

Measuring curriculum implementation using back-end app user data



*Stephen Taylor
South African Department of Basic Education
September 2024*



Motivation

- Implementation matters
 - For programme impacts
 - For scaling
- Curriculum coverage is a key aspect of implementation
 - For schooling to cause learning
 - Structured Learning Programmes have a coherent ToC targeting curriculum implementation
- Imperfect traditional ways of measuring curriculum coverage
 - Learner exercise books
 - Lesson observations
 - Self-reported information
- So, what about back-end user data from an app-based SLP?

BACKGROUND: INTERVENTION DESIGN



IMPACT LEVEL FINDINGS | Regression Results

Dependent variable: Each Language Composite Score, OLS Clustered Standard Errors						
	Grade 2, Wave 2 COHORT C (NEW) Grade 2 in 2023		Grade 3 COHORT A Grade 3 in 2023		Grade 4 COHORT B Grade 4 in 2023	
	HL	EFAL	HL	EFAL	HL	EFAL
External coaching	0.255*** (0.088)	0.323*** (0.090)	0.128 (0.081)	0.175** (0.081)	0.068 (0.077)	0.064 (0.082)
DH coaching	0.090 (0.105)	0.038 (0.096)	0.046 (0.074)	0.011 (0.074)	0.029 (0.068)	0.014 (0.068)
Observations	1,057	1,057	2,249	2,249	2,260	2,260
R ²	0.106	0.089	0.125	0.093	0.101	0.094
Adjusted R ²	0.094	0.076	0.119	0.087	0.095	0.088

Note: All regressions include individual, strata and district controls.

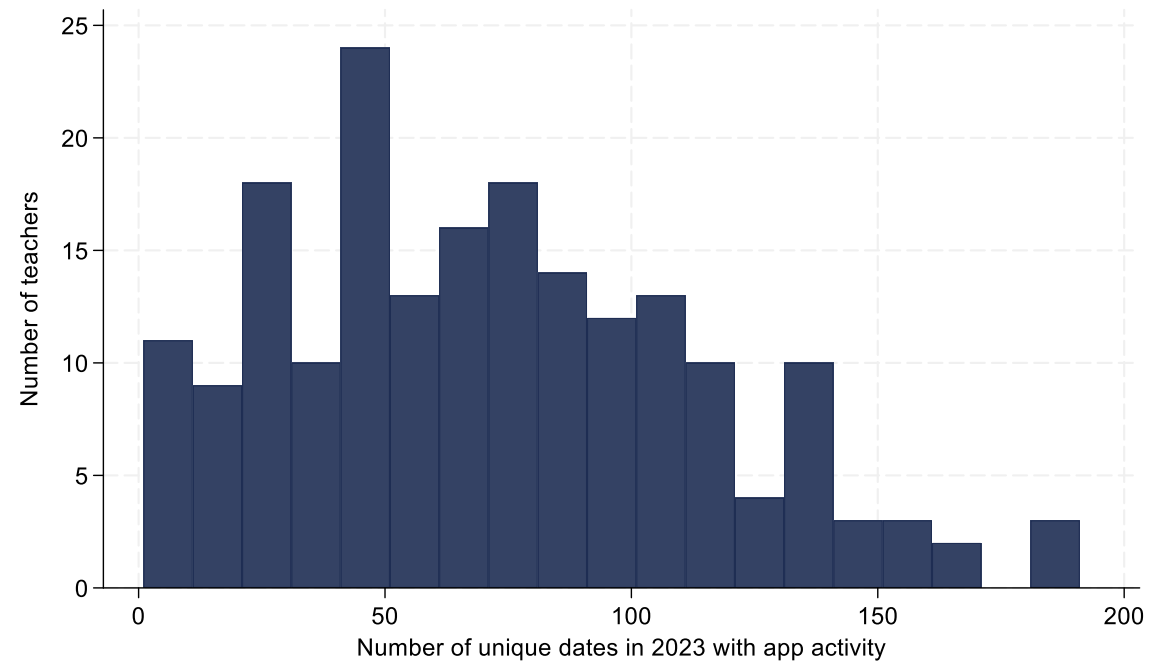
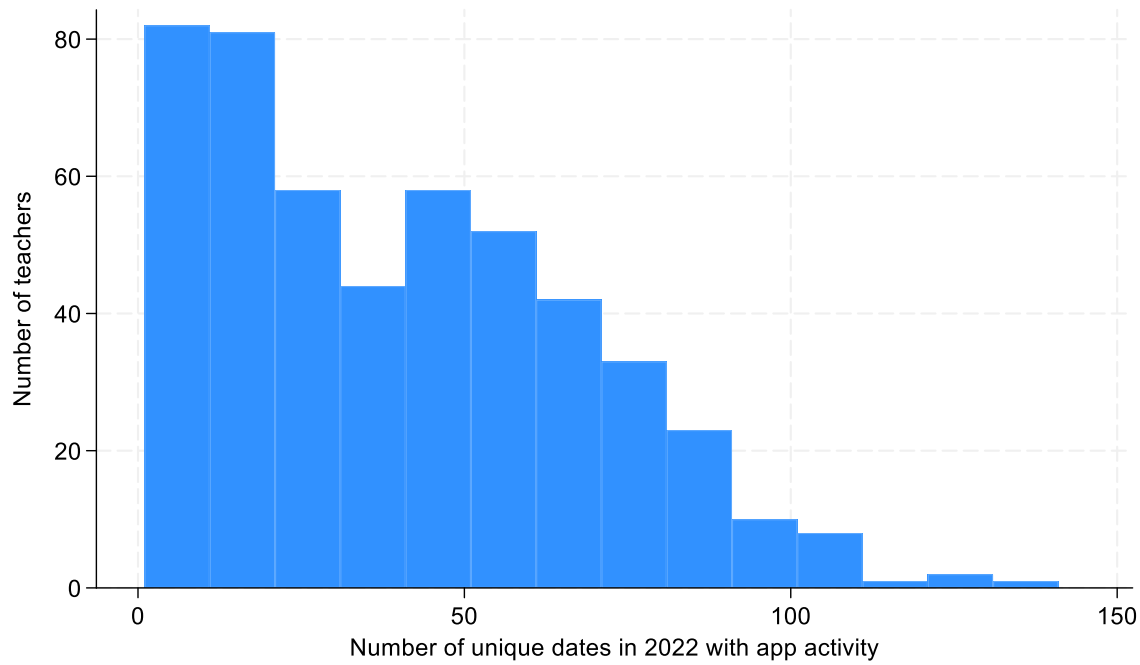
Statistical significance is indicated as follows * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$



