

Online Supplement to “How Effective Is Subsidizing Access to Broader Digital Educational Content? Evidence from a Large-Scale Field Study on a Reading App”

Jingcun Cao, Xiru Pan, Catherine Tucker, Yifei Wang

OA1. Appendix to the Main Effect

OA1.1. Alternative Ways of Clustering Standard Errors

Table OA1 reports results with standard errors clustered at the treatment-cohort level.

Table OA1 Short-term Effect of Digital Education Content Subsidy on Children's Reading Time (Clustered SEs)

	Full Sample	No Short Sessions	No Long Sessions	Alternative DV	Including Missing Cities	First Day Active User
	(1) Reading Time	(2) Reading Time	(3) Reading Time	(4) Number of Materials	(5) Reading Time	(6) Reading Time
FreeMembership	4.658*** (0.167)	4.504*** (0.164)	4.649*** (0.166)	0.653*** (0.015)	4.718*** (0.170)	10.473*** (1.683)
Observations	361,015	361,015	361,015	361,015	388,752	4,837
Mean Dep Var	2.645	2.564	2.642	0.393	2.691	6.978

Notes. The short term is defined as the observation period of seven days after the treatment. Standard error clustered at treatment cohort level. The sample includes treated users who were first observed at least seven days before the treatment and who were treated at least seven days (we excluded users whose post-treatment period is less than 7 days) from September 2019 to the end of March 2020, and all users who were treated in April 2020 that essentially served as the “never-treated” control group in our analysis. All columns in this table use a seven-day window before and after the treatment for each user. The data is aggregated at user-day level. * p<0.10 ** p<0.05 *** p<0.01.

Table OA2 Longer-term Effect of Digital Education Content Subsidy on Children's Reading Time (Clustered SEs)

	Bottom GRP	Middle GRP	Top GRP
	(1) Reading Time	(2) Reading Time	(3) Reading Time
FreeMembership	2.033*** (0.259)	2.514*** (0.211)	2.637*** (0.229)
Observations	352,027	349,128	360,732
Mean Dep Var	3.146	2.983	2.864

Notes. The long term is defined as the observation period of 60 days after the treatment. Standard error clustered at treatment cohort level. The sample includes treated users who were first observed at least seven days before the treatment and who were treated at least 60 days from September 2019 to the end of March 2020, and all users who were treated in April 2020 that essentially served as the “never-treated” control group in our analysis. All columns in this table use a seven-day window before and 60-day window after the treatment for each user. We also exclude users with missing values on city GRP. The data is aggregated at user-day level. * p<0.10, ** p<0.05, *** p<0.01

OA1.2. Testing the Parallel Trends Assumption

Table OA3 presents the results for testing the parallel trends assumption.

Table OA3 Checking for Pre-treatment Trends

	Main Result	Curriculum Related		Curriculum Unrelated	Reading Patterns	
	(1) Reading Time	(2) Reading Time Humanities	(3) Reading Time Science	(4) Reading Time	(5) Time per Session	(6) Number of Long-reading Sessions
Estimated effect, pre-treatment	−0.125 (0.105)	−0.139 (0.091)	−0.019 (0.022)	0.033 (0.033)	−0.082 (0.052)	−0.002 (0.002)
Estimated effect, post-treatment	4.658*** (0.167)	3.758*** (0.152)	0.410*** (0.033)	0.492*** (0.057)	2.663*** (0.083)	0.076*** (0.003)
Observations	361,015	361,015	361,015	361,015	361,015	361,015
Mean Dep Var	2.645	2.188	0.208	0.249	1.531	0.041

Notes. Bootstrap standard errors are reported. The sample includes treated users who were first observed at least seven days before the treatment and who were treated at least seven days (we excluded users whose post-treatment period is less than 7 days) from September 2019 to the end of March 2020, and all users who were treated in April 2020 that essentially served as the “never-treated” control group in our analysis. All columns in this table use a seven-day window before and after the treatment for each user. The data is aggregated at user-day level. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

OA1.3. Churn Rate and Control Group Stability

Figure OA1 plots the daily average reading time of control users by event time. Reading time remains stable throughout, with no discernible upward trend. This suggests that the composition of the control group does not systematically shift toward more engaged readers over the post-treatment horizon. Figure OA2 plots the pre-treatment churn rates of the treated and never-treated groups from 40 to 14 days before treatment, where churn is defined as recording no reading activity in the subsequent seven days.¹ Together, these figures provide supporting evidence that the control group does not experience systematic compositional changes over time.

