
Microdelays disrupt online learning

In the format provided by the
authors and unedited

Table of Contents

Section S1 | Media Analyses - ACP Coverage

Section S2 | Preliminary Survey

Section S3 | Randomized Experiment

Section S4 | Math Students - Preregistered Multiverse Analyses

Section S5 | Math Students – Discussion of Non-Significant Multiverse Results

Section S6 | Aspiring Coders - Association between Visit Length and Focus Score

Section S7 | Critical-Effect-Size Value Calculations

Section S8 | Multiverse Specification Selection Rationale

Section S9 | Aspiring Coders - Pooled Multiverse Analyses

Section S10 | Math Students – Giving Up

References

Tables

Table S1 | Math Students - Descriptive Statistics for Key Variables

Table S2 | Math Students - Model Specifications Included in Multiverse Analyses

Table S3 | Math Students - School-Level Demographics

Table S4 | Math Students - Geographic Distribution of Schools

Table S5 | Math Students - Multiverse Analysis Robustness Across Specifications - Outlier Addressing Method and Covariate Inclusion

Table S6 | Math Students - Exploratory Distraction Analyses

Table S7 | Aspiring Coders - Descriptive Statistics for Key Variables

Table S8 | Aspiring Coders - Model Specifications Included in Multiverse Analyses

Table S9 | Aspiring Coders - Survival Analyses - Likelihood of Completing First Project

Table S10 | Aspiring Coders - Distribution of Users by Country

Table S11 | Aspiring Coders - Results of Linear Models Using Pooled (4-Month) Dataset

Table S12 | Aspiring Coders - Results of Milestone Attainment Using Pooled (4-Month) Dataset

Table S13a | Math Students – Marginal and Conditional R^2

Table S13b | Aspiring Coders – Marginal and Conditional R^2

Table S14 | Math Students – Giving Up

Figures

Figure S1 | Math Students - Descriptive Statistics

Figure S2 | Math Students - Forest Plot - School-Level Load Times (Median and IQR)

Figure S3 | Math Students – Multiverse Analyses – Effect of Load Times on Assignment Completion

Figure S4 | Math Students – Multiverse Analyses – Effect of Load Times on Sessions to Complete

Figure S5 | Math Students – Multiverse Analyses – Effect of Load Times on Distraction

Figure S6 | Math Students – Multiverse Analyses – Effect of Load Times on Giving Up

Figure S7 | Math Students – Multiverse Analyses – Effect of Load Times on Assignment Scores

Figure S8 | Aspiring Coders – Descriptive Statistics

Figure S9 | Aspiring Coders – Multiverse Analyses – Effect of Load Times on Engagement

Figure S10 | Aspiring Coders – Multiverse Analyses – Effect of Load Times on Focus

Figure S11 | Aspiring Coders – Multiverse Analyses – Effect of Load Times on the Likelihood of Visiting more than One Page

Figure S12 | Aspiring Coders – Multiverse Analyses – Effect of Load Times on Number of Lessons Completed

Figure S13 | Aspiring Coders – Multiverse Analyses – Effect of Load Times on Number of Pages Visited

Figure S14 | Aspiring Coders – Multiverse Analyses – Effect of Load Times on Active Time

Figure S15 | Aspiring Coders – Multiverse Analyses – Effect of Load Times on Total Visit Length

Section S1 | Media Analyses - ACP Coverage

To quantify media coverage of the Affordable Connectivity Program (ACP), we first queried the New York Times API. Specifically, we searched for any articles that (a) contained the term “Affordable Connectivity Program” anywhere (in the title or text), and (b) were published between Jan 1, 2023 and Jan 1, 2025 (for reference, the program started in 2022 and ended in 2024). The first author then read through the results and determined how many of them referred to the ACP, finding only a single result.

In additional analyses of the ACP’s media coverage, we used the Global Database of Events, Language, and Tone (GDELT) to query mentions of the ACP across three major news networks (CNN, Fox News, and MSNBC). All in all, between Jan 1, 2024 and Jan 1, 2025, across all programming, the ACP was mentioned 31 times on CNN, 17 times on Fox News, and only 5 times on MSNBC.

Section S2 | Preliminary Survey

To explore perceptions of the role of loading delays in online learning, we recruited from Prolific Academic a sample of U.S. adults that matches the general U.S. population on age, gender, and race/ethnicity (N = 500; preregistration link: <https://aspredicted.org/8mym-ff5g.pdf>).

Participants received \$0.30 for their participation.

Participants responded to the following prompt:

"Many people who start online courses end up not sticking with the course for as long as they planned.

In your opinion, what are any obstacles or challenges that could prevent someone from sticking with an online course?

You can list as few or as many as you'd like, between 1-20 entries.

Some obstacles to sticking with an online course are..."

Participants could then enter up to 20 responses in open-text boxes.

As an exploratory item, they then completed a check-all-that-apply item asking them to indicate whether they had experience participating in a range of online activities; for instance, whether they "Participated in an online course (for example, via Coursera, edX, or Udemy)".

Participants' open-ended responses were analyzed using the OpenAI API. We used GPT-4o to rate each open-text response on whether it indicates issues with internet speed, lags, or loading time (1) or not (0). Specifically, the AI system prompt first introduced the context for this request, and included the verbatim survey open-text item that participants responded to. The main part of the prompt was as follows:

Rate the response on whether it indicates issues with internet speed, lags, or loading time (1) or not (0).

For example, responses such as 'slow internet' or 'long loading times' would count as '1'. Everything else counts as '0'.

Issues that are about having internet access at all (rather than having slow internet) count as '0'.

For example "no internet access" or "lack of wifi" would count as '0'.

Only 29 participants (5.8%) mentioned barriers related to internet speed, lags, or load times.

To validate this AI coding, on an exploratory basis, participants' open-text responses were additionally coded by two coders blind to our hypothesis. The coders' ratings agreed with each other for 99% of our participants, with remaining disagreements resolved by discussion. In this resolved coding, only 17 participants (3.4%) mentioned barriers related to internet speed, lags, or load times.

AI and human coders' ratings were in agreement 97.2% of the time, with satisfactory interrater reliability for an imbalanced binary measure (Cohen's $\kappa = 0.68$, representing "substantial agreement").¹

Section S3 | Randomized Experiment

We conducted a controlled online experiment via Prolific ($N = 598$; preregistration: <https://aspredicted.org/jq36-8yw9.pdf>). In a pre-screening survey, potential participants saw the title and cover picture for 9 educational blog posts (e.g., “Goblin Sharks and Spider Crabs: Inside the Strange World of Deep-Sea Creatures”), and rated their interest in each on a 1-7 scale. If they rated at least one blog post as 6 or 7, they were asked if they would like to read that post on their own time, without receiving any compensation for doing so. Participants were not compensated to ensure that they were only reading the blog post if they were intrinsically motivated to do so. If a participant agreed, they qualified to participate in the study, and a new tab opened with the short blog post, which included approximately 525 words across 8 pages.

On the new tab, participants were randomly assigned to either read the blog post as usual (Control condition) or with an incremental 3-second delay before each page (Delay condition). Our key dependent measure was whether a participant read the full blog post, which we measured as whether a participant progressed through and viewed all 8 pages.

To analyze differences between conditions, we used the R BayesFactor package to compute a Bayes factor for the non-independence of a contingency table, using Dirichlet priors for an independent multinomial design with $\alpha = 1$ (the default).²

Participants in the delay condition were 14 percentage points less likely to read the full blog post, compared with participants in the control condition (54% vs. 70% who read the full blog post, respectively), $BF = 559.06$, representing very strong evidence supporting a difference between the two groups. The difference was similarly highly significant in an exploratory frequentist analysis, $p < .001$, Cohen’s $h = .35$.

Section S4 | Math Students - Preregistered Multiverse Analyses

The data is covered by two separate preregistrations. Our first preregistration (<https://aspredicted.org/fgfg-qjq8.pdf>) spanned ASSISTments data collected between Sep 18 - Oct 18, 2024; our second preregistration (<https://aspredicted.org/yn4zf8.pdf>) added all data until winter break of that year (Oct 18 - Dec 18, 2024).

As is to be expected in large field data, there were several features of the data that we did not account for in our preregistration plans. We note the following deviations from our preregistration plan:

1. **Outliers** The preregistration plan for our multiverse analyses included models that use untransformed load times, alongside models that account for outliers. However, the load time distribution turned out to be far more skewed than anticipated (the highest load time value after preregistered exclusions was still more than 24 standard deviations above the mean, with load times showing heavy skewness = 11.43). Thus, we highlight in the main text analyses that account for outliers, and report here the preregistered multiverse analyses, meaning that these analyses include models that do not account for outliers. We find that in the initially preregistered analyses, out of 90 models, 66 (73%) yielded a significant harmful effect of load times, 24 (27%) yielded a null effect, and 0 models (0%) yielded a significant beneficial effect. When excluding, on an exploratory basis, models that do not account for outliers, 48 out of 60 models (80%) yield a significant harmful effect of load times, and 12 models (20%) yield a null effect.
2. **Within-person fluctuation analyses** In our preregistration of data analyses from Sep 18 - Oct 18, 2024 (but not the subsequent preregistration for Oct 18 - Dec 18, 2024), the planned multiverse analysis simply predicted learning outcomes from load times. A drawback of this approach is that it does not allow distinguishing between the causal effect of microdelays on learning and confounds that are tied to both load times and learning, thereby potentially capturing spurious correlations between the two. To address this potential confounding, we add analyses in the main text that test the impact of within-person fluctuations in load times on learning. Nonetheless, both the initial and subsequent preregistration yield qualitatively similar results.
3. **Skillbuilders** During data analyses, we identified a subset of assignments with both exceptionally long load times and higher-than-average scores. Further investigation revealed that these were “skillbuilder” assignments—an adaptive assignment format that pre-loads a long sequence of problems, and has students work through the sequence until they answer three consecutive problems correctly. Because these assignments involve both long load times (due to the large number of problems loaded) and high scores, they may introduce a potential confound in the results by inducing a spurious positive correlation between load times and performance. To account for this, we excluded skillbuilder assignments from analyses on an exploratory basis.
4. **Inclusion of day-of-week as a time-varying covariate (in models that account for time-varying confounds)** To account for temporal patterns, our preregistered models included covariates for time-of-day as well as an indicator for whether a student started an assignment during school hours (with school hours defined as 7am-3pm on weekdays). Although the school hours indicator helps control for some day-level effects (e.g. weekend vs. weekday) it

may not fully capture systematic differences across days of the week. For this reason, we included day-of-week fixed effects in models that account for time-varying confounds. Nonetheless, the inclusion or exclusion of this covariate does not alter the significance of any of the results: when re-running multiverse analyses without day-of-week, 132 out of 150 models (88%) revealed a significant harmful effect of delays, 17 models (11%) revealed a null effect, and 1 model (1%) revealed a significant beneficial effect of delays.

5. **Giving up** We discovered technical artifacts in the measurement of giving up, one of our planned outcome measures. Consequently, we report analyses that include Giving Up only in the Supplementary Information (See Supplementary Section S10 for more information).

Section S5 | Math Students – Discussion of Non-Significant Multiverse Results

Models that predict assignment score and assignment completion sometimes reveal non-significant microdelay effects. Specifically, out of models predicting assignment completion, 26 (72%) reveal a negative effect and 10 (28%) reveal a null effect, with no models (0%) suggesting a positive effect. Out of models predicting assignment scores, 20 models (56%) reveal a negative effect and 15 models (42%) reveal a null effect, with one model (3%) suggesting a positive effect.

Importantly, non-significance is not randomly distributed, with models that produce non-significant results falling into two categories:

1) **Models that do not address outliers**

Given the heavy right skew in the distribution of load times (skewness = 11.39 even after preregistered exclusions), analyses that fail to address outliers could lead to biased estimates. Notably, once key covariates are controlled for (see following point), the results hold regardless of the method of addressing outliers (log-transform, winsorization, exclusion) as well as the specific threshold used (95th or 99th percentile).