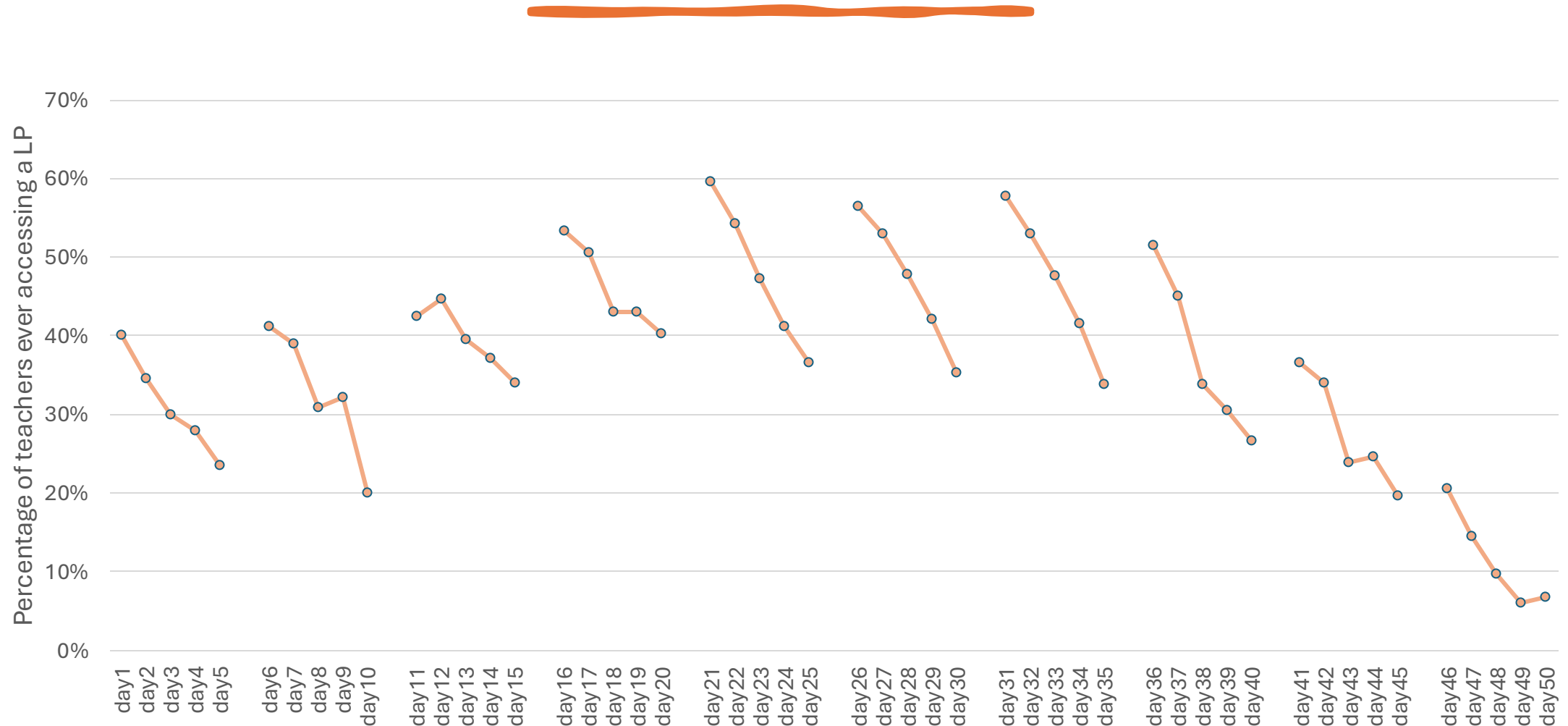
Back-end app user data

- All Grade 1-3 teachers given electronic tablets with an EGRP app
- App included daily lesson plans, each with a few slides
- Each row in the back-end data shows:
 - School
 - Username (teacher)
 - What slide was accessed (e.g., Term 2, Week 3, Tuesday, Shared Reading)
 - Start time and end time
- For 2022 and 2023
- We find 495 users in 2022 and only 193 in 2023 (compared to about 1000 teachers we expected)
- Not sure why so many teachers do not show up in the data, especially in 2023
 - Never used vs user data not uploading?
 - We find 134 out of 140 schools, so unlikely to be signal;
 - Likelihood of being in the data unrelated to proxies for signal strength (distance to town or major road)
 - Broken or lost tablets?

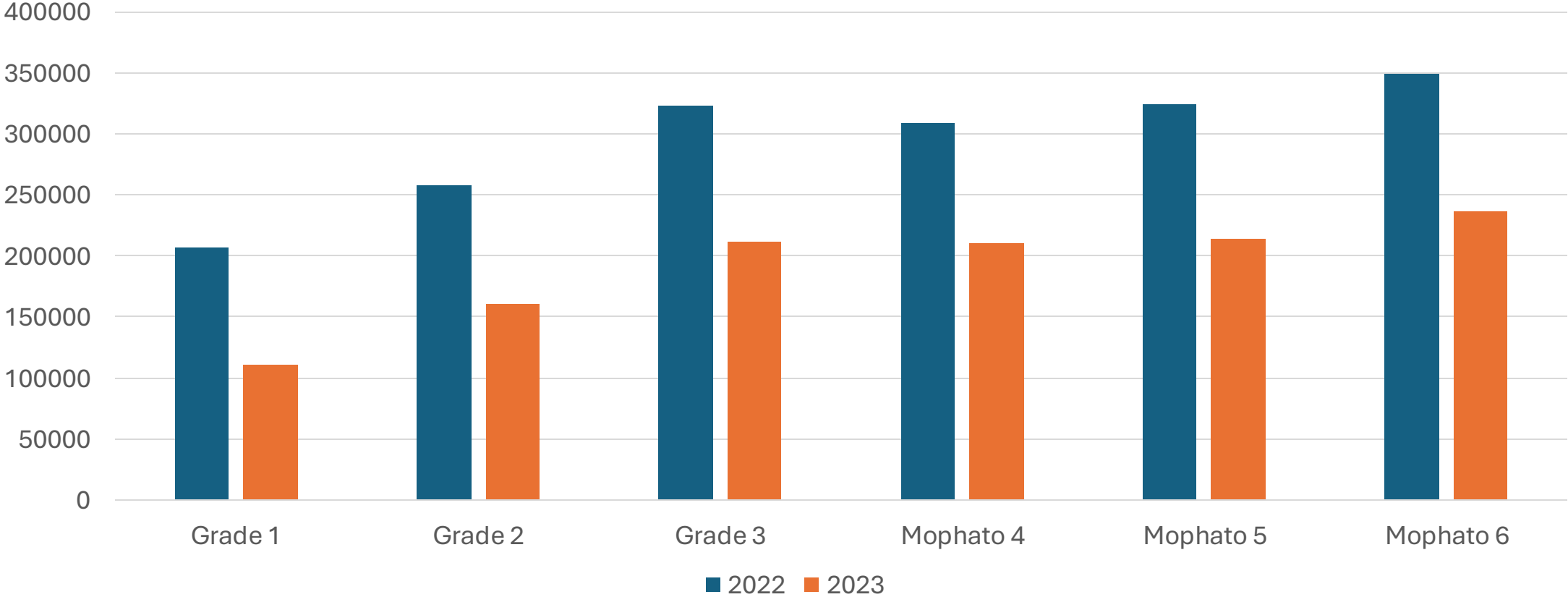
Overall, app use was low (contrary to self-reported use)



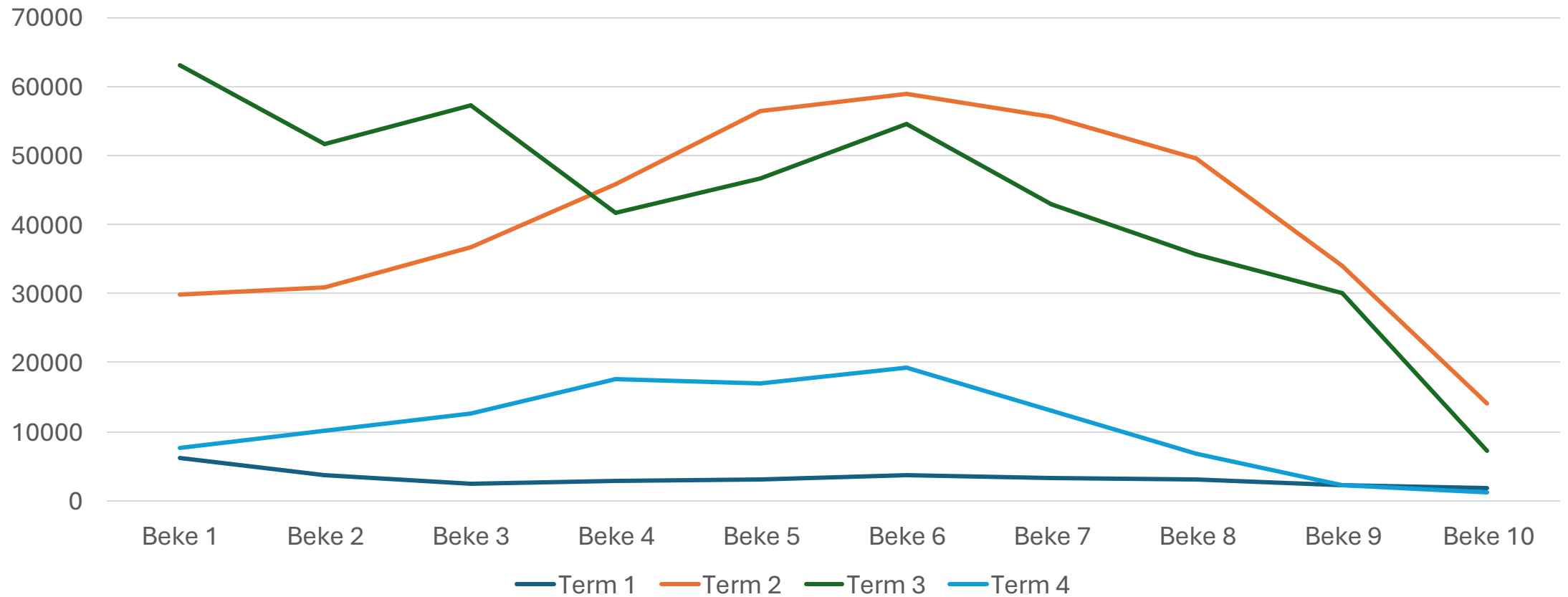
Percentage of lessons ever accessed (Term 2, 2022)