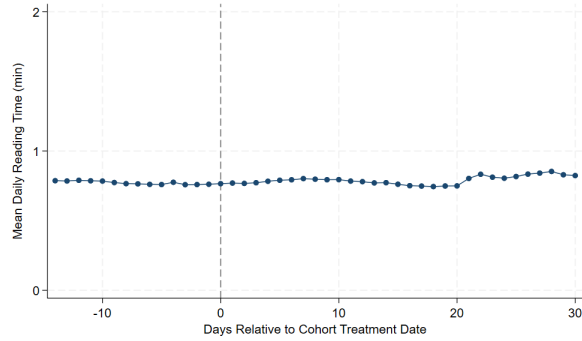
OA1.4. Alternative DiD Specification

We implement two alternative specifications to demonstrate robustness. First, following Wooldridge (2021), we restructure TWFE as a pooled OLS with cohort fixed effects, allowing treatment effects to vary across cohorts and time. Column (1) of Table OA4 confirms free VIP assignment increases reading time by 4.946 minutes. Second, we implement the stacked DiD method (Wing et al. 2024), which constructs a separate dataset for each treatment cohort and stacks them for joint estimation. Both alternatives yield results consistent with our main specification.

The stacked DiD method could help us address the concern that users needed to open the app to receive the free membership after receiving a push notification from the firm. The set of treated users who opened the app to receive the free membership might be substantially different from the not-yet-treated users who were not necessarily active on a focal date of VIP activation (day 0), which might introduce bias to our estimate of the effect of the VIP membership on day 0. We conducted a robustness check based on the stacked DiD method to address potential activity bias in our estimates.

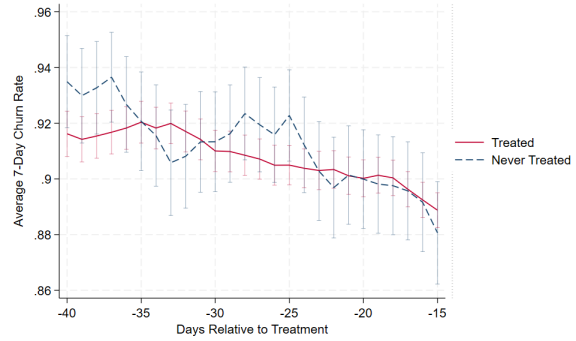
¹ We focus on the pre-treatment window from 40 to 14 days before the treatment date because treated users were selected based on their activity in the seven days immediately preceding treatment, as discussed in the paper and explicitly accounted for in our analysis. As a result, churn rates closer to the treatment date could be mechanically affected by the firm’s selection rule, thereby confounding the comparison between treated and never-treated users.

Figure OA1 Average Reading Time of Control Users by Event Time



Notes. This figure plots the daily average reading time of control users from $t = -14$ to $t = 30$ relative to the cohort treatment date. Reading time is in minutes.

Figure OA2 Pre-treatment Churn Rates of Treated and Never-treated Users



Notes. This figure plots the daily churn rate of treated and never-treated users from 40 to 14 days before treatment. Churn on a given day is defined as recording no reading activity in the subsequent seven days. We use never-treated users (those not treated in our observation period) rather than not-yet-treated users as the comparison group in this figure. Because treatment is assigned conditional on recent activity, a user's proximity to their own treatment date is informative about their activity and churn rates. When a not-yet-treated user is placed on an already-treated cohort's timeline, day $-t$ refers to t days before that cohort's treatment date, not t days before the user's own. As a result, this user is typically still far from their own treatment date, whereas the treated users they are compared to are close to theirs. This makes it difficult to interpret a direct comparison of raw churn rates between the treated and not-yet-treated users at a given event time for the treated users without further regression adjustments.

In Column (3) of Table OA4, for each cohort, we only include not-yet-treated users who were active on the date of VIP activation of the treated users (day 0). This approach ensures that both treated users and control users are active on the treatment date, thereby enhancing their comparability. The estimation results we obtained are also consistent with our main results.