2) **Models that omit assignment-level covariates**

Exploratory follow-up analyses suggest that models that use assignment length (the number of problems in an assignment) as a covariate reveal significant negative microdelay effects, whereas models that do not control for assignment length produce either a negative or null effect. We believe that this reflects a spurious positive correlation between assignment length and these outcomes. That is, longer assignments (a) take longer to load, simply because they contain more problems, and (b) tend to correlate with higher scores (and, to a lesser degree, higher completion rates) in our data, perhaps due to systematic differences in the difficulty of short vs. long assignments.

Importantly, both for assignment scores and assignment completion, all 80 models (100%) that account for outliers in any way and control for assignment length reveal a significant harmful effect of microdelays. Table S5 reports in detail how effects on assignment scores and completion vary across the multiverse as a function of model specifications.

Section S6 | Aspiring Coders - Association between Visit Length and Focus Score

Focus score is calculated as engaged time in a given visit divided by overall visit length. Models that predicted focus used overall visit length as a covariate to account for the notion that it is easier to stay completely focused in a short visit compared with a longer visit. However, including overall visit length as a covariate could lead to concerns that visit length is included as both a predictor and outcome. We address this in the following ways: (a) First, we find that the association between overall visit length and focus score is fairly weak, $r = .01$, and in the opposite direction of the hypothesized effect (b) Analysis results are qualitatively identical when overall visit length is removed from the model: in both cases, the fixed effect of load time on focus is equal to $b = -0.02$, meaning a 1-second increase in load time decreases focus by 2 percentage points.

Section S7 | Critical-Effect-Size Value Calculations

Following Perugini et al.,³ we report the critical-effect-size value for each outcome — the smallest coefficient that would reach statistical significance, given our α criterion of .05 (two-tailed). In both datasets, the large sample sizes mean that our analyses were adequately powered to detect even small effects. We note below the critical value effect sizes for each outcome across both datasets, using unstandardized regression coefficients.

In the math student dataset, critical value effect sizes are: assignment completion ($b = 0.003$), sessions to complete ($b = 0.006$), distraction ($b = 0.004$), and assignment score ($b = 0.003$).

In the aspiring coder dataset, critical value effect sizes are: engagement ($b = 0.002$), visiting > 1 page ($b = 0.001$), focus ($b = 0.002$) and lessons completed ($b = 0.009$).

Section S8 | Multiverse Specification Selection Rationale

Multiverse analyses can help increase transparency by testing results' robustness across a broad range of model specifications. However, an important consideration in multiverse analyses is elucidating how different specifications were selected for inclusion in the multiverse.⁴ Below, we outline the different dimensions varied in our multiverse analyses (in both datasets) and the rationale for selecting different variants, in keeping with the guidance outlined by Del Giudice and Gangestad (2001).

Outlier Handling

As is often the case in field data, the distribution of raw load times in both datasets exhibited severe right skew. Because extreme outliers can exert a disproportionate pull on model estimates, actively addressing outliers is objectively more justified than leaving the data fully untransformed (*principled nonequivalence* per Del Giudice and Gangestad's framework). For this reason, we report in the main text only analyses that address outliers (while reporting preregistered analyses that do not address outliers in the Supplementary Information).

However, we faced genuine uncertainty regarding the optimal approach for addressing outliers. For instance, we had no *a priori* theoretical reason to deem 95th-percentile winsorization strictly superior to 99th-percentile winsorization, or to exclusion-based strategies. In such scenarios, multiverse analyses should be used to transparently map and explore the specification space.⁴ Consequently, we populated the multiverse for both studies with a diverse array of outlier-handling methods (including winsorization, exclusion, and log-transformations across multiple thresholds), and plot how effects vary as a function of different outlier handling methods, in Supplementary Figures S3-7 and S9-S15.

Covariate inclusion

Models with alternative covariate sets can estimate different causal effects.⁴ In both datasets, temporal variables—specifically time of day and day of week—may introduce systemic variance in both server load and user engagement. Furthermore, in the math students data, assignment length (the number of problems) acts as an additional critical confound, as longer assignments inherently take longer to load and systematically correlate with higher scores in our data. Therefore, we believe that models that include these covariates are clearly superior to models that omit them (*principled nonequivalence*). As we discuss in detail in Supplementary Section S5 and Supplementary Table S5, models that control for assignment length yield robust microdelay effects, whereas models that omit assignment length sometimes do not, which we ascribe to the spurious positive correlation they induce between load times and achievement. Nonetheless, our preregistered multiverse analyses still included specifications without covariates to demonstrate robustness.

Dichotomization cutoffs

When operationalizing binary variables such as distraction (in the math student dataset) and engagement (in the aspiring coder dataset), we frequently encountered decision points that lacked definitive, gold-standard operational definitions. For instance, in the math students data, we defined "distraction" based on the lag between the time an assignment finished loading and the time a student started the assignment, originally establishing dichotomization cutoffs at 10 seconds and 1 minute. Similarly, in the aspiring coder data, metrics of engagement—such as classifying whether a user engaged with the platform for more or less than one minute—rely on boundary thresholds that do not have a single, objectively superior value. We lacked a definitive *a priori* rationale to substantiate that a 10-second distraction threshold is superior to a 30-second threshold, or that one specific engagement cutoff is strictly better than another (a possible exception is our threshold for number of pages visited, separating users who visited 1 page from those who visited more than one page—equivalent to the widely-used *bounce rate* metric in online platforms). To ensure our findings were robust to these arbitrary thresholds, we treated these boundaries as exploratory parameters, testing a broad range of thresholds for engagement (30 seconds - 10 minutes) and distraction (10 seconds - 5 minutes, including in exploratory analyses, see Supplementary Table S6).

Section S9 | Aspiring Coders - Pooled Multiverse Analyses

To assess microdelay effects across a longer time horizon, we re-ran our key analyses on A pooled dataset that includes both our preregistered dataset (July 4 - Sep 4, 2023) and an earlier 2-month period which we had already analyzed prior (May 4 - July 4, 2023). Results of main linear models as well as survival models can be found in Supplementary Tables S11 and S12. Multiverse results revealed that of 180 models, 180 (100%) revealed statistically significant and harmful effects of microdelays.

Section S10 | Math Students – Giving Up

One of our preregistered outcomes in the math student analyses was giving up—the share of problems in an assignment for which a student elected to reveal the answer (i.e., used all three hints; see Score in the Methods). While analyzing the results, we uncovered several measurement artifacts in this variable. These included:

- a) Assignments in which the number of given-up problems exceeded the total number of problems in an assignment
- b) Assignments in which the number of given-up problems was too high given the number of hints used. Because a learner can only reveal the answer to a problem after requesting 2 hints, the number of given-up answers should not exceed half the number of requested hints (e.g., if a user requested 4 hints in an assignments, they could have at most revealed 2 answers).

Because of concerns about the validity of this measure, we do not report analyses predicting giving up in the main text. Nonetheless, for completeness, we report the results for all preregistered analyses including the giving up measure here (specifically, here, in Supplementary Table S14, and Supplementary Figure S6).

Giving up was also included in our preregistered multiverse. When models that predict giving up are included in the multiverse, 180 out of 210 models (86%) revealed a statistically significant harmful effect of load times, 25 (12%) revealed a null effect, and 5 (2%) revealed a statistically significant beneficial effect.

As mentioned, given the measurement issues we document above, we do not believe that these results are valid; they are only reported for transparency.

References

1. Landis, J. R. & Koch, G. G. The measurement of observer agreement for categorical data. *Biometrics* **33**, 159–174 (1977).
2. Gunel, E. & Dickey, J. Bayes Factors for Independence in Contingency Tables. *Biometrika* **61**, 545–557 (1974).
3. Perugini, A. *et al.* The Benefits of Reporting Critical-Effect-Size Values. *Adv. Methods Pract. Psychol. Sci.* **8**, 25152459251335298 (2025).
4. Del Giudice, M. & Gangestad, S. W. A Traveler’s Guide to the Multiverse: Promises, Pitfalls, and a Framework for the Evaluation of Analytic Decisions. *Adv. Methods Pract. Psychol. Sci.* **4**, 2515245920954925 (2021).

Tables

Table S1 | Math Students - Descriptive Statistics for Key Variables

Level	Variable	Mean (SD)	Median [IQR]
Session	Load time (seconds)	1.55 (1.30)	1.21 [0.94, 1.67]
	Completed assignments	0.80	
	Sessions to complete	1.41 (0.82)	1.00 [1.00, 2.00]
	Distraction	0.20	
	Assignment score	0.66 (0.34)	0.75 [0.48, 0.96]
User	Overall number of sessions per user	11.82 (12.22)	7.00 [3.00, 16.00]
	Overall number of assignments per user	8.37 (8.73)	5.00 [3.00, 11.00]

Note. SD = Standard deviation. IQR = Interquartile range.

For binary variables (*Completed assignments* and *Distraction*), the "Mean (SD)" column represents the share of observations where an assignment was completed, or the share of observations that involved a lag longer than 10 seconds (for alternative distraction thresholds, see Supplementary Table S6).

Table S2 | Math Students - Model Specifications Included in Multiverse Analyses

Variable	Values
Outcome	1. Assignment completion 2. Sessions to complete 3. Distraction a. 10-second threshold b. 1-minute threshold 4. Giving up a. % of given-up problems b. Count of given-up problems 5. Assignment score
Transformation for handling outliers in load time	1. Winsorizing (95th percentile) 2. Winsorizing (99th percentile) 3. Exclusion (95th percentile) 4. Exclusion (99th percentile) 5. Log-transform
Include covariates?	1. No covariates 2. Assignment-level covariate (number of problems in assignment; time of day; day of week) 3. Assignment-level and school-level covariates (% Non-Hispanic White students and % of students eligible for free/reduced-price lunch)
Include school-hours?	1. Exclude 2. Include

Note. *Sessions to complete* represents the number of sessions a student needed to complete an assignment. *Distraction* represents whether a student failed to start an assignment shortly after loading it. *Giving up* represents the share of problems in an assignment for which a student chose to reveal the answer (see Supplementary Section S10).

Table S3 | Math Students - School-Level Demographics

Variable	Mean (SD)	Median [IQR]
Number of Students	647.88 (439.34)	562.00 [354.00, 800.00]
% White	0.49 (0.34)	0.51 [0.13, 0.83]
% Black/African American	0.19 (0.27)	0.05 [0.01, 0.24]
% Hispanic	0.22 (0.25)	0.11 [0.05, 0.27]
% Asian/Pacific Islander	0.04 (0.09)	0.01 [0.00, 0.04]
% Amer. Indian/Alaska Native	0.01 (0.05)	0.00 [0.00, 0.01]
% Two or more races	0.05 (0.04)	0.04 [0.02, 0.07]
% Free or reduced-price lunch	0.55 (0.24)	0.55 [0.39, 0.73]
% Female	0.49 (0.03)	0.49 [0.47, 0.51]

Note. Demographic data were obtained from the National Center for Education Statistics Common Core of Data (NCES CCD).

Table S4 | Math Students - Geographic Distribution of Schools

Region	N	%	% of U.S. Population
Midwest	116	31.10	20.8%
Northeast	76	20.38	17.3%
South	113	30.29	38.1%
West	68	18.23	23.7%
Total	373	100%	100%

Note. % of U.S. Population represents estimates from the 2020 U.S. Census.