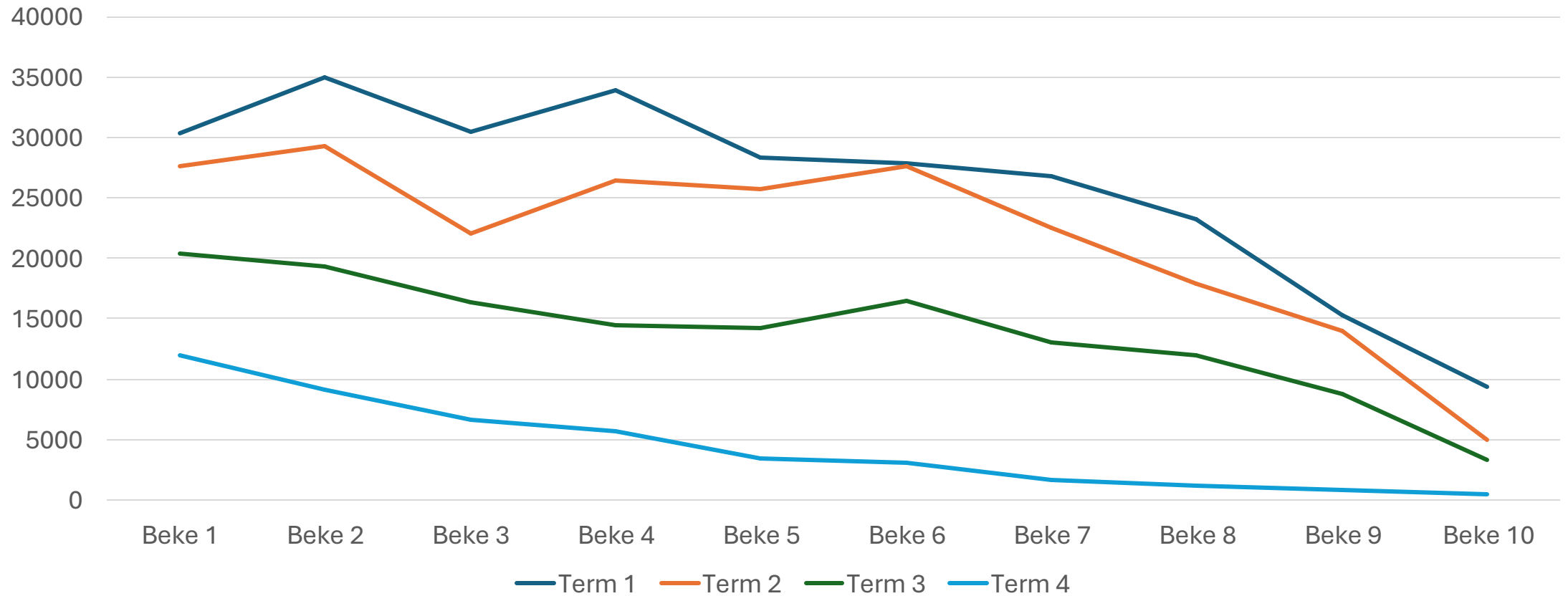
App usage events by grade and subject



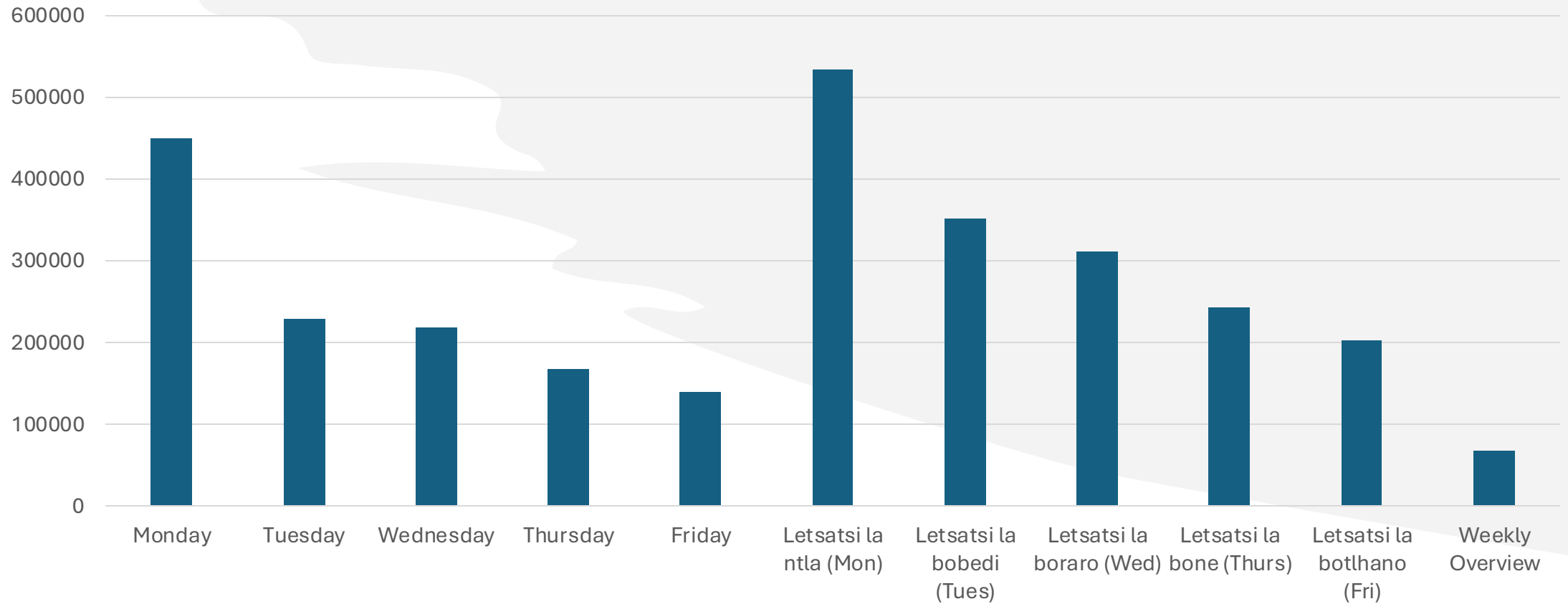
HL App usage events in 2022



HL App usage events in 2023



App usage events by Day of the Week



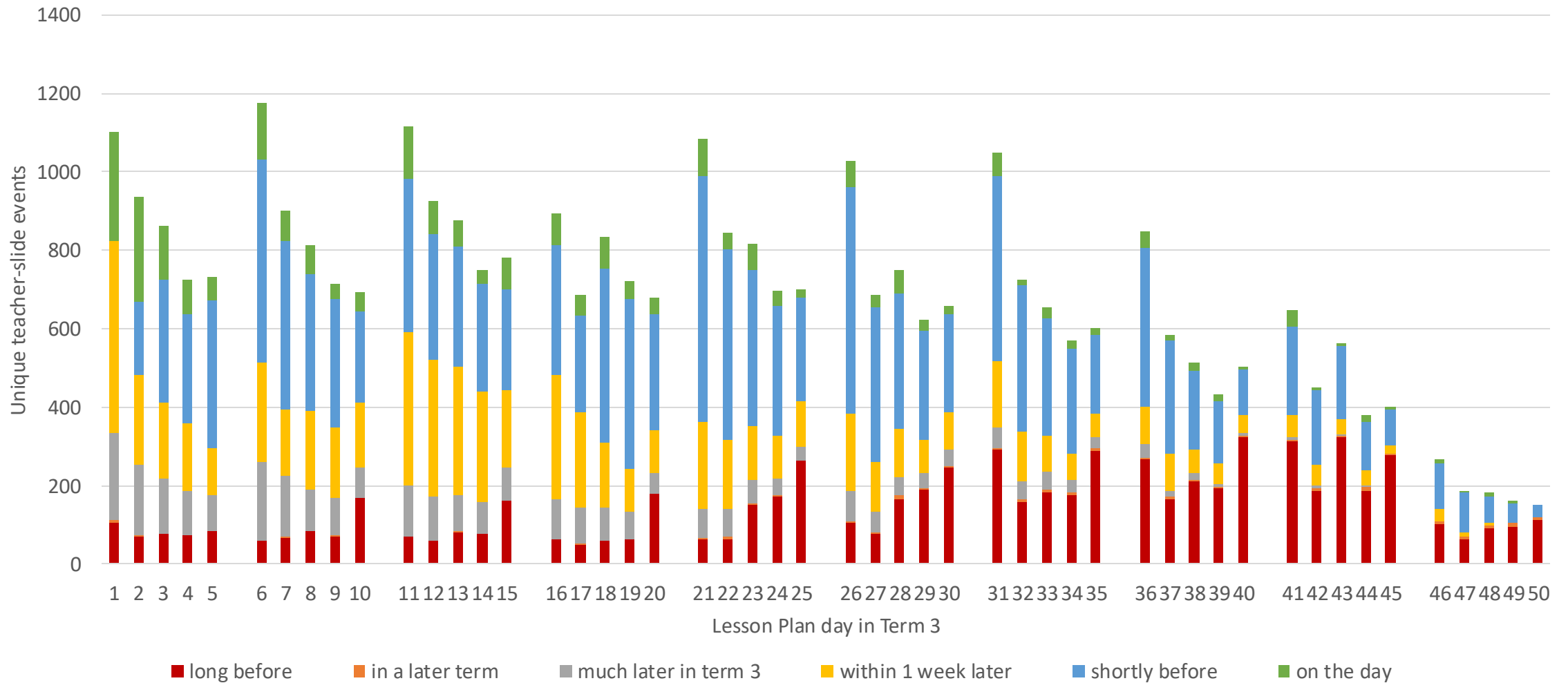
When did teachers access the lesson plans?

Day of the Lesson

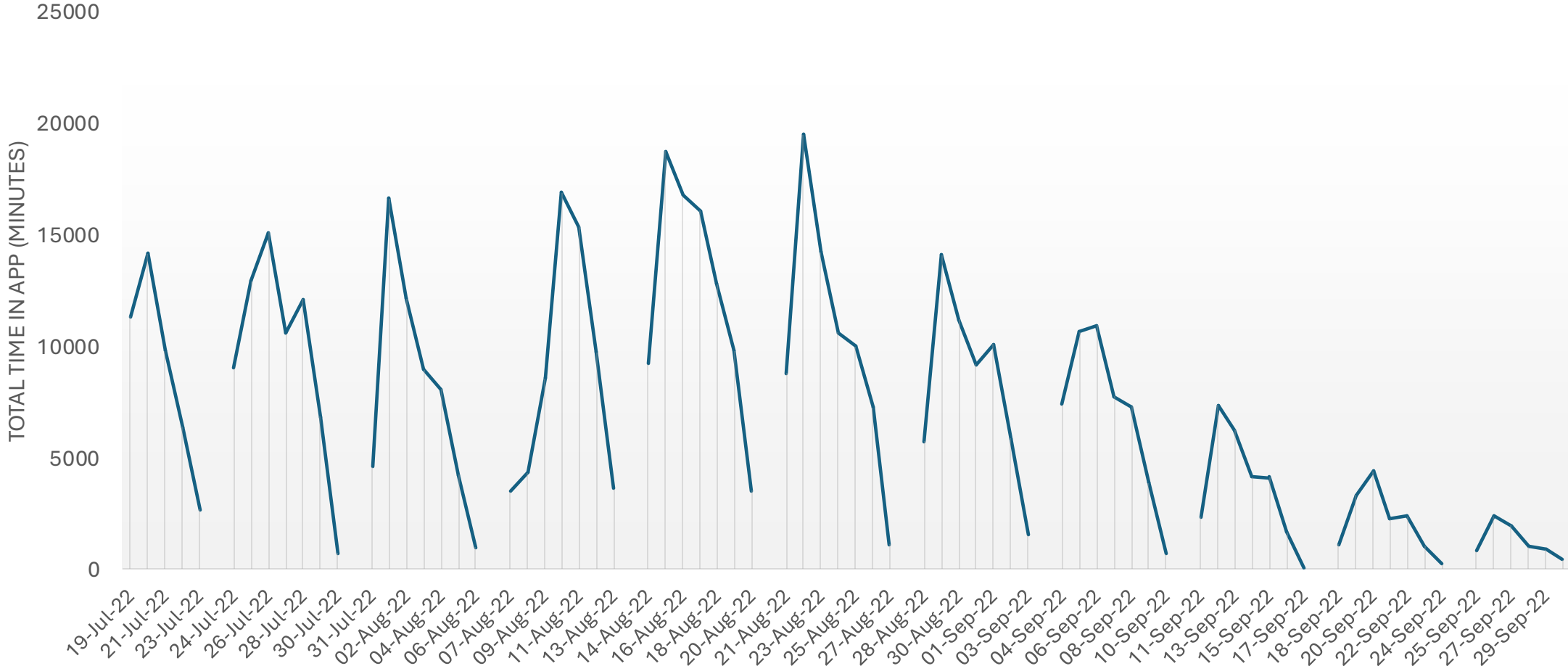
Day Accessed

	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Total
Monday	13.47	23.75	19.89	16.68	13.8	8.19	4.22	100
Tuesday	12.99	18.09	24.45	17.68	14.27	8.13	4.39	100
Wednesday	12.61	16.56	18.2	24.25	15.42	8.8	4.15	100
Thursday	13.58	16.7	15.83	16.89	22.17	10.1	4.74	100
Friday	14.04	17.73	16.72	15.19	15.52	16.03	4.77	100
Letsatsi la ntle (Mon)	12.52	27.32	19.92	15.19	13.31	7.52	4.21	100
Letsatsi la bobedi (T)	11.19	17.88	26.77	16.79	15.04	7.96	4.37	100
Letsatsi la boraro (W)	12.44	17.47	17.4	23.59	16.32	8.28	4.49	100
Letsatsi la bone (Thu)	13.19	16.57	16.44	16.14	23.9	8.95	4.8	100
Letsatsi la bothano	15.54	18.85	15.8	14.64	14.96	14.57	5.65	100
Weekly Overview	13.92	20.77	19.69	17.97	14.76	8.98	3.92	100
Total	13.03	20.37	19.65	17.57	15.7	9.21	4.47	100