Table OA4 Alternative Specifications for Measuring the Short-term Effect of Digital Education Content Subsidy

	Wooldridge DiD	Stacked DiD with Full Sample	Stacked DiD with Active Control Users
	(1) Reading Time	(2) Reading Time	(3) Reading Time
Main ATT	4.946*** (0.133)		
Treat $\times$ Post		5.089*** (0.134)	4.606*** (0.321)
Cohort-User FE		Yes	Yes
Cohort-Date FE		Yes	Yes
Observations	361,011	3,221,714	404,081
R^2	0.039	0.351	0.401

Notes. The short term is defined as the observation period of 7 days after the treatment. The sample includes treated users who were first observed at least 7 days before the treatment and who were treated at least 7 days (We excluded users whose post-treatment period is less than 7 days) from September 2019 to the end of March 2020, and all users who were treated in April 2020 that essentially served as the control group in our analysis. In Column (1), standard errors clustered at user level. In Columns (2)–(3) standard errors clustered at cohort-user level. The data is aggregated at user-day level. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

OA2. Appendix to the Heterogeneous Effects

OA2.1. Alternative Measures of Socioeconomic Status (SES)

Table OA5 Short-term Effect by Socioeconomic Status (Alternative SES Measures)

	Educational Investment			Teaching Force			Library Resources		
	(1) Bottom GRP	(2) Middle GRP	(3) Top GRP	(4) Bottom GRP	(5) Middle GRP	(6) Top GRP	(7) Bottom GRP	(8) Middle GRP	(9) Top GRP
FreeMembership	4.961*** (0.332)	4.732*** (0.246)	4.311*** (0.297)	4.728*** (0.288)	5.106*** (0.301)	4.174*** (0.254)	4.748*** (0.312)	5.286*** (0.296)	4.100*** (0.291)
Observations	116,200	121,670	123,145	116,391	115,181	129,443	111,146	110,658	118,339
Mean Dep Var	3.218	2.592	2.157	3.093	2.604	2.279	3.191	2.777	2.177

Notes. The short-term is defined as the observation period of seven days after the treatment. Bootstrap standard errors are reported. The outcome variable of this table is the daily reading time of each user, measured in minutes. The sample includes treated and control users who were first observed at least seven days before the treatment and who were treated at least seven days prior to the end of our sample period. All columns in this table use a seven-day window before and after the treatment for each user. Columns (1) to (3) present the results for users in the bottom, middle, and top terciles of city educational investment. Columns (4) to (6) present results for users in the bottom, middle, and top terciles of city-level counts of primary- and middle-school teachers. Columns (7) to (9) present results for users in the bottom, middle, and top terciles of city-level public library book holdings (measured by total number of books). The data is aggregated at user-day level. * p<0.10, ** p<0.05, *** p<0.01

Table OA6 Longer-term Effect by Socioeconomic Status (Alternative SES Measures)

	Educational Investment			Teaching Force			Library Resources		
	(1) Bottom GRP	(2) Middle GRP	(3) Top GRP	(4) Bottom GRP	(5) Middle GRP	(6) Top GRP	(7) Bottom GRP	(8) Middle GRP	(9) Top GRP
FreeMembership	2.000*** (0.299)	2.516*** (0.190)	2.658*** (0.238)	1.896*** (0.266)	2.646*** (0.224)	2.618*** (0.217)	1.934*** (0.261)	2.541*** (0.230)	2.635*** (0.250)
Observations	339,342	371,169	351,376	347,953	335,274	378,660	329,093	326,943	346,608
Mean Dep Var	3.138	2.957	2.902	3.039	3.023	2.935	3.042	2.978	3.015

Notes. The longer-term is defined as the observation period of sixty days after the treatment. Bootstrap standard errors are reported. The outcome variable of this table is the daily reading time of each user, measured in minutes. The sample includes treated and control users who were first observed at least seven days before the treatment and who were treated at least seven days prior to the end of our sample period. All columns in this table use a seven-day window before and sixty-day window after the treatment for each user. Columns (1) to (3) present the results for users in the bottom, middle, and top terciles of city educational investment. Columns (4) to (6) present results for users in the bottom, middle, and top terciles of city-level counts of primary- and middle-school teachers. Columns (7) to (9) present results for users in the bottom, middle, and top terciles of city-level public library book holdings (measured by total number of books). The data is aggregated at user-day level. * p<0.10, ** p<0.05, *** p<0.01

OA2.2. Alternative Model Specification

Our previous analysis suggests the treatment effect varies across users with different socioeconomic backgrounds in both the short and the long term. In this section, we conduct robustness checks to formally assess whether these differences are statistically significant. Since our main estimator, the CSDID model, does not support interaction terms, we estimate a two-way fixed effects model that includes interaction terms between the treatment indicator and various indicators of socioeconomic status. We estimate the following specification:

$$y_{it} = \alpha + \beta_1 Treated_{it} + \beta_2 Treated_{it} \times X_i + \gamma_i + \lambda_t + \epsilon_{it} \quad (1)$$

where y_{it} stands for the reading time (min) of user i at day t . $Treated_{it}$ is an indicator for treatment status which equals 1 if user i started the free VIP membership at day t and 0 otherwise. X_i represents

the socioeconomic status indicators used in the main analysis, including city-level GRP, educational investment, teaching force, and library resources. γ_i and λ_t represent individual fixed effects and day fixed effects, respectively. β_1 captures the average treatment effect, whereas β_2 assesses whether the treatment effect varies by socioeconomic status.