Table S5 | Math Students - Multiverse Analysis Robustness Across Specifications - Outlier Addressing Method and Covariate Inclusion

Outcome	Transformed ?	Covariates ?	Total Models	Negative Sig.	Positive Sig.
Completion	No	No	2	0 (0%)	0 (0%)
	No	Yes	4	0 (0%)	0 (0%)
	Yes	No	10	6 (60%)	0 (0%)
	Yes	Yes	20	20 (100%)	0 (0%)
Score	No	No	2	0 (0%)	0 (0%)
	No	Yes	4	0 (0%)	0 (0%)
	Yes	No	10	0 (0%)	1 (10%)
	Yes	Yes	20	20 (100%)	0 (0%)
Sessions	No	No	2	2 (100%)	0 (0%)
	No	Yes	4	4 (100%)	0 (0%)
	Yes	No	10	10 (100%)	0 (0%)
	Yes	Yes	20	16 (80%)	0 (0%)
Distraction	No	No	4	4 (100%)	0 (0%)
	No	Yes	8	8 (100%)	0 (0%)
	Yes	No	20	20 (100%)	0 (0%)
	Yes	Yes	40	40 (100%)	0 (0%)
Giving Up	No	No	4	4 (100%)	0 (0%)
	No	Yes	8	8 (100%)	0 (0%)

Yes	No	20	18 (90%)	0 (0%)
Yes	Yes	40	30 (75%)	4 (10%)

Note. This table breaks down multiverse results for assignment completion and assignment score by whether models addressed outliers (*Transformed?*) and whether models included covariates (*Covariates?*). *Transformed: No* includes models that use untransformed load times; *Transformed: Yes* includes models that applied winsorization, exclusion, or log-transformation. *Covariates: No* includes models without covariates; *Covariates: Yes* includes models with assignment-level covariates (number of problems, time of day, day of week) and, in some specifications, school-level demographics. *Negative Sig.* and *Positive Sig.* indicate the number (and percentage) of models yielding a statistically significant negative or positive effect of load times, respectively ($p < .05$, two-sided).

Table S6 | Math Students - Exploratory Distraction Analyses

	Preregistered		Exploratory		
	> 10 Sec	> 1 Min	> 20 Sec	> 30 Sec	> 5 Min
Load time (unstandardized)	0.028 [0.024, 0.031]	0.015 [0.013, 0.018]	0.022 [0.019, 0.025]	0.021 [0.018, 0.023]	0.010 [0.008, 0.011]
Load time (standardized)	0.033 [0.028, 0.037]	0.029 [0.024, 0.033]	0.031 [0.026, 0.035]	0.032 [0.028, 0.037]	0.025 [0.020, 0.029]
$t(df)$	14.73 (151,245)	12.83 (165,015)	13.74 (158,725)	14.46 (162,253)	10.94 (162,562)
p	< .001	< .001	< .001	< .001	< .001
N (observations)	208,082	208,082	208,082	208,082	208,082
N (students)	23,766	23,766	23,766	23,766	23,766
N (classes)	1,334	1,334	1,334	1,334	1,334
Mean (outcome)	0.20	0.07	0.13	0.10	0.04
R^2	0.010	0.009	0.010	0.010	0.008

Note. This table shows regression results for models that predict distraction at different thresholds ranging from 10 seconds to 5 minutes. *Load time (unstandardized)* indicates the untransformed regression coefficient, or the impact of a 1-second increase in load times on the dependent variable. *Load time (standardized)* indicates the same coefficient, but transformed to represent standard deviations. The 10 second and 1 minute thresholds are part of the preregistered analyses, whereas the 20 second, 30 second, and 5 minute thresholds are exploratory. All models used load times winsorized to the 95th percentile, included time-varying and school-level covariates, and include random intercepts for students nested within classes and cross-classified random intercepts for each unique assignment. R^2 represents marginal R^2 . Statistical significance was assessed using two-sided t -tests with Satterthwaite's approximation for degrees of freedom.

Table S7 | Aspiring Coders - Descriptive Statistics for Key Variables

Variable^a	Mean (SD)	Median [IQR]
Load time (seconds)	4.34 (5.39)	2.1 [1.2, 4.5]
Visit length (minutes)	23.45 (67.76)	4 [0, 30]
Engaged time in visit (minutes)	12.8 (20.33)	1.33 [0.02, 17.83]
% Engaged time in visit (% focus)	0.62 (0.33)	0.7 [0.34, 0.93]
Number of pages visited	9.75 (18.34)	3 [1, 10]
Number of lessons completed	6.18 (15.93)	0 [0, 5]
Engaged $\geq$ 30 sec	0.58	
Engaged $\geq$ 1 min	0.52	
Engaged $\geq$ 5 min	0.38	
Engaged $\geq$ 10 min	0.32	
Visited $>$ 1 page	0.63 (0.48)	

Note. SD = Standard deviation. IQR = Interquartile range.

^a For binary variables (engagement thresholds and visiting $>$ 1 page), the *Mean (SD)* column represents the share of visits meeting that criterion.

Table S8 | Aspiring Coders - Model Specifications Included in Multiverse Analyses

Variable	Values
Outcome	1. Engagement (i.e., active time over): a. 30 seconds b. 1 minute c. 5 minutes d. 10 minutes 2. Multi-page visit 3. Focus 4. Number of completed lessons 5. Active time 6. Number of pages visited 7. Visit length in minutes
Transformation for handling outliers	1. 80th percentile winsorization 2. 90th percentile winsorization 3. 95th percentile winsorization 4. 99th percentile winsorization 5. 80th percentile exclusion 6. 90th percentile exclusion 7. 95th percentile exclusion 8. 99th percentile exclusion 9. Log+1 transformation
Include time-varying covariates (time of day, day of week)?	1. Yes 2. No

Note. *Engagement* represents whether a user's active time in a given visit exceeded a given threshold (30 seconds, 1 minute, 5 minutes, or 10 minutes). *Multi-page visit* represents whether a user visited more than one page. *Focus* represents active time divided by total visit length. *Active time* represents time during which the platform was in focus on the user's computer. All models included random intercepts for users nested within countries and cross-classified random intercepts for each unique starting page URL.

Table S9 | Aspiring Coders - Survival Analyses - Likelihood of Completing First Project

Transform	HR	95% CI	<i>p</i>
Log+1 transform	0.900	[0.856, 0.947]	< .001
80th percentile (winsorization)	0.873	[0.831, 0.917]	< .001
90th percentile (winsorization)	0.912	[0.867, 0.959]	< .001
95th percentile (winsorization)	0.934	[0.888, 0.982]	0.008
99th percentile (winsorization)	0.961	[0.913, 1.011]	0.122
80th percentile (exclusion)	0.914	[0.870, 0.960]	< .001
90th percentile (exclusion)	0.921	[0.876, 0.968]	0.001
95th percentile (exclusion)	0.928	[0.882, 0.976]	0.004
99th percentile (exclusion)	0.943	[0.895, 0.993]	0.027

Note. This table represents the results of a Cox proportional hazards models predicting time to completing the first course project in freeCodeCamp's web design course from a user's mean load time. Each row represents a single model using a different outlier-handling approach, denoted in the *Transform* column. *HR [95% CI]* indicates the hazard ratio of a one standard deviation increase in load times, and the 95% confidence interval around that slope, computed using the Wald method. *p* represents the p-value, with statistical significance assessed using two-sided Wald z-tests.

Table S10 | Aspiring Coders - Distribution of Users by Country

Country	N	%
Argentina	14,123	25.4%
Colombia	8,818	15.9%
Mexico	6,941	12.5%
Spain	6,087	11.0%
Peru	3,496	6.3%
Chile	2,474	4.5%
Venezuela	2,254	4.1%
Dominican Republic	1,603	2.9%
Uruguay	1,513	2.7%
United States	1,203	2.2%
Ecuador	920	1.7%
Bolivia	646	1.2%
Guatemala	566	1.0%
Costa Rica	559	1.0%
Honduras	445	0.8%
Paraguay	438	0.8%
El Salvador	422	0.8%
Brazil	287	0.5%
Panama	253	0.5%
Cuba	237	0.4%
Nicaragua	227	0.4%
Other	2,031	3.7%

Note. *N* refers to the number of users in our data from each country, and % denotes the share of users from each country. The *Other* category comprises countries with fewer than 200 visits during the study period

Table S11 | Aspiring Coders - Results of Linear Models Using Pooled (4-Month) Dataset

	Engagement	Visiting > 1 page	Focus (%)	# Lessons completed
Load time (unstandardized)	-0.049 [-0.050, -0.048]	-0.055 [-0.056, -0.054]	-0.017 [-0.018, -0.016]	-0.215 [-0.223, -0.207]
Load time (standardized)	-0.099 [-0.102, -0.096]	-0.116 [-0.118, -0.113]	-0.052 [-0.056, -0.048]	-0.075 [-0.078, -0.072]
t(df)	-71.50 (273,374)	-91.92 (260,105)	-26.88 (166,717)	-54.10 (290,347)
<i>p</i>	< .001	< .001	< .001	< .001
N (users)	102,468	102,468	74,697	102,468
N (countries)	36	36	36	36
N (first pages)	397	397	397	397
N (visits)	356,625	356,625	218,822	356,625
Mean (outcome)	0.523	0.634	0.618	1.972
R ²	0.010	0.016	0.003	0.007

Note. This table reports results from multilevel linear models using the pooled 4-month dataset (May 4 to September 4, 2023), extending the primary 2-month analysis. *Load time (unstandardized)* indicates the untransformed regression coefficient, or the impact of a 1-second increase in load times on the dependent variable. *Load time (standardized)* indicates the same coefficient, but transformed to represent standard deviations. *Engagement* represents whether a user's active time exceeded one minute. *Visiting > 1 page* represents whether a user visited more than one page. *Focus* represents active time divided by total visit length. *Lessons completed* refers to the number of lessons completed per visit. All models use load times winsorized to the 80th percentile, control for time-varying covariates (day of week and time of day), and include random intercepts for users nested within countries and cross-classified random intercepts for each unique starting page URL. *R*² represents marginal *R*². Statistical significance was assessed using two-sided *t*-tests with Satterthwaite's approximation for degrees of freedom.

Table S12 | Aspiring Coders - Results of Milestone Attainment Using Pooled (4-Month) Dataset

Transform	<i>HR</i>	95% CI	<i>p</i>
Log+1 transform	0.877	[0.846, 0.909]	< .001
80th percentile (winsorization)	0.855	[0.826, 0.886]	< .001
90th percentile (winsorization)	0.891	[0.859, 0.924]	< .001
95th percentile (winsorization)	0.914	[0.881, 0.948]	< .001
99th percentile (winsorization)	0.938	[0.904, 0.973]	< .001
80th percentile (exclusion)	0.891	[0.859, 0.923]	< .001
90th percentile (exclusion)	0.891	[0.859, 0.924]	< .001
95th percentile (exclusion)	0.895	[0.863, 0.929]	< .001
99th percentile (exclusion)	0.929	[0.895, 0.964]	< .001

Note. This table reports results of Cox proportional hazards models predicting time to completing the first course project in freeCodeCamp’s web design course, using the pooled 4-month dataset (May 4 to September 4, 2023). Each row represents a single model using a different outlier-handling approach, denoted in the *Transform* column. *HR* indicates the hazard ratio of a one-standard-deviation increase in a user’s mean load time, with the 95% confidence interval computed using the Wald method. *p* represents the *p*-value, with statistical significance assessed using two-sided Wald *z*-tests.

Table S13a | Math Students – Marginal and Conditional R^2

Outcome	Marginal R^2	Conditional R^2
Completion	0.287	0.803
Sessions	0.032	0.354
Distraction	0.009	0.132
Score	0.145	0.610

Note. R^2 values for the models in Table 1. *Marginal R^2* reflects the variance explained by the fixed effects alone; *conditional R^2* reflects the variance explained by the fixed and random effects together.

Table S13b | Aspiring Coders – Marginal and Conditional R^2

Outcome	Marginal R^2	Conditional R^2
Engagement	0.010	0.358
Visiting > 1 page	0.016	0.460
Focus (%)	0.003	0.360
# Lessons completed	0.006	0.364

Note. R^2 values for the models in Table 2. *Marginal R^2* reflects the variance explained by the fixed effects alone; *conditional R^2* reflects the variance explained by the fixed and random effects together.