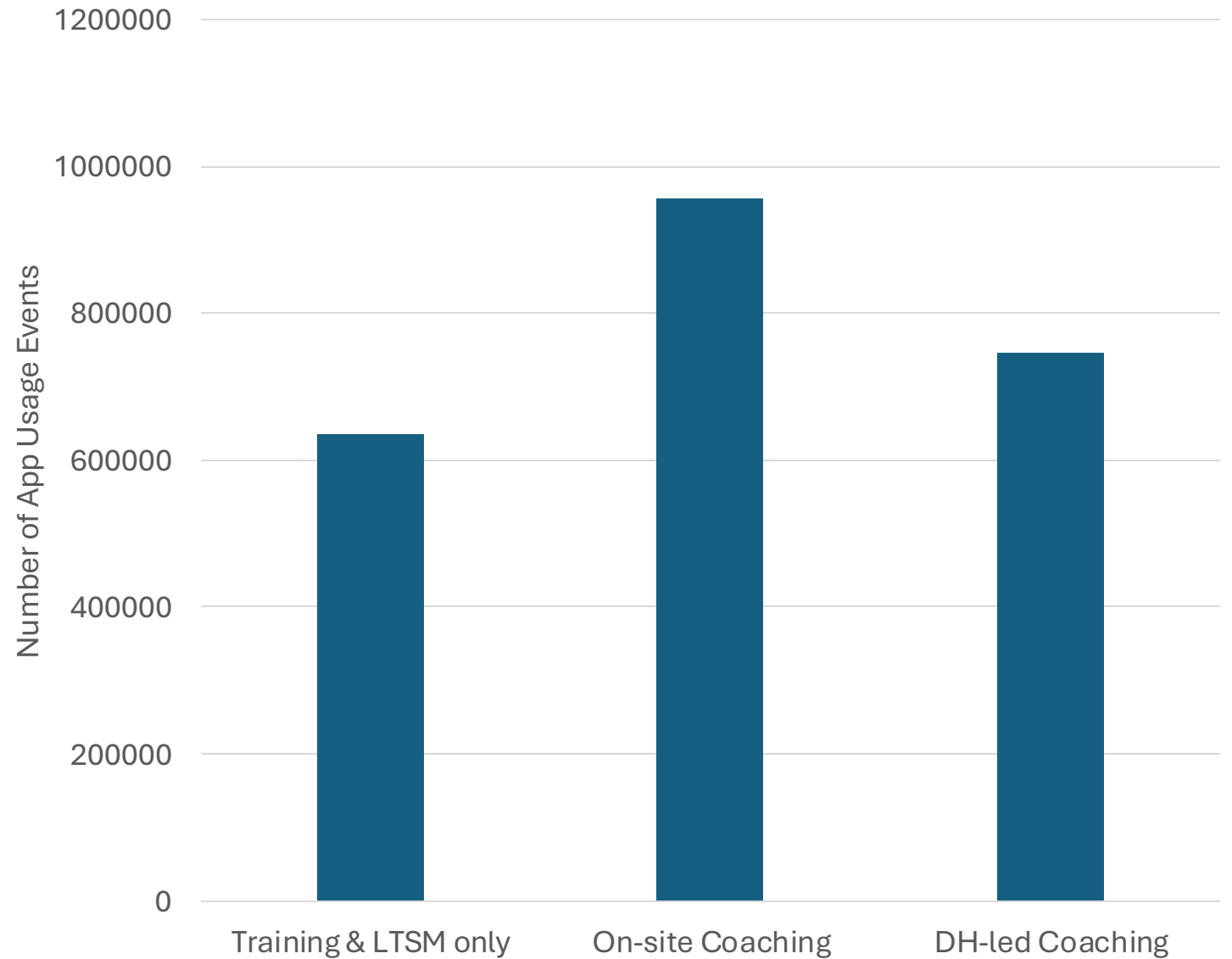
When were Term 3's (2022) lessons accessed?