Table OA7 reports the results for the immediate (short-term) effect. In Column (1), we use city-level GRP as a proxy for household income. The interaction term, $Treated \times GRP$, is negative and statistically significant, suggesting that the treatment effect is smaller in more economically developed cities. Columns (2)–(4) present results using alternative indicators of socioeconomic status—educational investment, teaching force, and library resources. Across all specifications, the interaction terms remain negative and statistically significant, supporting our conclusion in Table ?? that, in the short term, the treatment has a greater impact on children from less developed cities.

Table OA8 presents the longer-term results using the same TWFE specification. Column (1) uses city-level GRP as a proxy for annual household income, while Columns (2)–(4) incorporate the same alternative socioeconomic indicators as in the short-term analysis—namely educational investment, teaching force, and library resources. In all cases, the interaction terms are positive and statistically significant, suggesting that in the long run, the treatment effect is larger for users from more developed regions. These findings provide formal statistical support for the presence of heterogeneous treatment effects and align with the patterns shown in Table 5.

Table OA7 Short-term Effect of the Free Membership (TWFE)				
	(1) Reading Time	(2) Reading Time	(3) Reading Time	(4) Reading Time
Treated	5.055*** (0.296)	5.052*** (0.292)	5.237*** (0.353)	4.961*** (0.296)
Treated $\times$ GRP	−0.003*** (0.001)			
Treated $\times$ Educational investment		−0.102*** (0.035)		
Treated $\times$ Teaching force			−7.089** (3.257)	
Treated $\times$ Library resources				−0.002*** (0.001)
Observations	360,959	360,959	360,959	340,088
R^2	0.367	0.367	0.367	0.371

Notes. The outcome variable of this table is the daily reading time of each user, measured in minutes. The short term is defined as the observation period of seven days after the treatment. We control for whether an observation falls within the seven-day period preceding a user's VIP activation to account for potential pre-treatment activity bias in our analysis. The sample includes treated users who were first observed at least seven days before the treatment and who were treated at least seven days (we excluded users whose post-treatment period is less than 7 days) from September 2019 to the end of March 2020, and all users who were treated in April 2020 that essentially served as the 'never-treated' control group in our analysis. All columns in this table use a seven-day window before and after the treatment for each user. GRP, Educational Investment, and Teaching Force are rescaled to units of millions, and Library Resources are rescaled to units of thousands for ease of interpretation. Standard error clustered at user level. The data is aggregated at user-day level. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

Table OA8 Longer-term Effect of the Free Membership (TWFE)

	(1) Reading Time	(2) Reading Time	(3) Reading Time	(4) Reading Time
Treated	4.496*** (0.306)	4.539*** (0.302)	4.437*** (0.334)	4.532*** (0.309)
Treated $\times$ GRP	0.002** (0.001)			
Treated $\times$ Educational investment		0.067** (0.029)		
Treated $\times$ Teaching force			4.374* (2.430)	
Treated $\times$ Library resources				0.001** (0.001)
Observations	1,061,831	1,061,831	1,061,831	1,002,589
R^2	0.285	0.285	0.285	0.287

Notes. The long term is defined as the observation period of 60 days after the treatment. Standard error clustered at user level. We control for whether an observation falls within the seven-day period preceding a user's VIP activation to account for potential pre-treatment activity bias in our analysis. The sample includes treated users who were first observed at least seven days before the treatment and who were treated at least 60 days from September 2019 to the end of March 2020, and all users who were treated in April 2020 that essentially served as the "never-treated" control group in our analysis. All columns in this table use a seven-day window before and 60-day window after the treatment for each user. GRP, Educational Investment, and Teaching Force are rescaled to units of millions, and Library Resources are rescaled to units of thousands for ease of interpretation. The data is aggregated at user-day level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

OA2.3. Alternative Outcome Variable

Table OA9 Short-term and Longer-term Effects by Socioeconomic Status (Log Reading Time)