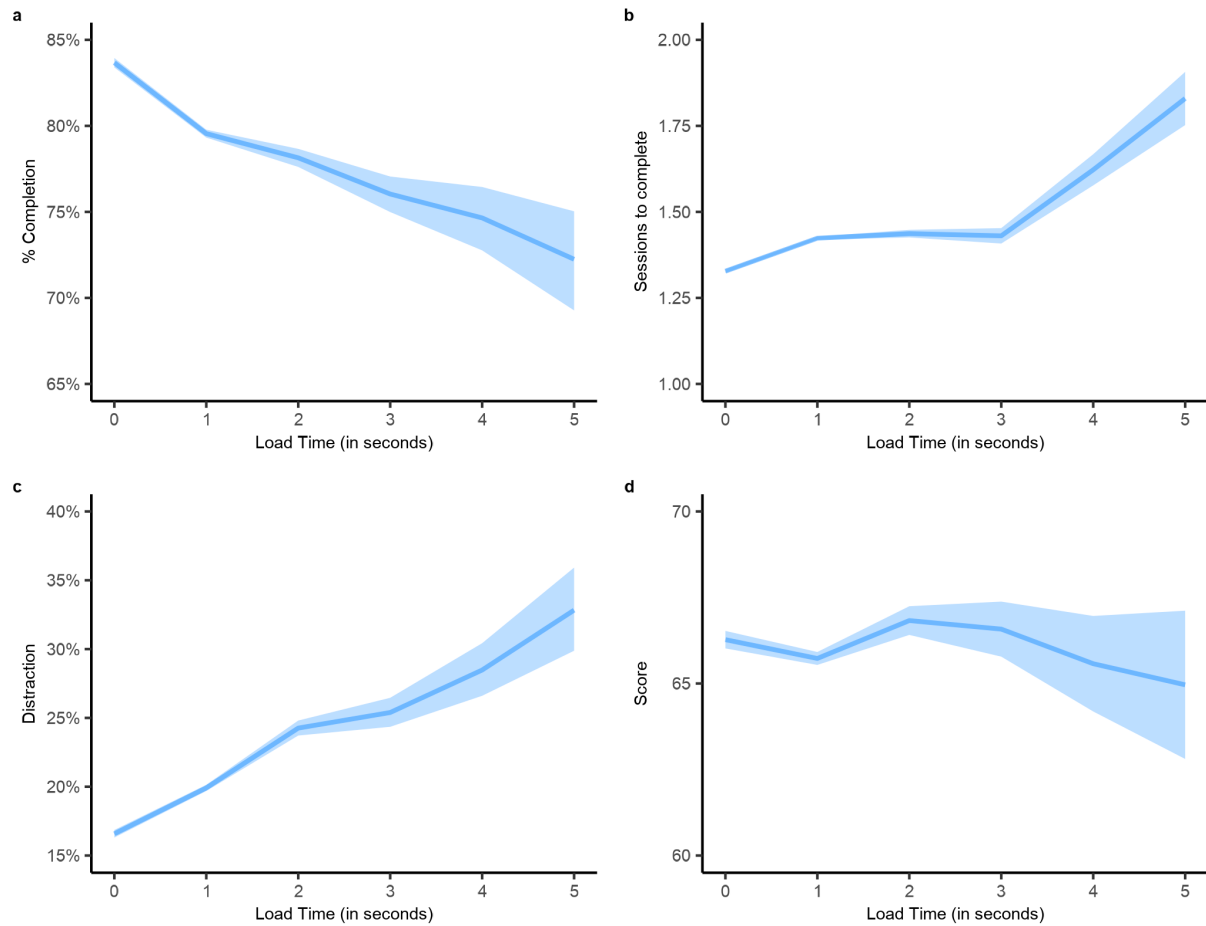
Table S14 | Math Students – Giving Up

	Giving up
Load time (unstandardized)	0.010 [0.005, 0.015]
Load time (standardized)	0.009 [0.005, 0.014]
t(df)	3.96 (182,451)
p	<.001
N (observations)	208,082
N (students)	23,766
N (classes)	1,334
Outcome mean	0.11
R ²	0.002

Note. This model uses the same specification as the models in Table 1: namely, load times are winsorized to the 95th percentile, and the. Model includes time-varying covariates (number of problems in an assignment, time of day, day of week, and an indicator for school hours) as well as school-level demographic covariates (share of students eligible for free or reduced-price lunch, in addition to the share of non-Hispanic White students). The model also includes random intercepts for students nested within classes as well as cross-classified random intercepts for each unique assignment. Giving up refers to the share of problems in an assignment for which a student chose to view the answer instead of trying to solve the problem themselves. R² represents marginal R². Statistical significance was assessed using a two-sided t-test with Satterthwaite's approximation for degrees of freedom. As described in Supplementary Section S10, we uncovered technical artifacts in the measurement of this column, such that we do not believe that these results are valid; they are only reported for transparency.

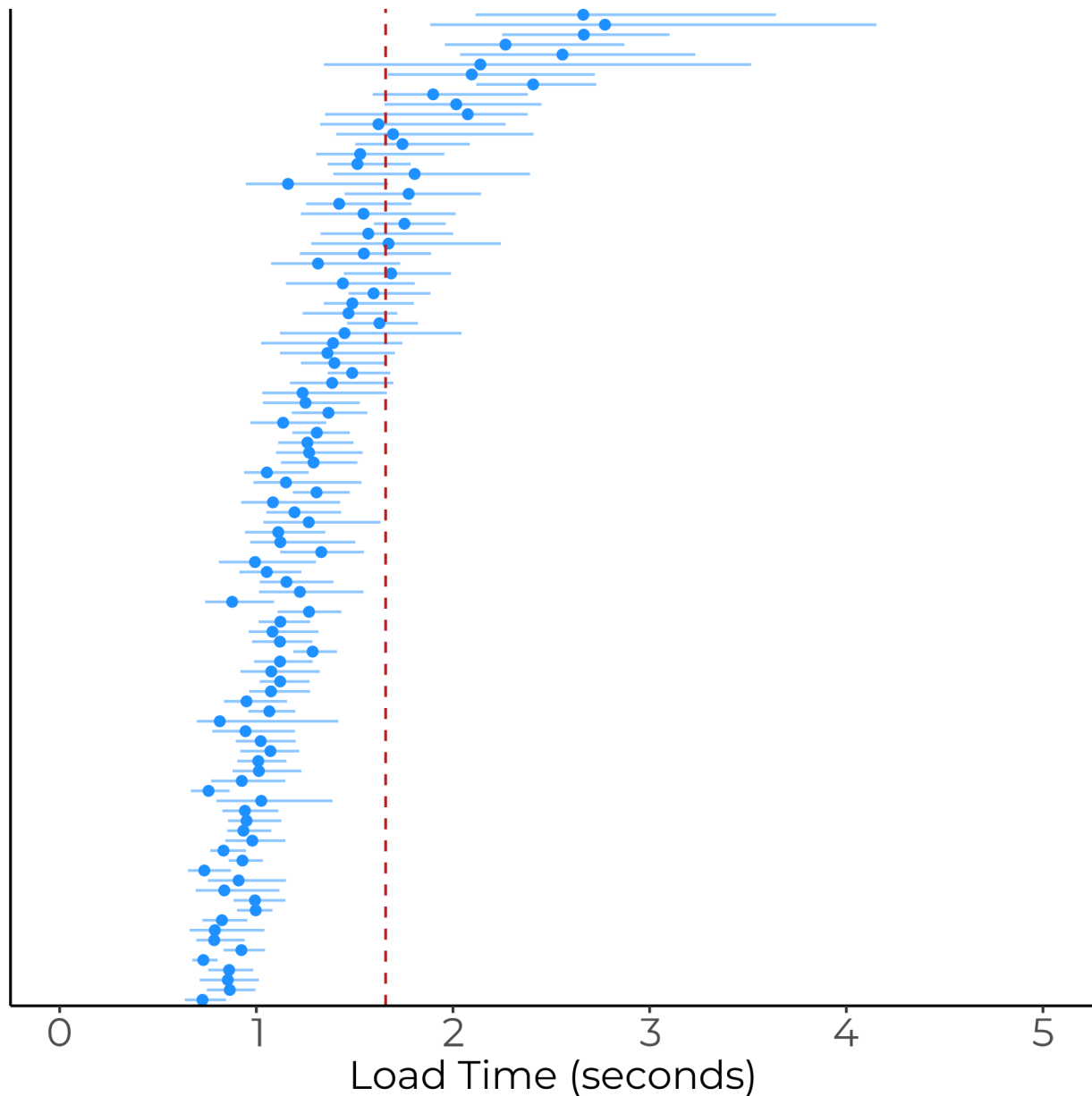
Figures

Figure S1 | Math Students - Descriptive Statistics



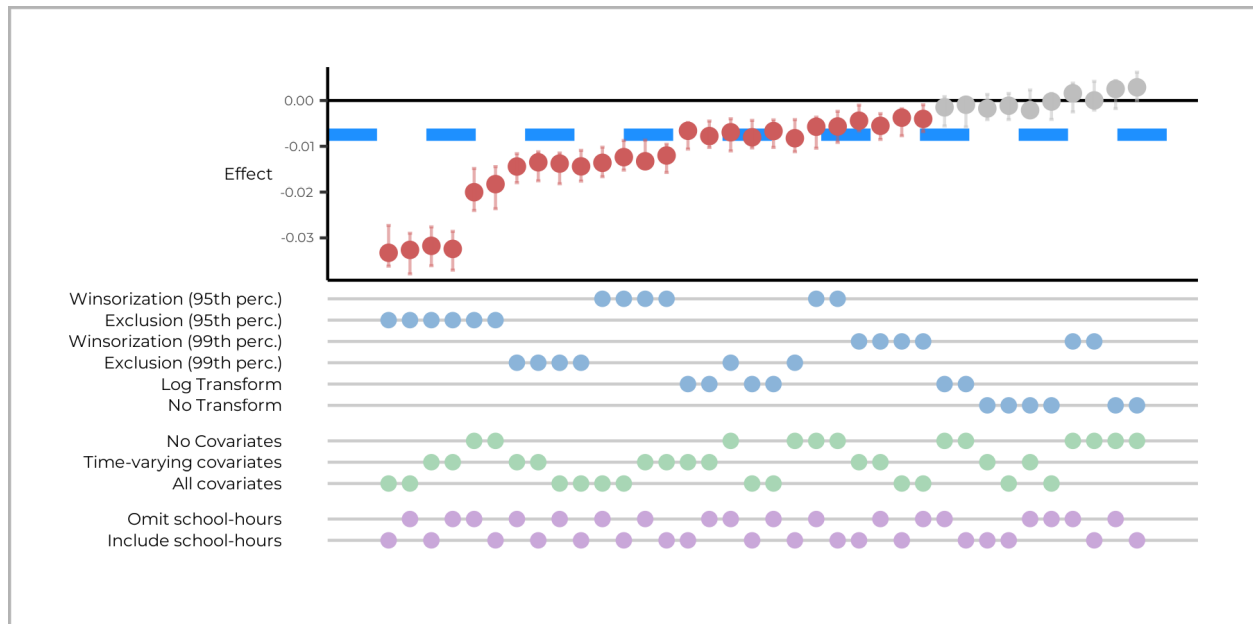
Note. Each x-axis tick represents a single one-sec bin (e.g., “0” represents the 0-1 second bin). Line height represents the average of each outcome within each bin, and ribbons represent the 95% confidence interval. These descriptive statistics do not adjust for clustering or covariates; see Table 1 (main text) for model-based estimates that account for these influences. $N = 228,734$ assignments (% completion, distraction, giving up); $N = 184,129$ assignments (sessions to complete); $N = 220,411$ assignments (score).