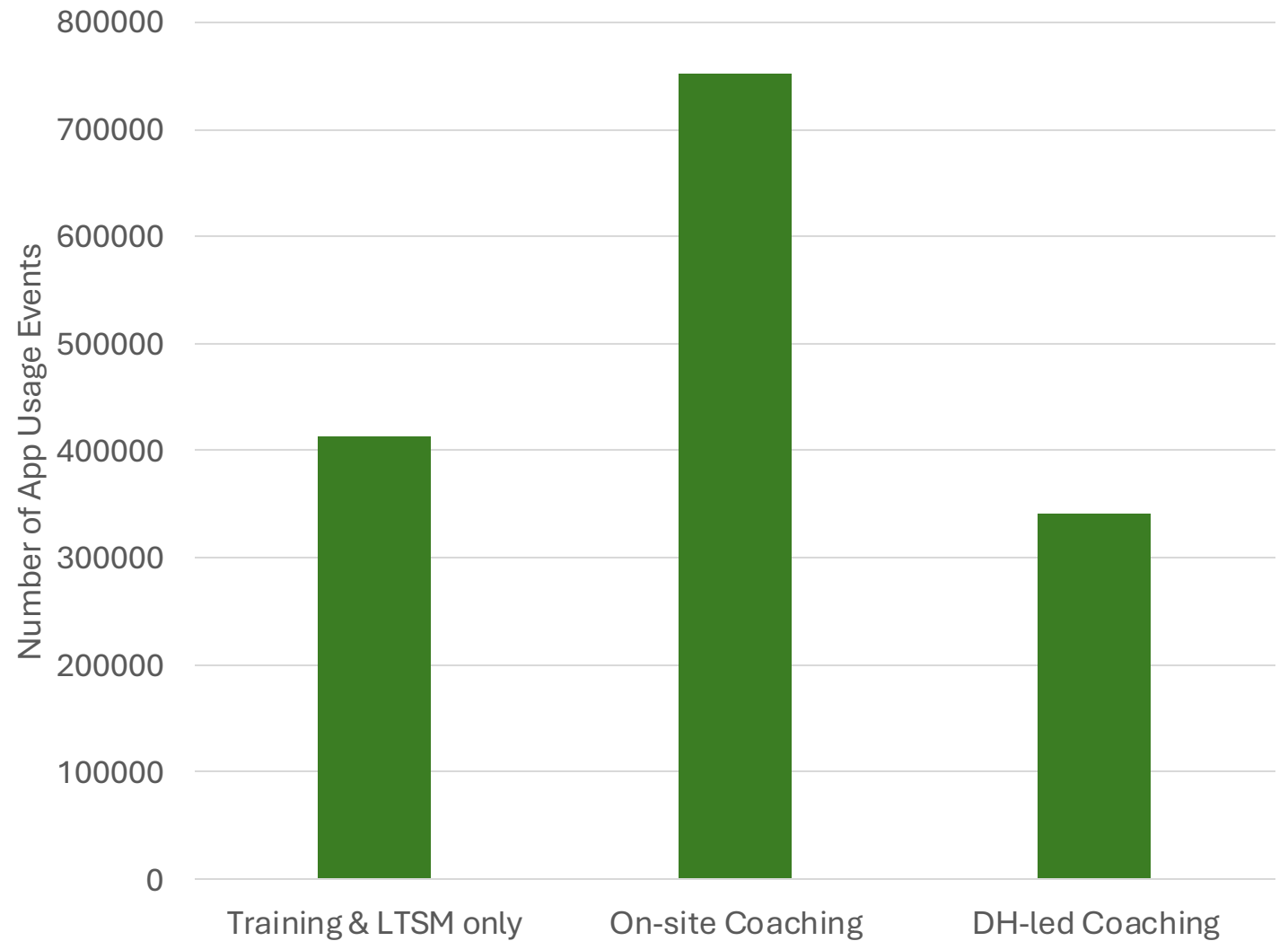
Term 3 (2022) app use by *date*



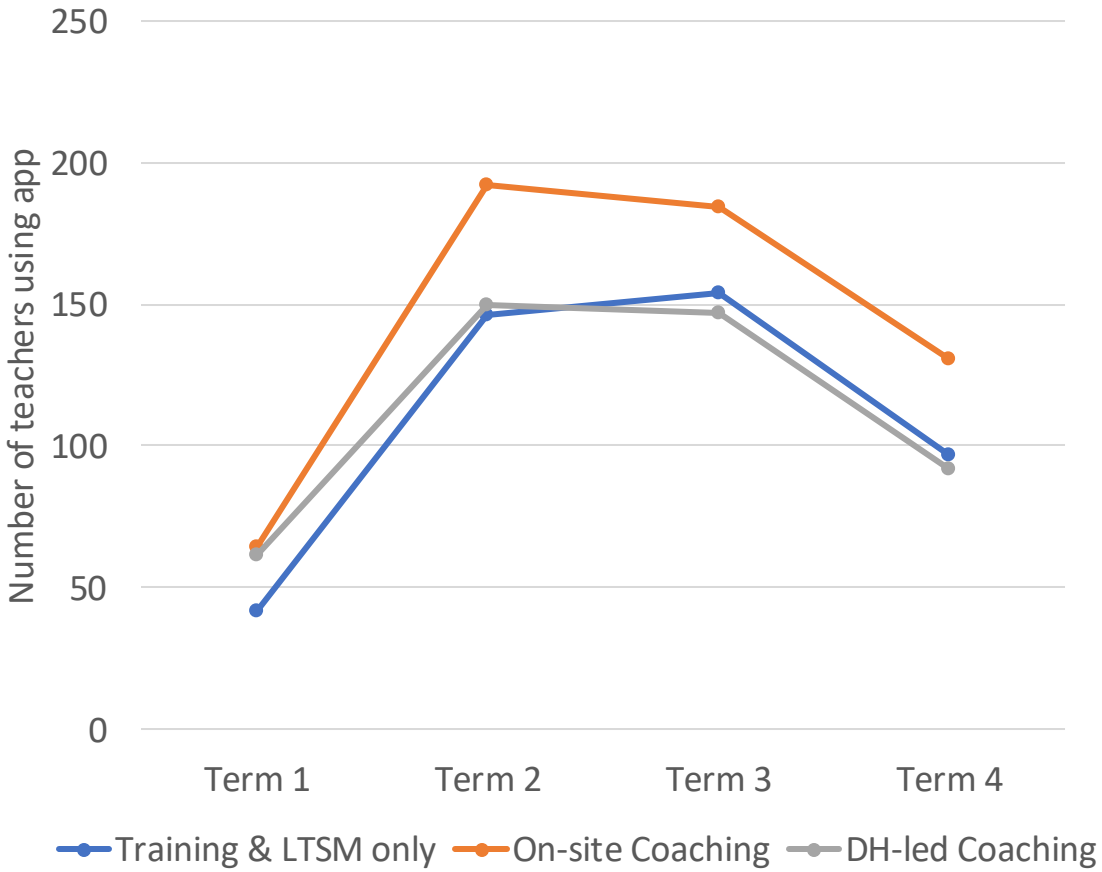
App usage by
treatment
group in 2022
(Adjusted for
differential numbers of
schools in each group)



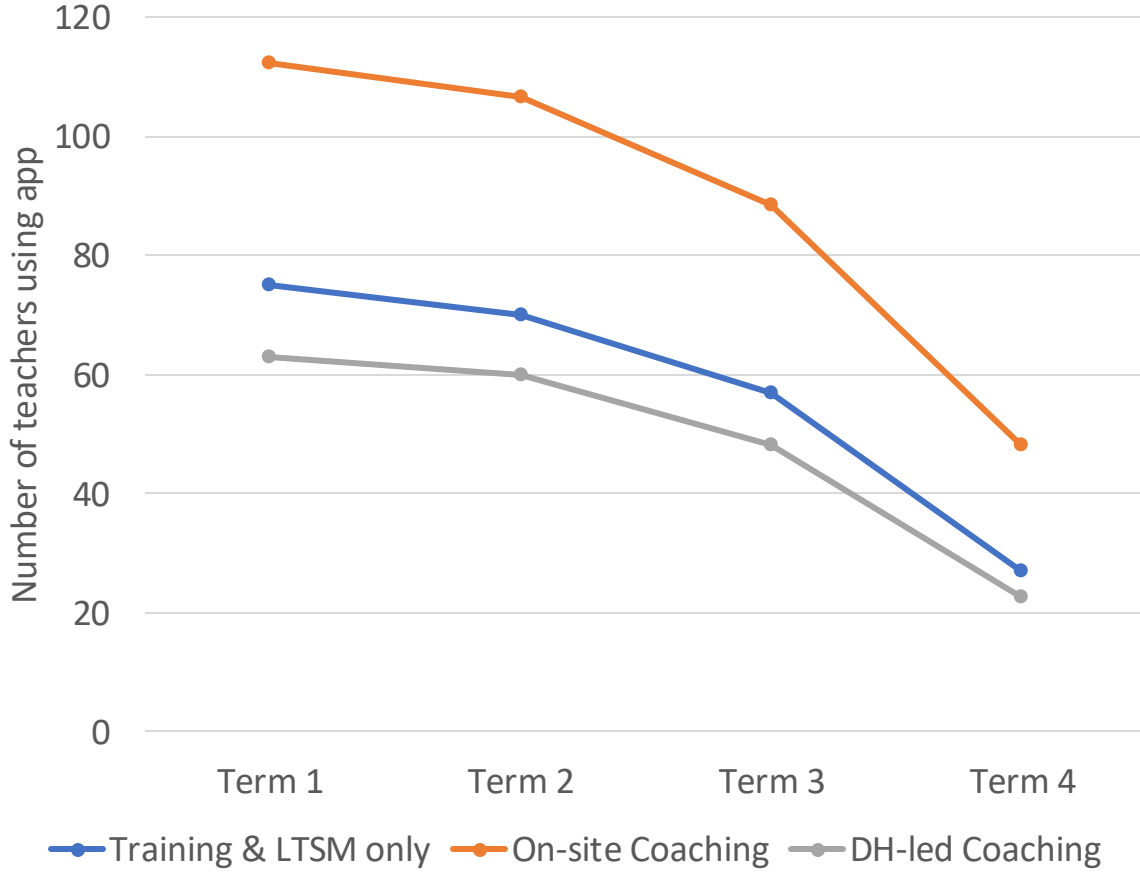
App usage by
treatment
group in 2023
(Adjusted for
differential numbers of
schools in each group)