Panel A. Short-term Effect (7 Days)			
	City GRP		
	(1) Bottom GRP	(2) Middle GRP	(3) Top GRP
FreeMembership	0.426*** (0.014)	0.429*** (0.012)	0.393*** (0.013)
Observations	120,681	113,869	126,465
Mean Dep Var	0.267	0.228	0.190
Panel B. Longer-term Effect (60 Days)			
	City GRP		
	(1) Bottom GRP	(2) Middle GRP	(3) Top GRP
FreeMembership	0.179*** (0.013)	0.246*** (0.012)	0.248*** (0.011)
Observations	352,027	349,128	360,732
Mean Dep Var	0.265	0.270	0.256

Notes. The outcome variable in this table is the log-transformed daily reading time for each user. Panel A reports short-term effects, defined as a seven-day observation window following treatment. The sample includes users who were observed for at least seven days before and after treatment, between September 2019 and March 2020. Panel B reports longer-term effects, based on a 60-day post-treatment window. This sample includes users observed at least seven days prior to and 60 days after treatment during the same period. Bootstrap standard errors are reported. Users with missing data on GRP or education are excluded. The unit of analysis is the user-day. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

OA2.4. Robustness Checks for Heterogeneous Effects by Level of Prior Engagement.

OA2.4.1. Alternative Specification To formally test whether the treatment effects differ significantly between light and heavy users, we re-estimated the model using a two-way fixed effects (TWFE) specification that includes an interaction term between the treatment indicator and the prior engagement indicator. The results are presented in Table OA10. Panel A shows that in the short run, both light and heavy users significantly increased their reading time following the treatment, with heavy users experiencing a larger short-term benefit than light users. Panel B shows that in the long run, the interaction term is positive and statistically significant, confirming that light users sustain significantly larger longer-term benefits from the free membership compared to heavy users. These results are consistent with our main findings in Section 4.4.

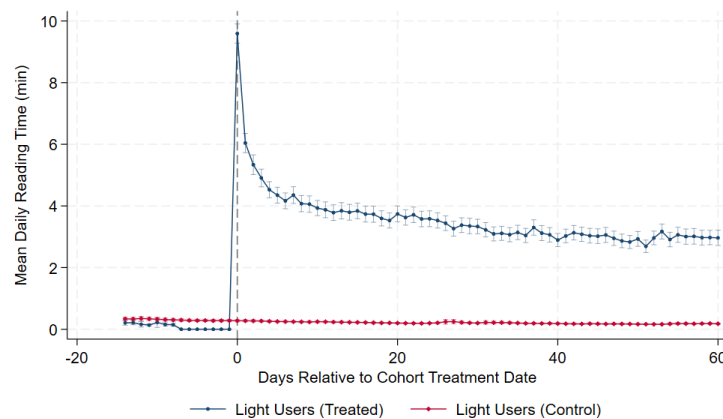
OA2.4.2. Addressing Mean Reversion as an Alternative Explanation One concern is that the larger longer-term treatment effect for light users reflects mean reversion rather than a genuine treatment response. Light users are defined based on zero reading time in the single week prior to treatment, which could reflect a transient dip in availability rather than a stable trait. If so, their reading time would naturally rebound in subsequent weeks even absent treatment, simply reverting to a higher underlying mean. This concern is amplified by the fact that treatment effects are reported in percentage terms, which mechanically enlarges gains for users starting near zero.

Table OA10 Short-term and Longer-term Effects by Users' Prior Engagement (TWFE)

	Short-term Effect	Long-term Effect
	(1) Reading Time	(2) Reading Time
FreeMembership $\times$ Light user	-2.193*** (0.269)	0.868*** (0.224)
FreeMembership	5.703*** (0.290)	3.172*** (0.263)
Observations	360,959	1,061,831

Notes. The outcome variable of this table is the daily reading time of each user, measured in minutes. Column (1) reports the short-term effect, defined using a seven-day observation window following treatment. Column (2) reports longer-term effects, based on a 60-day post-treatment window. Standard error clustered at user level. The data is aggregated at user-day level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We address this concern by providing graphical evidence on the pre- and post-treatment trajectory of light users' reading time. Figure OA3 plots the average daily reading time of treated and control light users over an extended window spanning 14 days before through 60 days after treatment. Two features speak against mean reversion. First, control light users do not exhibit a natural upward drift in reading activity absent treatment. Their reading time remains stable both before and after the cohort treatment date, providing no evidence of a temporary negative shock that would mechanically reverse. Second, for treated light users, while reading time rises sharply at treatment onset and gradually declines thereafter, it stabilizes at approximately 3 minutes per day by day 60 which is much larger than the pre-treatment baseline. If the post-treatment trajectory reflected a simple return to the true mean, we would expect reading time to converge toward pre-treatment levels. The sustained elevation well above baseline is inconsistent with this prediction.