Figure S2 | Math Students - Forest Plot - School-Level Load Times (Median and IQR)



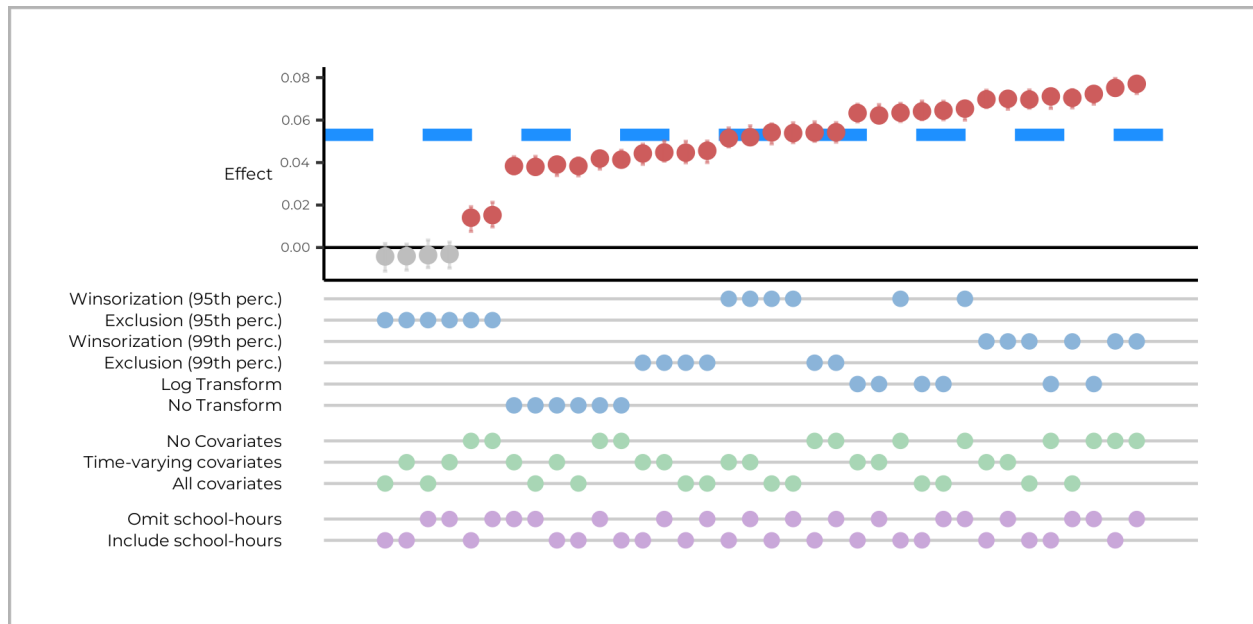
Note. Each point and its associated error bars represent a single school. Points represent the median load time in a given school, and the ends of the error bars represent the interquartile range (25th and 75th percentiles) of load times. For legibility, this figure only includes the top 100 schools, based on the number of students in the data. $N = 162,124$ observations from 17,824 students.

Figure S3 | Math Students – Multiverse Analyses – Effect of Load Times on Assignment Completion

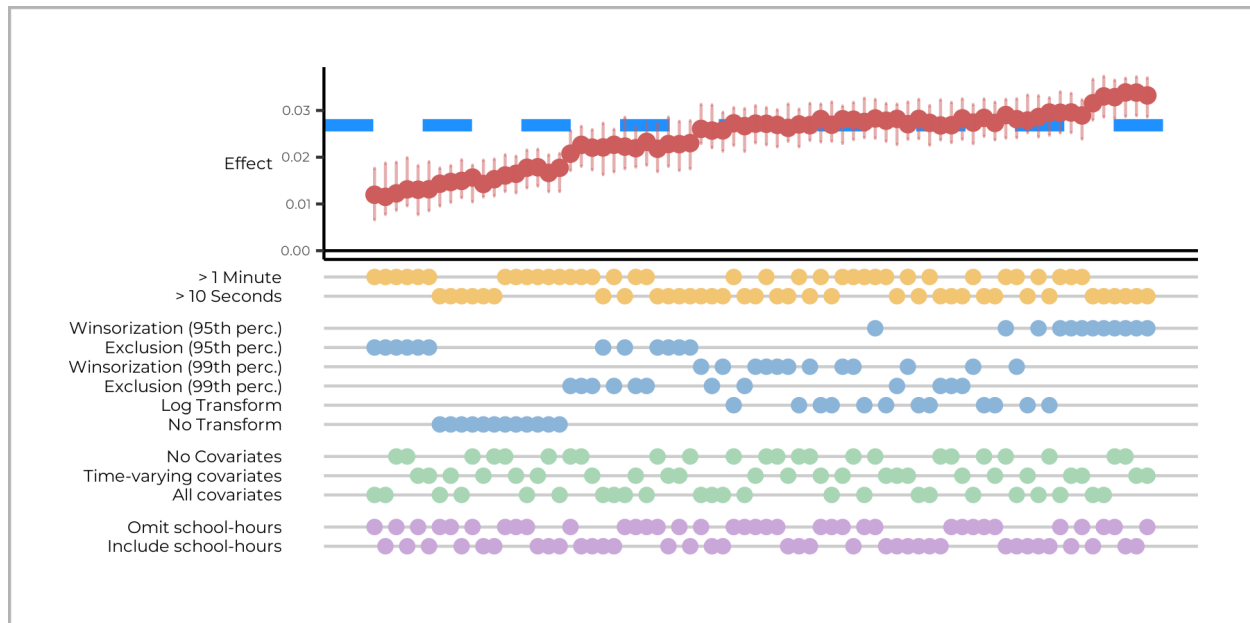


Note. This figure presents results from multiverse analyses examining the relationship between load times and assignment completion rates among math students. Each column or vertical slice represents a single model. The bottom panel delineates the model specification, and the top panel plots each model's corresponding effect size (standardized regression coefficient), with points representing the effect size and error bars representing 95% confidence intervals. Red points indicate statistically significant effects, while gray points represent non-significant effects. The dashed blue line represents the median effect size. The multiverse dimensions plotted include: different outlier handling methods (none, 95th- or 99th-percentile winsorization, 95th- or 99th-percentile exclusion, log+1 transformation); covariate sets included (none, time-varying, time-varying and school-level covariates); and whether an indicator for school-hours is included or omitted.

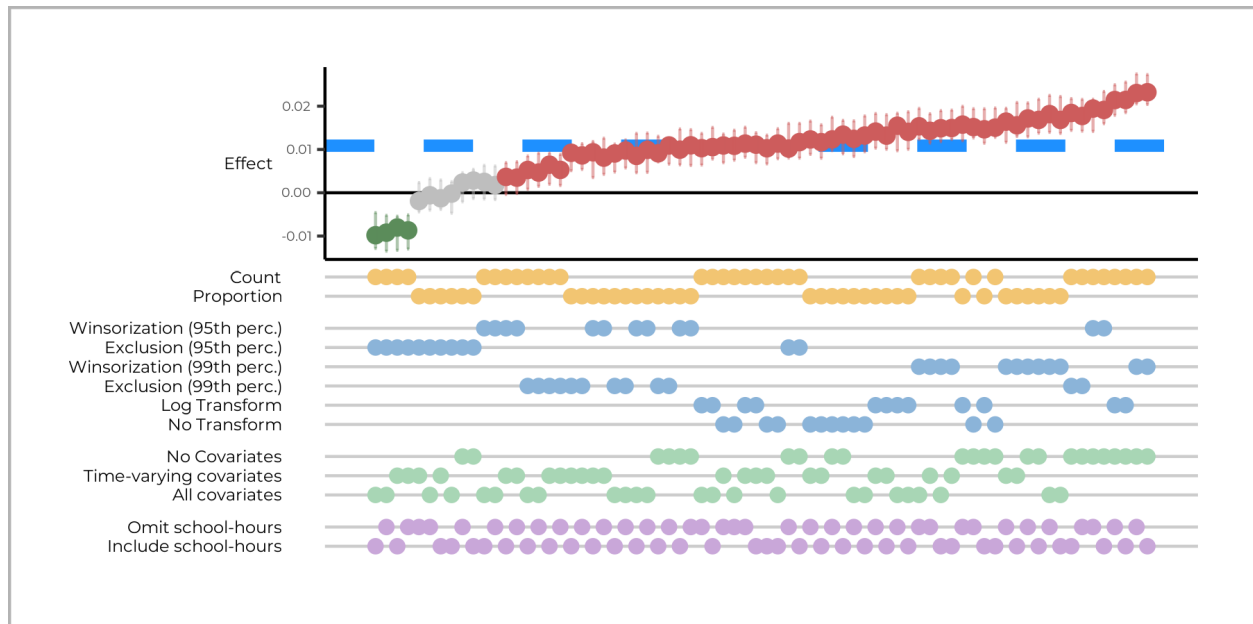
Figure S4 | Math Students – Multiverse Analyses – Effect of Load Times on Sessions to Complete



Note. This figure presents results from multiverse analyses examining the relationship between load times and the number of sessions required to complete a math assignment among math students. Each column or vertical slice represents a single model. The bottom panel delineates the model specification, and the top panel plots each model's corresponding effect size (standardized regression coefficient), with points representing the effect size and error bars representing 95% confidence intervals. Red points indicate statistically significant effects, while gray points represent non-significant effects. The dashed blue line represents the median effect size. The multiverse dimensions plotted include: different outlier handling methods (none, 95th- or 99th-percentile winsorization, 95th- or 99th-percentile exclusion, log+1 transformation); covariate sets included (none, time-varying, time-varying and school-level covariates); and whether an indicator for school-hours is included or omitted.

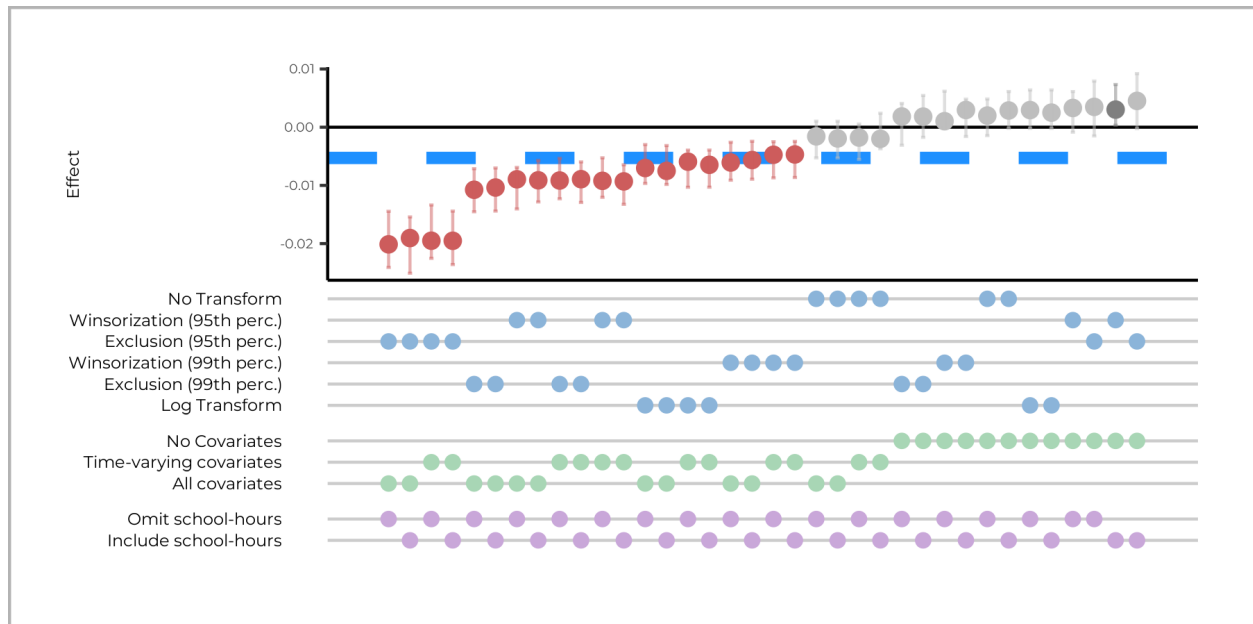
Figure S5 | Math Students – Multiverse Analyses – Effect of Load Times on Distraction

Note. This figure presents results from multiverse analyses examining the relationship between load times and distraction among math students, with distraction measured as whether a student failed to start an assignment within (10 seconds / one minute) of its loading. Each column or vertical slice represents a single model. The bottom panel delineates the model specification, and the top panel plots each model's corresponding effect size (standardized regression coefficient), with points representing the effect size and error bars representing 95% confidence intervals. Red points indicate statistically significant effects, while gray points represent non-significant effects. The dashed blue line represents the median effect size. The multiverse dimensions plotted include: whether the threshold for distraction is 1 minute or 10 seconds; different outlier handling methods (none, 95th- or 99th-percentile winsorization, 95th- or 99th-percentile exclusion, log+1 transformation); covariate sets included (none, time-varying, time-varying and school-level covariates); and whether an indicator for school-hours is included or omitted.