Number of teachers using app

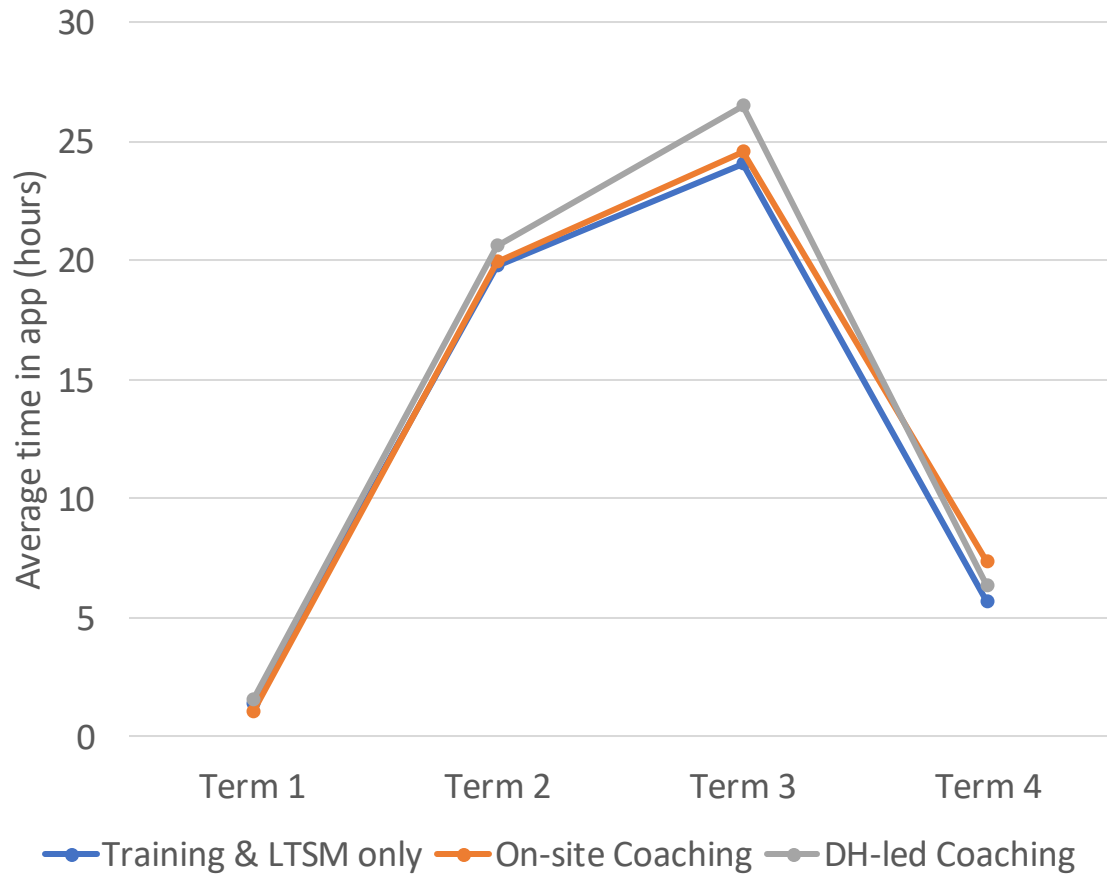


2022

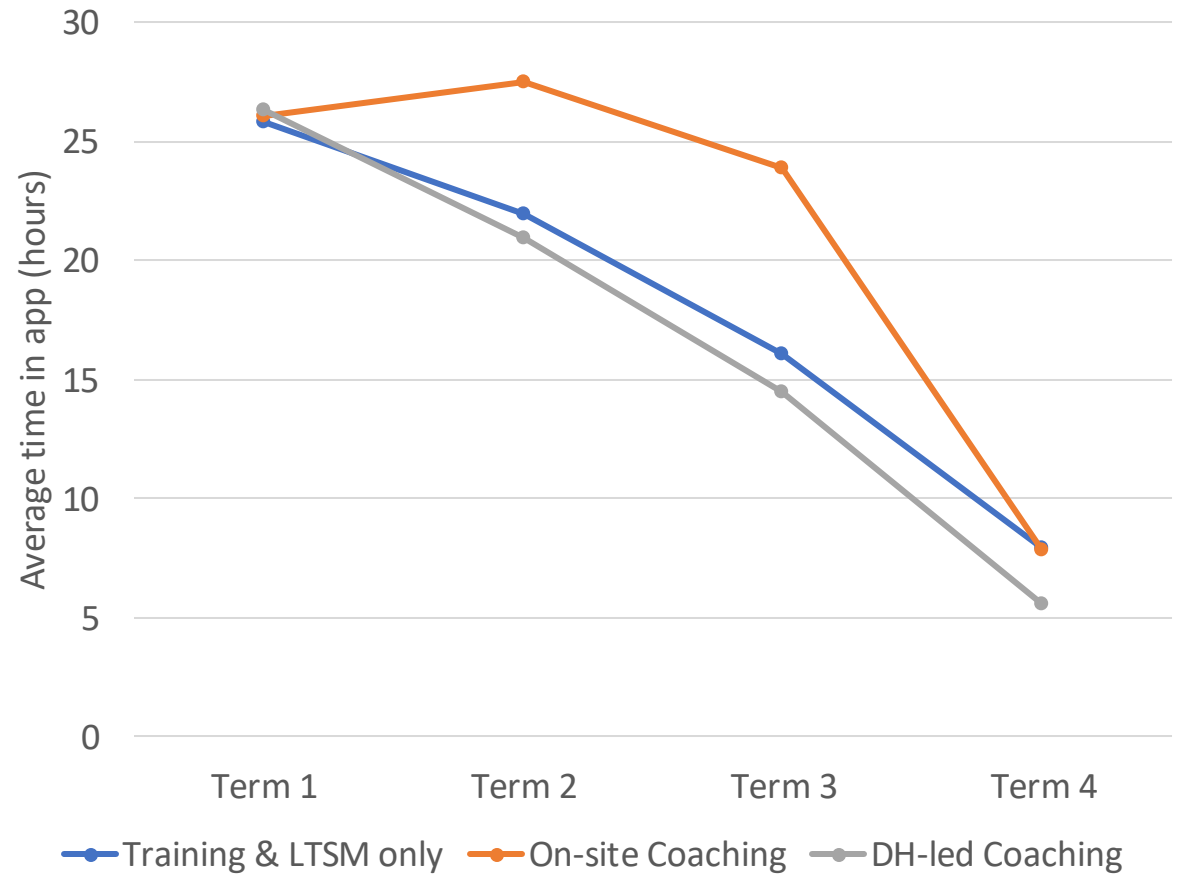


2023

Average time in the app

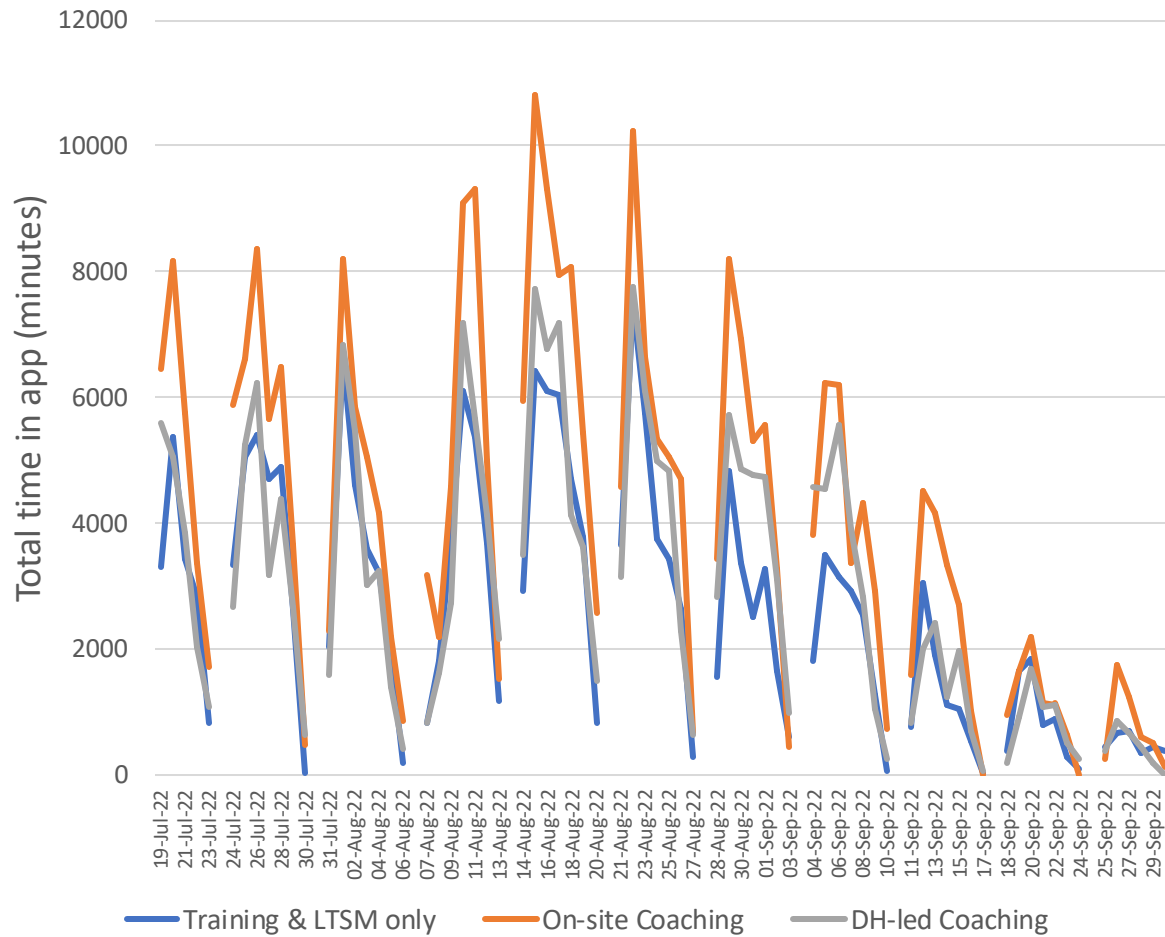


2022

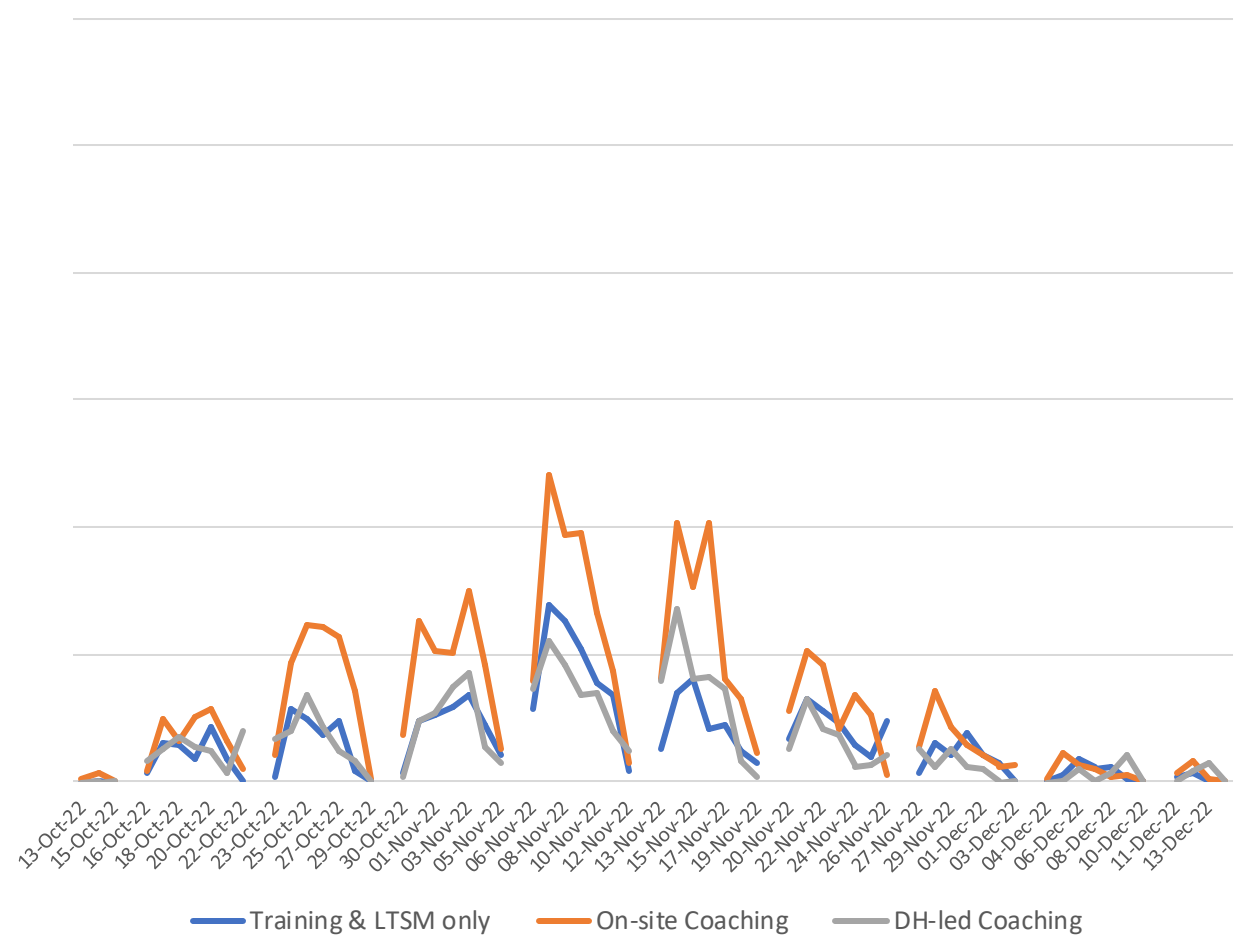


2023

Total time using app by *date*



Term 3, 2022



Term 4, 2022

Treatment effect on the better implementers?

- Problem: How to observe control teachers who would have had strong implementation?
- Instrument for a *particular* measure of implementation using treatment assignment?
- IV assumes impact runs only through that measure of implementation
- EGRP: All treatment groups had tablets so we can observe different levels of implementation across groups



Heterogeneous effects by implementation fidelity – Grade 2 Home Language Literacy

	(1) Main	(2) Impl_1B	(3) Impl_1G	(4) Impl_2B	(5) Impl_2G	(6) Impl_StrataB	(7) Impl_StrataG
DH Coaching	0.07	0.03	0.12	-0.15	0.34*	-0.01	0.15
Ext. Coaching	0.25***	0.12	0.30**	0.08	0.44***	0.24**	0.27**
Observations	1,058	524	471	525	470	650	408
R-squared	0.11	0.14	0.13	0.16	0.15	0.10	0.10

Robust standard errors not shown

*** p<0.01, ** p<0.05, * p<0.1

Heterogeneous effects by implementation fidelity – Grade 3 Home Language Literacy

	(1) Main	(2) Impl_1B	(3) Impl_1G	(4) Impl_2B	(5) Impl_2G	(6) Impl_StrataB	(7) Impl_StrataG
DH Coaching	0.05	-0.04	0.07	0.01	0.03	-0.00	0.10
Ext. Coaching	0.13	0.11	0.14	0.09	0.08	0.09	0.18