Figure OA3 Average Daily Reading Time of Treated and Control Light Users

Notes. This figure plots the daily average reading time of light users from $t = -14$ to $t = 60$ relative to the cohort treatment date, separately for treated and control users. Reading time is measured in minutes.

OA3. Additional Evidence for Mechanism

OA3.1. Supporting Evidence for Mechanism

Table OA11 Unfit Materials Disproportionately Harm Longer-term Reading for Poor Children

	Bottom GRP		Top GRP	
	(1) Long-term Reading	(2) Long-term Reading	(3) Long-term Reading	(4) Long-term Reading
Week1Reading	3.024*** (0.179)	3.026*** (0.179)	2.815*** (0.286)	2.814*** (0.286)
Week1Reading, BadFit	−0.846** (0.343)		−0.268 (0.491)	
Week1Reading, TooEasy		−0.954** (0.467)		−0.381 (0.675)
Week1Reading, TooHard		−0.764* (0.405)		−0.201 (0.549)
Gender Fixed Effects	Yes	Yes	Yes	Yes
Age Fixed Effects	Yes	Yes	Yes	Yes
Grade Fixed Effects	Yes	Yes	Yes	Yes
Observations	10,954	10,954	10,614	10,614
Mean Dep Var	203.700	203.700	226.640	226.640

Notes. The outcome variable of this table is the daily reading time of each user, measured in minutes. All columns in this table use a 7-day window before and 60-day window after the treatment for each treated user. Standard errors are clustered at user level. The data is aggregated at user-day level. * p<0.10, ** p<0.05, *** p<0.01.

Table OA12 Variation in Different Content Type Consumption across Rich and Poor Cities over Time

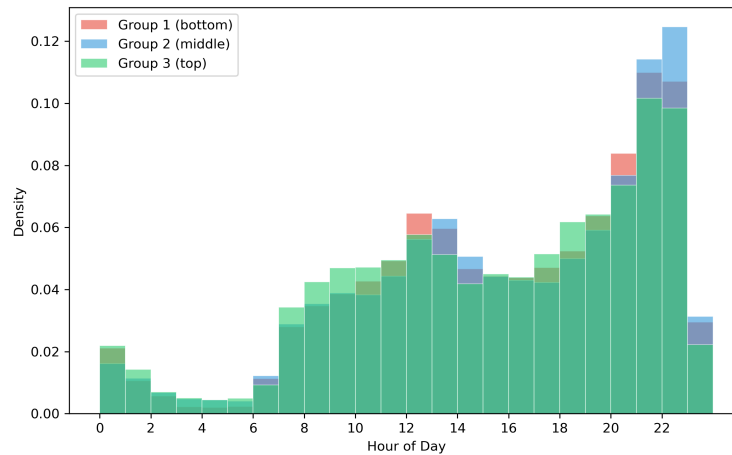
	Panel A. Audiobooks				Panel B. Ebooks		
	Bottom GRP	Middle GRP	Top GRP		Bottom GRP	Middle GRP	Top GRP
	(1) Time per Session	(2) Time per Session	(3) Time per Session		(1) Time per Session	(2) Time per Session	(3) Time per Session
Pre_avg	−0.001 (0.001)	−0.001 (0.001)	0.000 (0.001)	Pre_avg	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Day = 0	0.098*** (0.004)	0.084*** (0.004)	0.088*** (0.004)	Day = 0	0.009*** (0.001)	0.004*** (0.001)	0.004*** (0.001)
Day = 7	0.027*** (0.003)	0.029*** (0.003)	0.028*** (0.002)	Day = 7	0.004*** (0.001)	0.003*** (0.001)	0.003*** (0.001)
Day = 30	0.023*** (0.003)	0.022*** (0.002)	0.023*** (0.002)	Day = 30	0.001* (0.001)	0.001*** (0.000)	0.002*** (0.001)
Day = 60	0.009*** (0.003)	0.019*** (0.003)	0.023*** (0.003)	Day = 60	−0.000 (0.000)	0.002** (0.001)	0.001*** (0.000)
Observations	352,027	349,128	360,732	Observations	352027	349128	360732
Mean Dep Var	0.026	0.025	0.025	Mean Dep Var	0.002	0.001	0.001