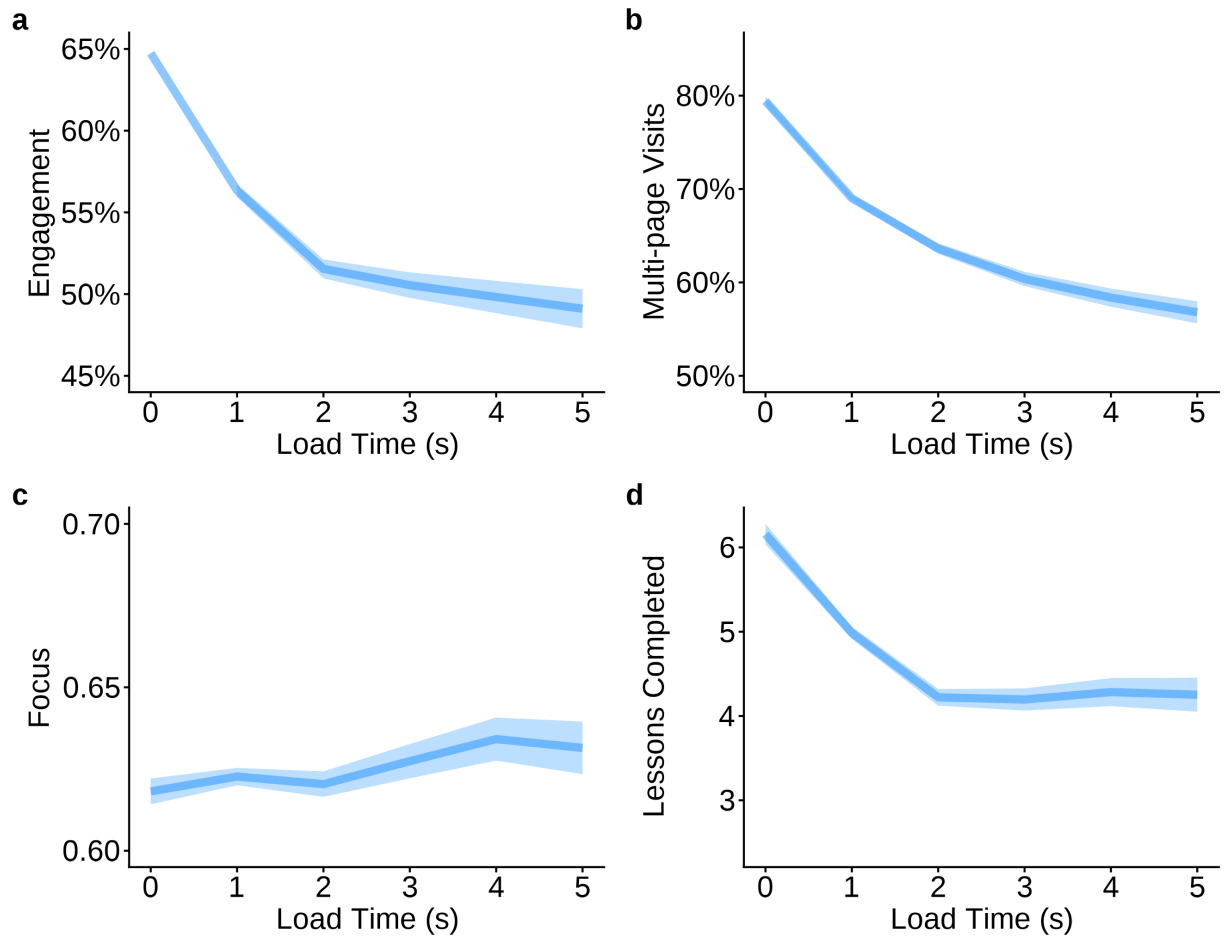
Figure S6 | Math Students – Multiverse Analyses – Effect of Load Times on Giving Up

Note. This figure presents results from multiverse analyses examining the relationship between load times and giving up among math students, with giving up measured as the share (or count) of problems in an assignment for which a student chose to reveal the answer. Each column or vertical slice represents a single model. The bottom panel delineates the model specification, and the top panel plots each model's corresponding effect size (standardized regression coefficient), with points representing the effect size and error bars representing 95% confidence intervals. Red points indicate statistically significant harmful effects, gray points represent non-significant effects, and green points represent statistically significant beneficial effects. The dashed blue line represents the median effect size. The multiverse dimensions plotted include: whether the outcome represents the count or proportion of problems in an assignment that a student has given up on; different outlier handling methods (none, 95th- or 99th-percentile winsorization, 95th- or 99th-percentile exclusion, log+1 transformation); covariate sets included (none, time-varying, time-varying and school-level covariates); and whether an indicator for school-hours is included or omitted. As explained above in Supplementary Section S10, given the measurement issues we document above, we do not believe that these results are valid; they are only reported for transparency.

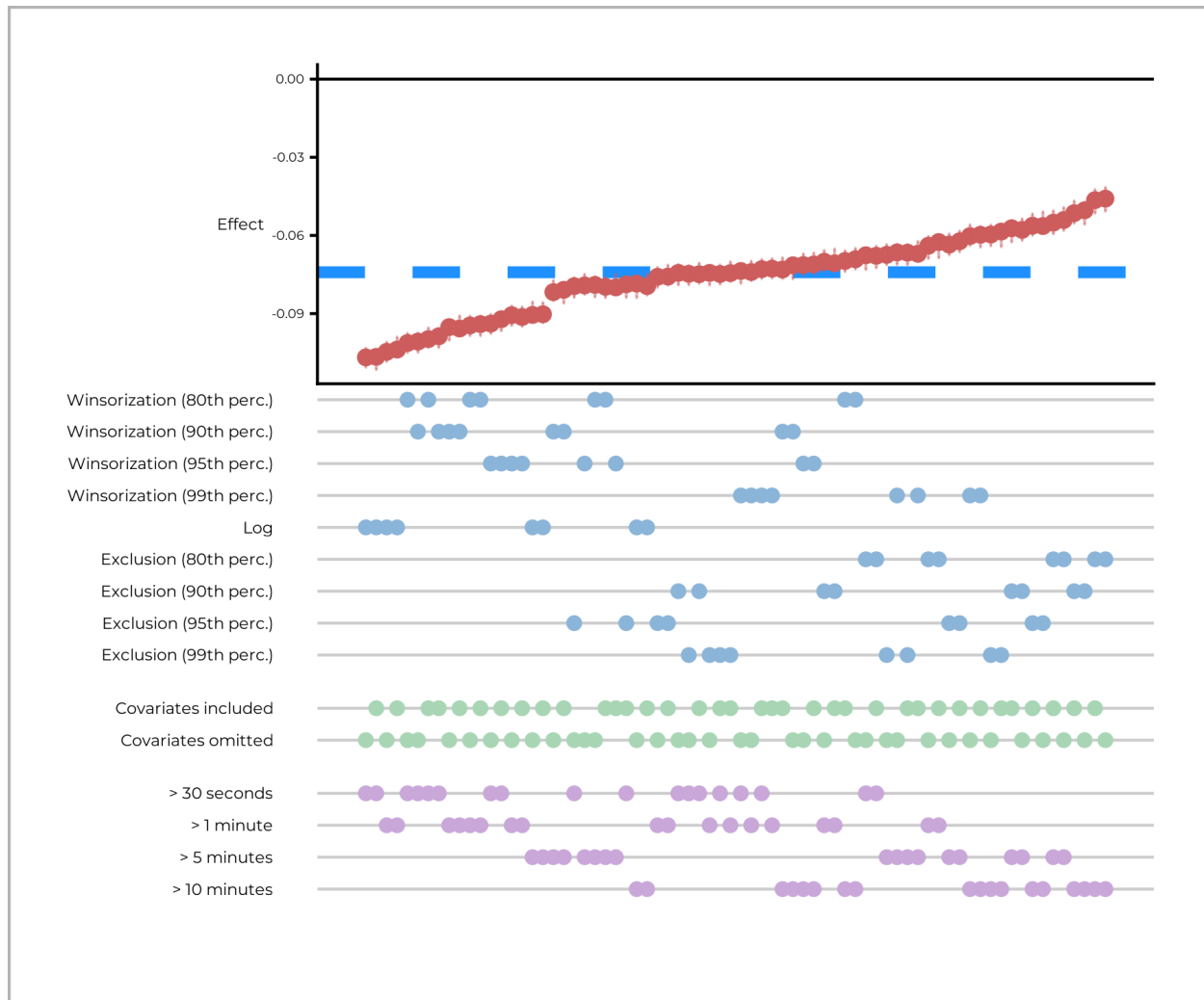
Figure S7 | Math Students – Multiverse Analyses – Effect of Load Times on Assignment Scores



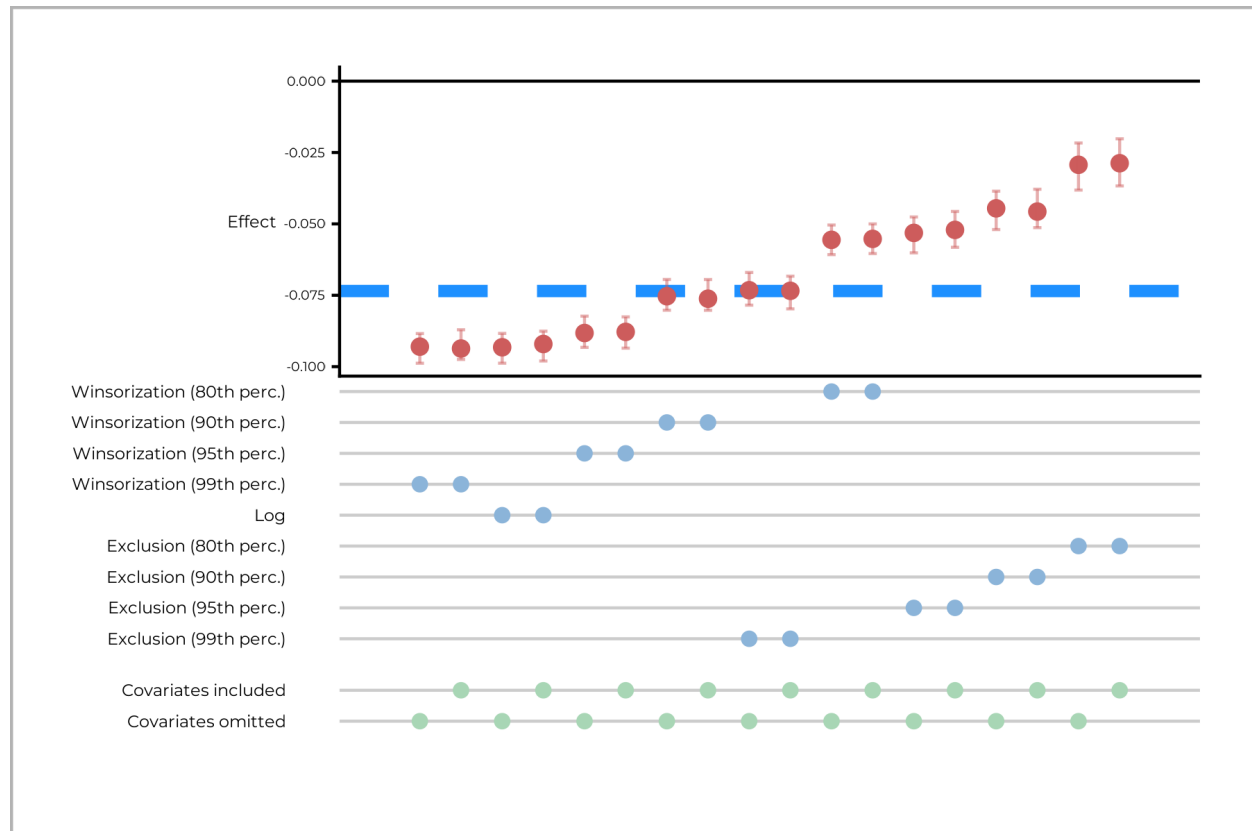
Note. This figure presents results from multiverse analyses examining the relationship between load times and math assignment scores among math students. Each column or vertical slice represents a single model. The bottom panel delineates the model specification, and the top panel plots each model's corresponding effect size (standardized regression coefficient), with points representing the effect size and error bars representing 95% confidence intervals. Red points indicate statistically significant harmful effects, gray points represent non-significant effects, and dark green points represent statistically significant beneficial effects. The dashed blue line represents the median effect size. The multiverse dimensions plotted include: different outlier handling methods (none, 95th- or 99th-percentile winsorization, 95th- or 99th-percentile exclusion, log+1 transformation); covariate sets included (none, time-varying, time-varying and school-level covariates); and whether an indicator for school-hours is included or omitted.