Observations	2,249	1,091	1,030	1,102	1,019	1,338	911
R-squared	0.12	0.15	0.14	0.14	0.16	0.12	0.13

Robust standard errors not shown
 *** p<0.01, ** p<0.05, * p<0.1

Heterogeneous effects by implementation fidelity – Grade 4 Home Language Literacy

	(1) Main	(2) Impl_1B	(3) Impl_1G	(4) Impl_2B	(5) Impl_2G	(6) Impl_StrataB	(7) Impl_StrataG
DH Coaching	0.03	-0.06	0.08	-0.06	0.09	-0.03	0.11
Ext. Coaching	0.07	-0.05	0.17*	-0.15	0.35***	0.03	0.11
Observations	2,262	1,093	1,043	1,111	1,025	1,353	909
R-squared	0.10	0.13	0.09	0.12	0.10	0.10	0.11

Robust standard errors not shown
 *** p<0.01, ** p<0.05, * p<0.1

Conclusion

- Coaching improved programme implementation
 - Low-order mechanistic aspect of the Coaching theory of change
- Impact of coaching was greater amongst more compliant teachers
 - lower bound given that I am using school-level estimates of implementation not teacher-level
- Ed-tech: Tablets vs Paper-based LPs?
 - Tech more of a behavioural obstacle than something (older) teachers are afraid of?
- Can user data be used by coaches/officials to monitor curriculum coverage?
 - Or will this lead to compliance behaviour or opposition?
- Can back-end user data be used more in similar evaluations?
 - We are doing so in EGRP 2
- New view of (low) time on task and curriculum coverage
 - building on literatures about teacher absenteeism in developing countries; overambitious curricula

If you could change one thing about EGRP....