Notes. Panel A reports dynamic effects for audiobooks. Panel B reports dynamic effects for ebooks. ‘Time per Session’ is defined as the average duration of each session. The long term is defined as the observation period of 60 days after the treatment. Bootstrap standard errors are reported. The sample includes treated users who were first observed at least seven days before the treatment and who were treated at least 60 days from September 2019 to the end of March 2020, and all users who were treated in April 2020 that essentially served as a “never-treated” control group in our analysis. All columns in this table use a seven-day window before and 60-day window after the treatment for each treated user. The data is aggregated at user-day level. * p<0.10, ** p<0.05, *** p<0.01

OA3.2. Empirical Evidence Against Device-Sharing as an Alternative Explanation

We present two pieces of empirical evidence that device-sharing is unlikely to be our underlying mechanism. First, we plot the distribution of reading activity across hours of the day for children from low-, middle-, and high-GRP cities (Figure OA4). If high-SES children read more because they have dedicated devices and therefore face less competition with parents during working hours, we would expect them to read relatively more during the day than low-SES children, who may only gain access to shared devices after their parents return home. Instead, children from all three groups exhibit a strikingly similar intraday distribution, with reading activity concentrated in the evening hours after 8 p.m. This pattern does not support a device-availability explanation. Second, we consulted with the platform manager, who has access to device-level information not available in our dataset. The manager confirmed that multi-device usage is relatively uncommon, with most users accessing the platform through a single device. This makes it unlikely that device-sharing across household members is a widespread phenomenon driving our findings.

Figure OA4 **Distribution of Reading Hour by Group**



Notes. This figure plots the distribution of reading activity by hour of day for children in low-, middle-, and high-GRP cities. The y-axis reports the density of reading sessions initiated at each hour.

OA4. Linking Reading Time to Educational Outcomes

Table OA13 Short-term and Longer-term Effects by Curriculum Relevance

Panel A. Short-term Effect (7 Days)						
	Curriculum Related			Curriculum Unrelated		
	Bottom GRP	Middle GRP	Top GRP	Bottom GRP	Middle GRP	Top GRP
	(1) Reading Time	(2) Reading Time	(3) Reading Time	(4) Reading Time	(5) Reading Time	(6) Reading Time
FreeMembership	4.479*** (0.250)	4.248*** (0.223)	4.165*** (0.248)	0.698*** (0.110)	0.410*** (0.067)	0.327*** (0.063)
Observations	121,756	112,794	126,465	121,756	112,794	126,465
Mean Dep Var	2.832	2.352	2.015	0.387	0.233	0.131
Panel B. Longer-term Effect (60 Days)						
	Curriculum Related			Curriculum Unrelated		
	Bottom GRP	Middle GRP	Top GRP	Bottom GRP	Middle GRP	Top GRP
	(1) Reading Time	(2) Reading Time	(3) Reading Time	(4) Reading Time	(5) Reading Time	(6) Reading Time
FreeMembership	1.815*** (0.250)	2.297*** (0.208)	2.471*** (0.222)	0.247*** (0.078)	0.221*** (0.038)	0.187*** (0.023)
Observations	355,728	345,427	360,732	355,728	345,427	360,732
Mean Dep Var	2.819	2.734	2.700	0.323	0.252	0.165

Notes. The outcome variable of this table is the daily reading time of each user, measured in minutes. Columns (1)–(3) present the results for curriculum-related content, and Columns (4)–(6) present the results for curriculum-unrelated content. Panel A reports short-term effects, defined as a seven-day observation window following treatment. Panel B reports longer-term effects, based on a 60-day post-treatment window. The data is aggregated at user-day level. Bootstrap standard errors are reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table OA14 Short-term and Longer-term Effects by Content Type

Panel A. Short-term Effect (7 Days)						
	Humanities			Science		
	(1) Bottom GRP	(2) Middle GRP	(3) Top GRP	(4) Bottom GRP	(5) Middle GRP	(6) Top GRP
	(1) Reading Time	(2) Reading Time	(3) Reading Time	(4) Reading Time	(5) Reading Time	(6) Reading Time
FreeMembership	3.957*** (0.267)	3.793*** (0.239)	3.552*** (0.285)	0.374*** (0.084)	0.465*** (0.049)	0.466*** (0.045)
Observations	121,756	112,794	126,465	121,756	112,794	126,465
Mean Dep Var	2.603	2.131	1.838	0.229	0.221	0.177