Figure S8 | Aspiring Coders – Descriptive Statistics

Note. Each x-axis tick represents a single one-sec bin (e.g., “0” represents load times between 0 and 1 second). Line height represents the average of each outcome within each bin, and ribbons represent the 95% confidence interval. These descriptive statistics do not adjust for clustering or covariates; see Table 2 (main text) for model-based estimates that account for these. $N = 181,790$ visits (for all outcomes except focus, where $N = 111,572$ visits).

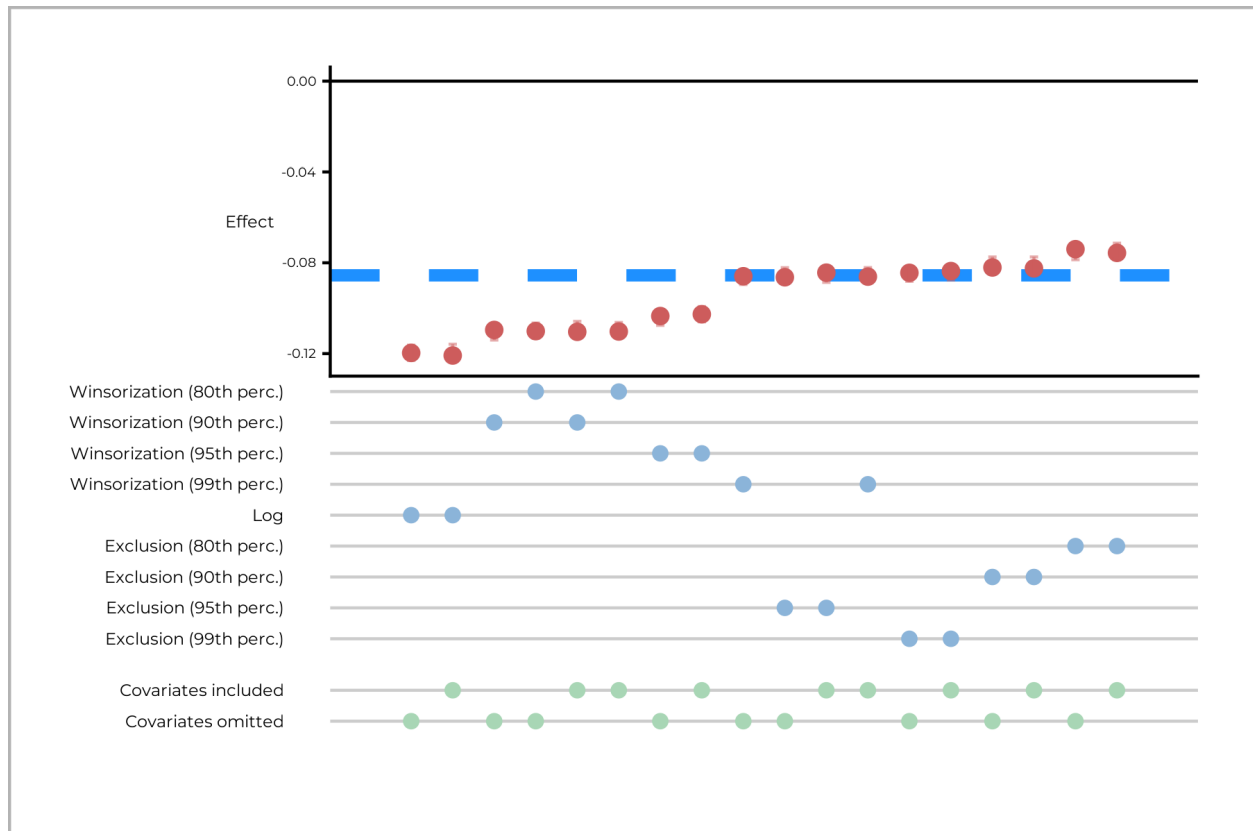
Figure S9 | Aspiring Coders – Multiverse Analyses – Effect of Load Times on Engagement

Note. This figure presents results from multiverse analyses examining the relationship between load times and engagement (whether a user engaged with the website over one minute) in a given visit. Each column or vertical slice represents a single model. The bottom panel delineates the model specification, and the top panel plots each model's corresponding effect size (standardized regression coefficient), with points representing the effect size and error bars representing 95% confidence intervals. Red points indicate statistically significant effects. The dashed blue line represents the median effect size. The multiverse dimensions plotted include: different outlier handling methods (80th-, 90th-, 95th-, or 99th-percentile winsorization; 80th-, 90th-, 95th-, or 99th-percentile exclusion; log+1 transformation) and whether covariates (time of day as well as day of week) were included in the models.

Figure S10 | Aspiring Coders – Multiverse Analyses – Effect of Load Times on Focus

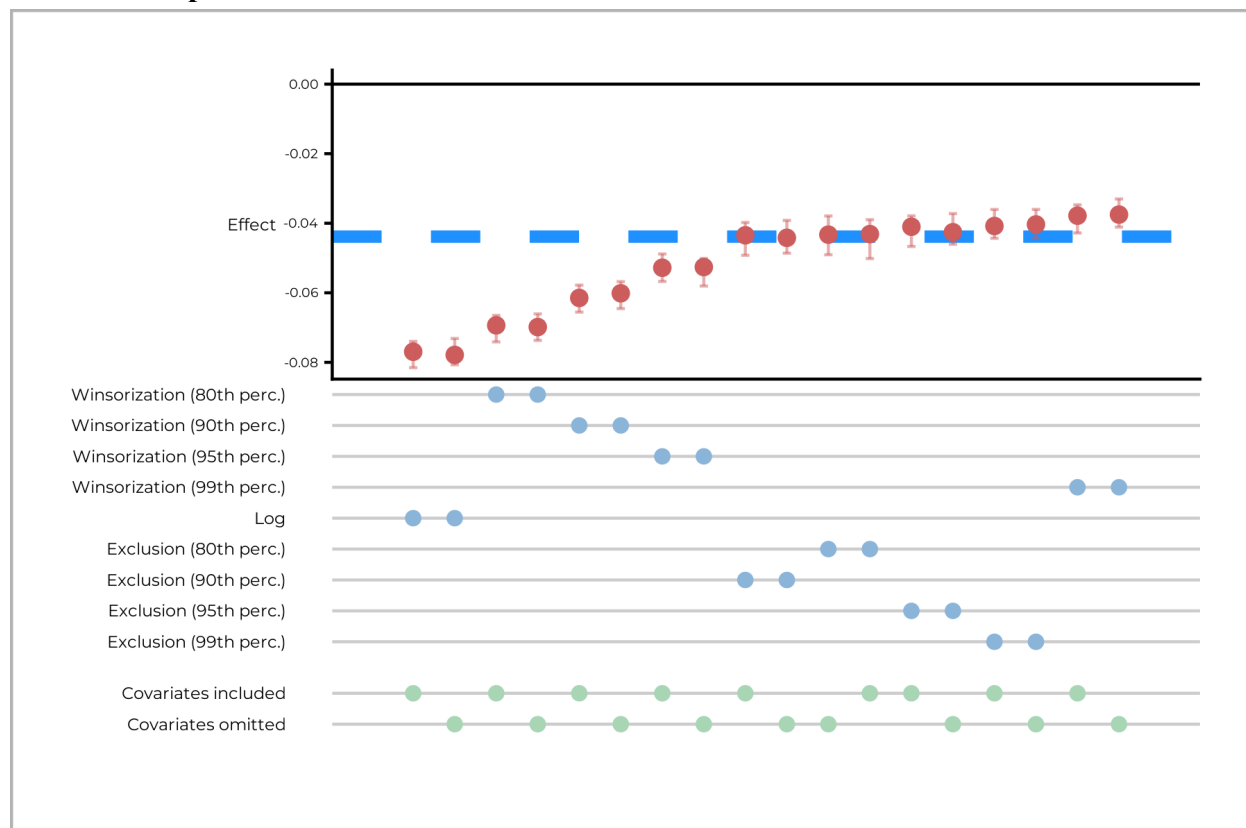
Note. This figure presents results from multiverse analyses examining the relationship between load times and focus in a given visit, with focus measured as active time (i.e., time when freeCodeCamp was in focus on a user's screen) divided by total visit length. Each column or vertical slice represents a single model. The bottom panel delineates the model specification, and the top panel plots each model's corresponding effect size (standardized regression coefficient), with points representing the effect size and error bars representing 95% confidence intervals. Red points indicate statistically significant effects. The dashed blue line represents the median effect size. The multiverse dimensions plotted include: different outlier handling methods (80th-, 90th-, 95th-, or 99th-percentile winsorization; 80th-, 90th-, 95th-, or 99th-percentile exclusion; log+1 transformation) and whether covariates (time of day as well as day of week) were included in the models.

Figure S11 | Aspiring Coders – Multiverse Analyses – Effect of Load Times on the Likelihood of Visiting more than One Page



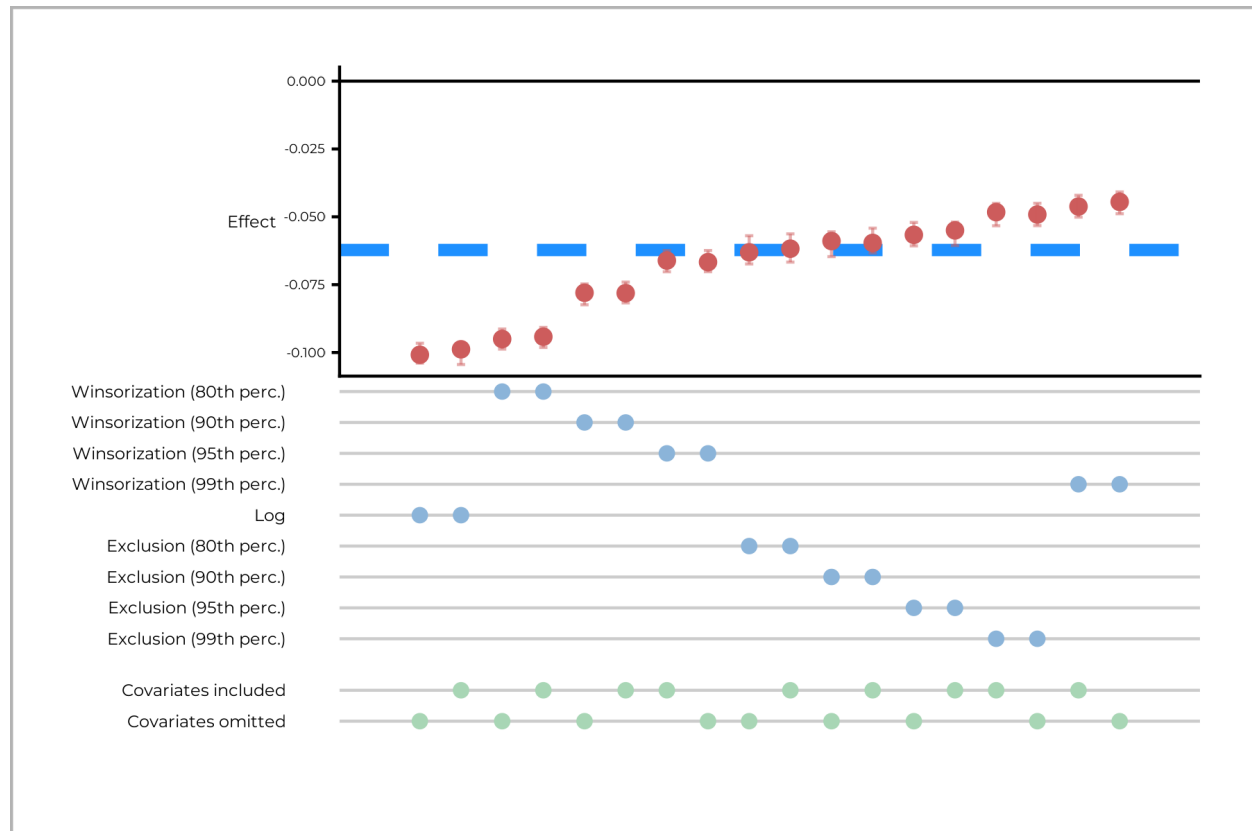
Note. This figure presents results from multiverse analyses examining the relationship between load times and the likelihood of a user visiting more than one page in a given visit. Each column or vertical slice represents a single model. The bottom panel delineates the model specification, and the top panel plots each model's corresponding effect size (standardized regression coefficient), with points representing the effect size and error bars representing 95% confidence intervals. Red points indicate statistically significant effects. The dashed blue line represents the median effect size. The multiverse dimensions plotted include: different outlier handling methods (80th-, 90th-, 95th-, or 99th-percentile winsorization; 80th-, 90th-, 95th-, or 99th-percentile exclusion; log+1 transformation) and whether covariates (time of day as well as day of week) were included in the models.

Figure S12 | Aspiring Coders – Multiverse Analyses – Effect of Load Times on Number of Lessons Completed

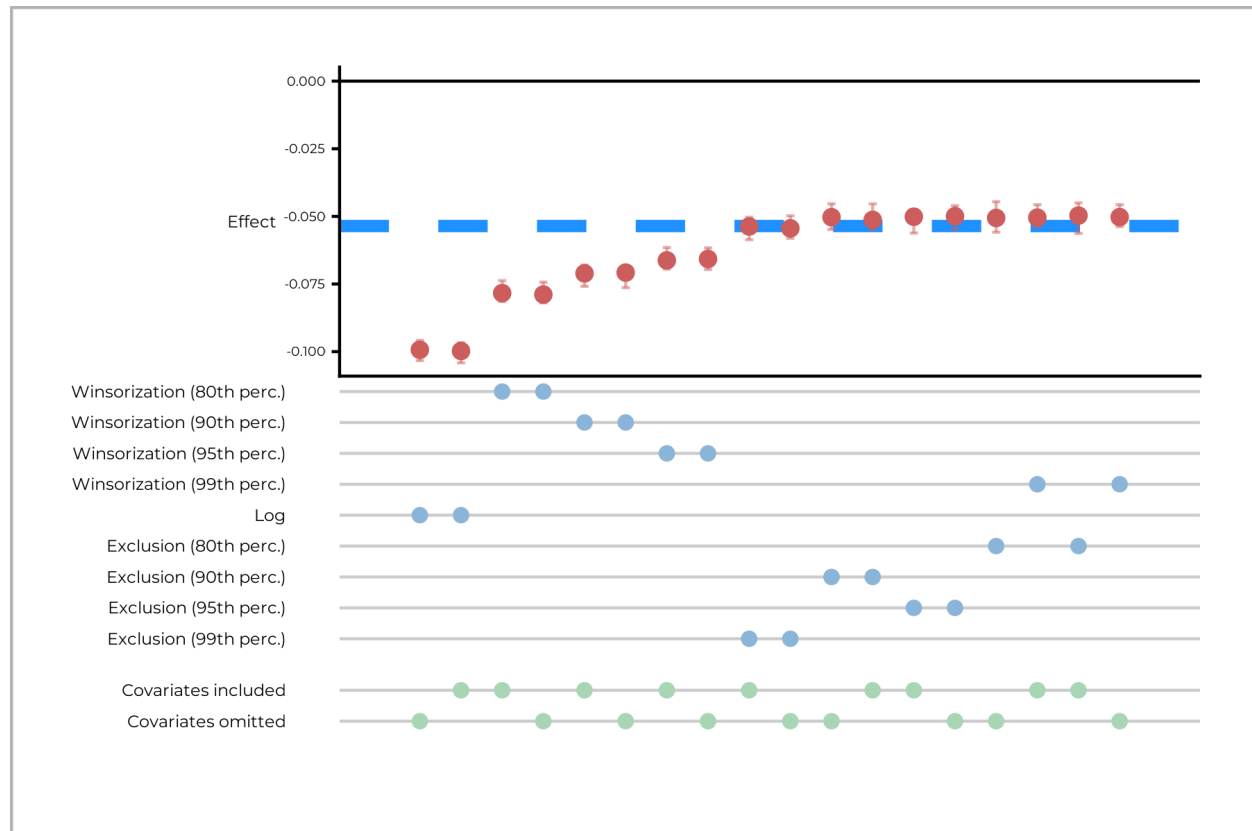


Note. This figure presents results from multiverse analyses examining the relationship between load times and the number of lessons a user completed in a given visit. Each column or vertical slice represents a single model. The bottom panel delineates the model specification, and the top panel plots each model's corresponding effect size (standardized regression coefficient), with points representing the effect size and error bars representing 95% confidence intervals. Red points indicate statistically significant effects. The dashed blue line represents the median effect size. The multiverse dimensions plotted include: different outlier handling methods (80th-, 90th-, 95th-, or 99th-percentile winsorization; 80th-, 90th-, 95th-, or 99th-percentile exclusion; log+1 transformation) and whether covariates (time of day as well as day of week) were included in the models.

Figure S13 | Aspiring Coders – Multiverse Analyses – Effect of Load Times on Number of Pages Visited

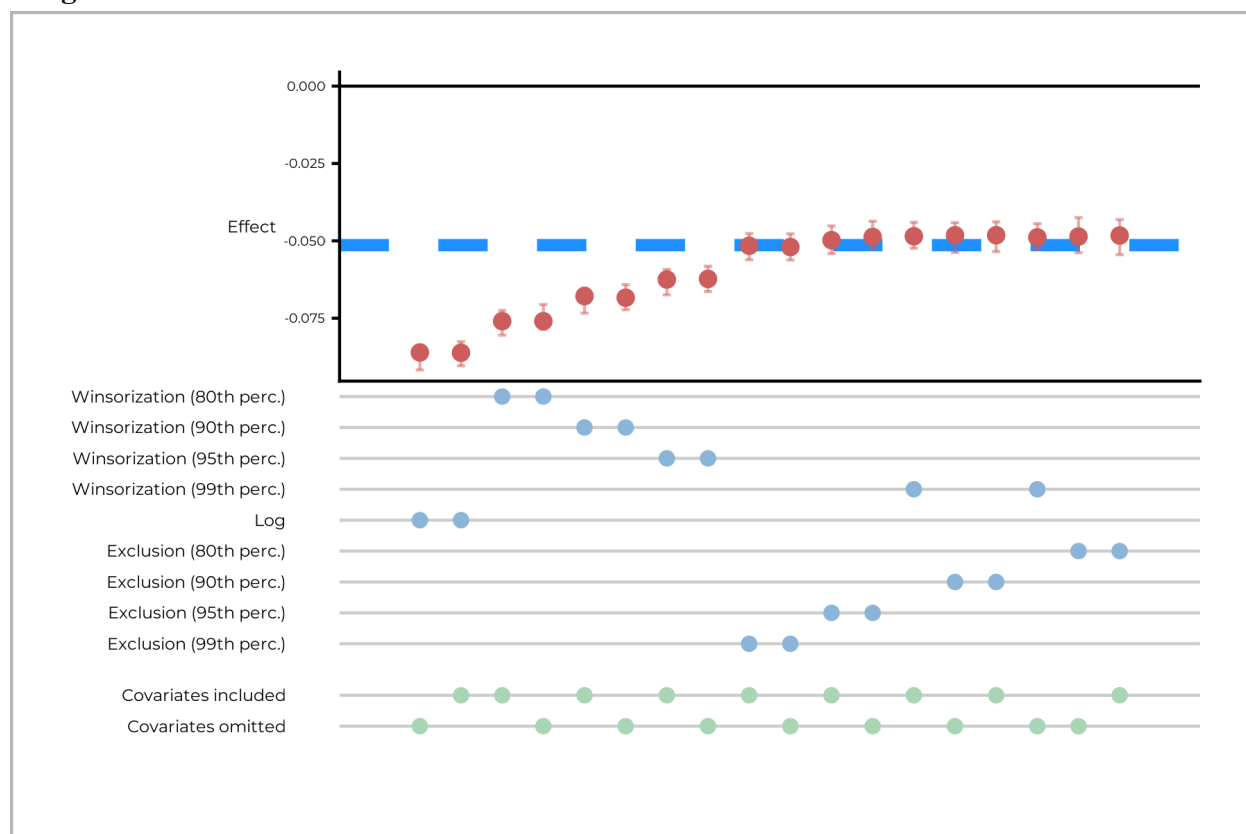


Note. This figure presents results from multiverse analyses examining the relationship between load times and the number of pages a user visited in a given visit. Each column or vertical slice represents a single model. The bottom panel delineates the model specification, and the top panel plots each model's corresponding effect size (standardized regression coefficient), with points representing the effect size and error bars representing 95% confidence intervals. Red points indicate statistically significant effects. The dashed blue line represents the median effect size. The multiverse dimensions plotted include: different outlier handling methods (80th-, 90th-, 95th-, or 99th-percentile winsorization; 80th-, 90th-, 95th-, or 99th-percentile exclusion; log+1 transformation) and whether covariates (time of day as well as day of week) were included in the models.

Figure S14 | Aspiring Coders – Multiverse Analyses – Effect of Load Times on Active Time

Note. This figure presents results from multiverse analyses examining the relationship between load times and the active time (in minutes) they spent engaging with the website in a given visit, as measured as time during which the freeCodeCamp tab was in focus on their screen. Each column or vertical slice represents a single model. The bottom panel delineates the model specification, and the top panel plots each model's corresponding effect size (standardized regression coefficient), with points representing the effect size and error bars representing 95% confidence intervals. Red points indicate statistically significant effects. The dashed blue line represents the median effect size. The multiverse dimensions plotted include: different outlier handling methods (80th-, 90th-, 95th-, or 99th-percentile winsorization; 80th-, 90th-, 95th-, or 99th-percentile exclusion; log+1 transformation) and whether covariates (time of day as well as day of week) were included in the models.

Figure S15 | Aspiring Coders – Multiverse Analyses – Effect of Load Times on Total Visit Length



Note. This figure presents results from multiverse analyses examining the relationship between load times and total visit length. Each column or vertical slice represents a single model. The bottom panel delineates the model specification, and the top panel plots each model's corresponding effect size (standardized regression coefficient), with points representing the effect size and error bars representing 95% confidence intervals. Red points indicate statistically significant effects. The dashed blue line represents the median effect size. The multiverse dimensions plotted include: different outlier handling methods (80th-, 90th-, 95th-, or 99th-percentile winsorization; 80th-, 90th-, 95th-, or 99th-percentile exclusion; log+1 transformation) and whether covariates (time of day as well as day of week) were included in the models.