehana Thabane; David J Torgerson; Sunita Vohra; Ian R White; Asbjørn Hróbjartsson. SPIRIT 2025 statement: updated guideline for protocols of randomised trials. BMJ, 389, e081477. 2025-04-28.

https://doi.org/10.1136/bmj-2024-081477

Citation only; no third-party text or instrument republished. Original authors and institutions retain their rights. Data and instrument use must follow the source terms.

Authoritative source / evidence | checked 2026-10-15

[14] Clare Rutterford; Andrew Copas; Sandra Eldridge. Methods for sample size determination in cluster randomized trials. International Journal of Epidemiology, 44(3), 1051-1067. 2015.

https://doi.org/10.1093/ije/dyv113

Citation only; no third-party text or instrument republished. Original authors and institutions retain their rights. Data and instrument use must follow the source terms.

Authoritative source / evidence | checked 2026-10-15

[15] OpenAI. ChatGPT [AI-assisted drafting software]. Software/tool citation, not an article author. Accessed 2026-10-15.

https://chatgpt.com/

Tool use disclosed; no model version asserted. Not a human peer reviewer.

Authoritative source / evidence | checked 2026-10-15

[16] National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1. 2023-01-26.

https://doi.org/10.6028/NIST.AI.100-1

Citation only; no Creative Commons reuse permission asserted. Original authors, agencies and rights holders retain applicable rights; source files are not redistributed.

Authoritative source / evidence | checked 2026-10-15

[17] John Barnard; Donald B. Rubin. Miscellanea. Small-sample degrees of freedom with multiple imputation. Biometrika, 86(4), 948-955. 1999-12-01.

https://doi.org/10.1093/biomet/86.4.948

Citation only; no third-party text or instrument republished. Original authors and institutions retain their rights. Data and instrument use must follow the source terms.

Authoritative source / evidence | checked 2026-10-15

This original primary-study protocol and completed methods investigation credit each cited author and data agency. BLS public-use source files were analyzed for a documented secondary audit and are not redistributed in this research package. The source studies are not republished as work by this manuscript author. The proposed new text license requires actual human author approval and the journal’s licensing decision; no source reuse license is inferred from free access.