Panel B. Longer-term Effect (60 Days)						
	Humanities			Science		
	(1) Bottom GRP	(2) Middle GRP	(3) Top GRP	(4) Bottom GRP	(5) Middle GRP	(6) Top GRP
	(1) Reading Time	(2) Reading Time	(3) Reading Time	(4) Reading Time	(5) Reading Time	(6) Reading Time
FreeMembership	1.695*** (0.224)	2.027*** (0.205)	2.159*** (0.227)	0.112 (0.076)	0.292*** (0.031)	0.338*** (0.035)
Observations	355,728	345,427	360,732	355,728	345,427	360,732
Mean Dep Var	2.596	2.462	2.443	0.223	0.272	0.256

Notes. The outcome variable of this table is the daily reading time of each user, measured in minutes. Columns (1)–(3) present the results for humanities content, and Columns (4)–(6) present the results for science content. Panel A reports short-term effects, defined as a seven-day observation window following treatment. Panel B reports longer-term effects, based on a 60-day post-treatment window. The data is aggregated at user-day level. Bootstrap standard errors are reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

OA5. Educator's Survey

Sample Survey

**The survey was conducted in Chinese, with the English translation provided below.*

Please randomly select any five students from your class whose student ID's last digit is 1 to 5. Based on your observation of each student's performance in the past semester, as well as the interactions with you and his or her parent(s), please complete one form for each of the five selected students. For each question, check only ONE option.

Please note that you need to fill out a separate survey questionnaire for each student.

Thank you for your support!

Last 4 digits of your phone number:

The school you are affiliated with:

Grade you serve as principal:

Please type in the student's ID:

Parental Involvement

A. Parent Involvement in School Activities

During the past semester, to what extent has the parent/guardian...

Q1. Participated in optional school or class activities/events beyond required parent meetings.

☐ Never ☐ Rarely ☐ Sometimes ☐ Often ☐ Very frequently

Q2. Visited school to meet me to discuss their child.

☐ Never ☐ Rarely ☐ Sometimes ☐ Often ☐ Very frequently

Q3. Volunteered to assist with class or school activities.

☐ Never ☐ Rarely ☐ Sometimes ☐ Often ☐ Very frequently

Q4. Visited the school to serve as a parent volunteer/in parent-teacher association.

☐ Never ☐ Rarely ☐ Sometimes ☐ Often ☐ Very frequently

Q5. Which of the following fruits is your favorite? For this question, please choose Banana.

☐ Apple ☐ Banana ☐ Orange ☐ Grapes

B. Parent-Teacher Contact

During the past semester, to what extent has the parent/guardian...

Q6. Responded to my messages about the child within a reasonable time.

☐ Never ☐ Rarely ☐ Sometimes ☐ Often ☐ Very frequently

Q7. Contacted me when questions or concerns about the child's learning or behavior arose.

☐ Never ☐ Rarely ☐ Sometimes ☐ Often ☐ Very frequently

Q8. Shared information from home that was helpful for supporting the child.

☐ Never ☐ Rarely ☐ Sometimes ☐ Often ☐ Very frequently

Q9. Worked with me on goals for the child's development and provided follow-up updates.

☐ Never ☐ Rarely ☐ Sometimes ☐ Often ☐ Very frequently

Student Persistence

C. Consistency of Interest

During the past semester, how true is each statement for this student?

Q10. This student tends to stick with goals they set over time.

☐ Strongly disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly agree

Q11. This student becomes very absorbed in a particular idea for a short time but later loses interest in it.

☐ Strongly disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly agree

Q12. This student can maintain focus on homework even when it requires more effort.

☐ Strongly disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly agree

Q13. New ideas sometimes distract this student from previous ones.

☐ Strongly disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly agree

D. Perseverance of Effort

During the past semester, how true is each statement for this student?

Q14. This student can finish the work they have started.

◦ Strongly disagree ◦ Disagree ◦ Neutral ◦ Agree ◦ Strongly agree

Q15. Setbacks can easily make this student frustrated.

◦ Strongly disagree ◦ Disagree ◦ Neutral ◦ Agree ◦ Strongly agree

Q16. This student is diligent.

◦ Strongly disagree ◦ Disagree ◦ Neutral ◦ Agree ◦ Strongly agree

Q17. The student consistently demonstrates avoidance when confronted with problems.

◦ Strongly disagree ◦ Disagree ◦ Neutral ◦ Agree ◦ Strongly agree

Please type in the student's ID:

...

(Each principal was asked to complete five individual questionnaires, one for each of five students. For ease of illustration, we display only the first student survey. The subsequent surveys are identical in content, apart from the omission of the Attention Check question (Q5).)

References

- Wing, C., S. M. Freedman, and A. Hollingsworth (2024). Stacked difference-in-differences. Technical report, National Bureau of Economic Research.
- Wooldridge, J. M. (2021). Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators.