
Supplementary Materials: Statistical Output, Data Dictionary, and Reproducibility

*AI-Adaptive Learning Paths and Task-Difficulty Synchronization:
A Longitudinal Study of Student Persistence and Engagement*

This document consolidates the supplied supplementary statistical report and the three accompanying machine-readable datasets. The complete CSV files are embedded as attachments in the PDF and are also included verbatim inside this single $\LaTeX$ source through `filecontents*` environments.

Contents

Overview	2
S1. Descriptive Statistics	2
S2. Reliability (Cronbach's α)	3
S3. Correlation Matrix	3
S4. Confirmatory Factor Analysis (Three-Factor SES-AI, T2)	3
S5. Latent Growth Curve Model (H1)	4
S6. Cox Proportional Hazards Regression (H2, H5)	4
S7. Polynomial (Flow-Arch) Regression (H3)	5
S8. Mediation Path Analysis (H4, H6)	5
S9. Cross-Lagged Panel Model (H7)	5
S10. Experience-Sampling Trends by Phase	6
S11. Data Dictionary	6
students.csv data dictionary	6
sesai_items.csv data dictionary	7
esm_pulse.csv data dictionary	7
S12. Reproducibility	7
Appendix A. Embedded Data Files and Integrity Manifest	9
Appendix B. Data Previews	9

Overview

This document reports the complete statistical output underlying the results section of the manuscript, together with a data dictionary for the analysis dataset and instructions for reproducing every reported value. All analyses were conducted in R 4.3 (packages: `lavaan`, `survival`, and `psych`). The analysis dataset comprises $N = 79$ students, 45 SES-AI item responses (15 items $\times$ 3 waves), and 1,896 experience-sampling observations. The random seed (20250724) fixes the dataset so that all values below are exactly reproducible.

Table A1. Manifest of data files embedded in this PDF.

File	Rows	Columns	Contents
students.csv	79	15	Participant-level predictors, engagement and persistence measures, and survival outcomes.
sesai_items.csv	79	46	Participant identifier plus 45 SES-AI item responses across T1, T2, and T3.
esm_pulse.csv	1,896	10	Twenty-four experience-sampling pulse checks per participant.

S1. Descriptive Statistics

Table S1. Means and standard deviations of SES-AI subscales and the total engagement composite at each measurement wave (1–5 Likert metric), and of the remaining study variables.

Variable	Baseline (T1) M (SD)	Midpoint (T2) M (SD)	Final (T3) M (SD)
Behavioral engagement	3.83 (0.38)	4.06 (0.49)	4.25 (0.64)
Emotional engagement	3.84 (0.38)	4.15 (0.54)	4.28 (0.63)
Cognitive engagement	3.85 (0.36)	4.10 (0.51)	4.31 (0.65)
Total engagement (SES-AI)	3.84 (0.31)	4.10 (0.46)	4.28 (0.60)

Variable	M (SD) / summary
AI-adaptive path quality (1–5)	3.57 (0.56)
Perceived competence (1–5)	3.60 (0.63)
Task-difficulty synchronization (proportion optimal-flow)	0.59 (0.12)
Persistence index (standardized, 0–100 metric)	50.0 (10.0)
Time to disengagement (days)	$M = 127.2$, $SD = 103.8$, $Mdn = 84$
Disengagement events / censored	67 events, 12 censored

S2. Reliability (Cronbach's α)

Table S2. Internal consistency of the SES-AI subscales at each administration (raw Cronbach's α ; average inter-item correlation in brackets).

Subscale	Baseline (T1)	Midpoint (T2)	Final (T3)
Behavioral	.85 [.54]	.83 [.49]	.85 [.53]
Emotional	.87 [.57]	.88 [.60]	.87 [.59]
Cognitive	.86 [.54]	.88 [.61]	.90 [.65]

Note. All coefficients fall in the acceptable-to-excellent range (.83–.90).

S3. Correlation Matrix

Table S3. Zero-order Pearson correlations among key study variables ($N = 79$).

Variable	1	2	3	4	5	6	7
1. Engage- ment T1	1.00						
2. Engage- ment T2	.64	1.00					
3. Engage- ment T3	.53	.80	1.00				
4. Path quality	.27	.74	.76	1.00			
5. Compe- tence	.79	.63	.54	.23	1.00		
6. Synchro- nization (opt %)	.44	.27	.28	.18	.46	1.00	
7. Persistence index	.35	.48	.49	.39	.27	.08	1.00

S4. Confirmatory Factor Analysis (Three-Factor SES-AI, T2)

Model fit: $\chi^2(87) = 100.14$, $p = .16$; $\chi^2/df = 1.15$; CFI = .98; TLI = .98; RMSEA = .044 (90% CI [.00, .078]); SRMR = .053. Inter-factor correlations: Behavioral–Emotional = .80, Behavioral–Cognitive = .84, Emotional–Cognitive = .82.

Table S4. Standardized factor loadings (all $p < .001$).

Item	Factor	Std. loading
B1	Behavioral	.67
B2	Behavioral	.72
B3	Behavioral	.68
B4	Behavioral	.79
B5	Behavioral	.62
E1	Emotional	.68
E2	Emotional	.75
E3	Emotional	.83
E4	Emotional	.83
E5	Emotional	.78
C1	Cognitive	.76
C2	Cognitive	.81
C3	Cognitive	.82
C4	Cognitive	.77
C5	Cognitive	.74

S5. Latent Growth Curve Model (H1)

Table S5. Unstandardized estimates from the linear latent growth model of total engagement, with AI-adaptive path quality (grand-mean centered) predicting the latent slope.

Parameter	Estimate (B)	SE	z	p
Intercept mean (initial engagement)	3.85	0.03	112.2	< .001
Slope mean (rate of change)	0.23	0.02	12.7	< .001
Path quality → slope	0.40	0.03	12.9	< .001
Intercept variance	0.07	0.01	5.2	< .001

S6. Cox Proportional Hazards Regression (H2, H5)

Table S6. Cox model for time to disengagement (≥ 7 consecutive days without login). Overall model: likelihood-ratio $\chi^2(3) = 35.9$, $p < .001$; Harrell's $C = .71$. The proportional-hazards assumption held (global scaled-Schoenfeld test $p = .96$).

Predictor	<i>B</i>	HR	95% CI	<i>p</i>
Task-difficulty synchronization (group)	−0.87	0.42	[0.25, 0.70]	< .001
AI path quality (<i>z</i>)	−0.25	0.78	[0.57, 1.06]	.109
Synchronization × path quality	−0.71	0.49	[0.29, 0.84]	.010

S7. Polynomial (Flow-Arch) Regression (H3)

Standardized quadratic regression of momentary engagement on perceived task difficulty across 1,896 experience-sampling observations. Model $R^2 = .16$, $F(2, 1893) = 179.7$, $p < .001$. Linear term $\beta = -0.16$ ($p < .001$); quadratic term $\beta = -0.37$ ($p < .001$). The fitted curve peaks at an estimated difficulty of ≈ 4.0 on the 1–7 scale (optimal/flow), confirming the inverted-U relationship.

S8. Mediation Path Analysis (H4, H6)

Table S8. Standardized path estimates with 2,000 bias-corrected bootstrap resamples.

Effect	Estimate	SE	95% CI	<i>p</i>
H4: Path quality → Engagement (<i>a</i>)	0.72	0.08	[0.57, 0.87]	< .001
H4: Engagement → Persistence (<i>b</i>)	0.48	0.13	[0.26, 0.76]	< .001
H4: Direct (path quality → persistence, <i>c'</i>)	0.05	0.12	[−0.19, 0.27]	.686
H4: Indirect ($a \times b$)	0.34	0.09	[0.19, 0.55]	< .001
H6: Synchronization → Competence (<i>a</i>)	0.76	0.08	[0.60, 0.92]	< .001
H6: Competence → Engagement (<i>b</i>)	0.62	0.15	[0.30, 0.88]	< .001
H6: Direct (synchronization → engagement, <i>c'</i>)	0.11	0.15	[−0.17, 0.42]	.468
H6: Indirect ($a \times b$)	0.47	0.12	[0.23, 0.70]	< .001

S9. Cross-Lagged Panel Model (H7)

Table S9. Standardized cross-lagged and autoregressive estimates.

Path	β	SE	z	p
Engagement T1 → Persistence T2 (cross-lag)	0.27	0.09	2.86	.004
Persistence T2 → Engagement T3 (cross-lag)	0.33	0.06	5.55	< .001
Persistence T1 → Persistence T2 (autoregressive)	0.45	0.09	4.80	< .001
Engagement T2 → Engagement T3 (autoregressive)	0.71	0.06	12.07	< .001

S10. Experience-Sampling Trends by Phase

Table S10a. *Distribution of task-difficulty states (% of ESM responses).*

Difficulty state	Phase 1	Phase 2	Phase 3
Optimal / flow	41.9	61.1	73.3
Anxiety-inducing	46.8	28.8	11.1
Boring	11.2	10.1	15.7

Table S10b. *Endorsement of affective states (% of ESM responses; multiple selections allowed).*

Affective state	Phase 1	Phase 2	Phase 3
Curious	71.0	73.4	69.9
Frustrated	30.1	16.0	9.7
Excited	12.5	35.8	57.3
Neutral	40.0	31.6	23.9

S11. Data Dictionary

The analysis dataset is distributed as three CSV files.

students.csv (79 rows—one per participant)

Field	Description
id	Participant identifier (1–79).
path_quality	AI-adaptive learning path quality, mean rating (1–5).

Field	Description
competence	Perceived competence, mean rating (1–5).
sync_score	Task-difficulty synchronization index (standardized).
sync_prop	Proportion of ESM responses in the optimal/flow state (0–1).
sync_group	Synchronized-difficulty group (1 = high, 0 = low; median split).
eng_t1, eng_t2, eng_t3	SES-AI total engagement composite at each wave (1–5).
pers_t1, pers_t2, pers_t3	Persistence index at each wave (standardized, 0–100 metric).
persistence_index	Overall persistence index.
survival_days	Days until first disengagement episode (or censoring at 300).
event	1 = disengagement episode observed; 0 = censored.

sesai_items.csv (79 rows × 46 columns)

id plus 45 item columns named <Item>_<Wave>, for example, B1_T1 through C5_T3, where B1–B5 = behavioral, E1–E5 = emotional, and C1–C5 = cognitive engagement items. Each item uses a 1–5 Likert scale at waves T1 (baseline), T2 (midpoint), and T3 (final).

esm_pulse.csv (1,896 rows—24 pulse checks per participant)

Field	Description
id	Participant identifier.
phase	Study phase (1 = initial, 2 = advanced synchronization, 3 = mastery).
ping	Pulse-check number within participant (1–24).
difficulty_state	boring / optimal / anxiety (Part A categorical response).
difficulty	Continuous perceived difficulty (1–7).
momentary_engagement	Continuous momentary engagement/flow rating.
aff_frustrated, aff_curious, aff_excited, aff_neutral	Part B affect selections (1 = selected, 0 = not).

S12. Reproducibility

All reported statistics can be regenerated from the accompanying R scripts:

1. `01_generate_data.R` — builds the analysis dataset with a fixed seed (20250724) and writes the three CSV files to a `./data/` folder. Base R only.

2. `02_analysis.R` — reads the CSV files and runs every model reported in the manuscript (reliability, CFA, latent growth curve model, Cox regression, polynomial regression, mediation with bootstrapping, and the cross-lagged panel model), writing a full results log to `results_dump.txt`. Requires the `lavaan`, `survival`, and `psych` packages.

From a terminal:

```
Rscript 01_generate_data.R && Rscript 02_analysis.R
```

Because the seed is fixed, the tables above are reproduced exactly; minor differences from rounded values in the main text reflect rounding only.

Assembly note. The two R scripts referenced above were not among the files supplied for this compilation. The PDF embeds the three supplied CSV datasets in full.

Appendix A. Embedded Data Files and Integrity Manifest

The three complete datasets are embedded in the PDF as file attachments. In compatible PDF readers, open the attachments panel to save the original CSV files. The same CSV contents are included verbatim near the beginning of the accompanying `.tex` source, making the source package self-contained.

File	SHA-256 checksum
students.csv	3df0c5a84418e5b8dd574194926d7282225979e6c8de1db15215b0757e66abd3
sesai_items.csv	1507b3ede75de19cd085f14f52e95d8cafc37490aeddec338304d1eb78623e42
esm_pulse.csv	5da7d2100b0c116d6d4f284ab90fcf418aea49acf0124df57b2e529378f0b5b5

Appendix B. Data Previews

The tables below provide compact previews only. The embedded CSV attachments contain every row and column.

B1. students.csv preview (first 8 rows)

id	path quality	competence	sync prop	eng T1	eng T2	eng T3	persistence	survival	event
1	3.975	3.533	0.458	4.166	4.521	4.887	64.46	151	1
2	4.597	4.724	0.667	4.275	5	5	43.27	300	0
3	2.878	3.835	0.5	3.711	3.794	4.139	47.65	25	1
4	2.912	4.151	0.667	4.181	4.005	4.027	45.69	226	1
5	3.222	3.083	0.5	3.568	3.952	3.559	34.23	76	1
6	3.439	3.246	0.542	3.624	4.014	3.978	55.06	73	1
7	2.397	2.81	0.542	3.621	3.336	3.001	43.68	28	1
8	2.846	3.939	0.625	3.777	3.563	3.351	32.39	4	1

B2. sesai_items.csv preview (selected columns, first 6 rows)

id	B1 T1	B5 T1	E1 T1	E5 T1	C1 T1	C5 T1	B1 T2	E1 T2	C1 T2	B1 T3	E1 T3	C1 T3
1	4	4	4	4	4	4	5	5	5	4	5	4
2	5	4	4	5	5	4	5	5	4	5	4	5
3	4	4	4	4	3	4	4	4	4	4	4	4
4	4	4	4	4	4	4	4	4	4	4	4	3
5	4	3	3	3	4	4	3	4	4	4	3	3
6	4	3	4	4	4	4	5	4	5	5	4	3

B3. esm_pulse.csv preview (first 12 rows)

id	phase	ping	state	difficulty	engagement	frustrated	curious	excited	neutral
1	1	1	anxiety	5.49	3.41	1	1	0	1
1	1	2	optimal	4.24	4.96	0	0	0	1
1	1	3	anxiety	6.17	6.29	0	1	0	0
1	1	4	optimal	4.67	5.58	0	0	0	1
1	1	5	optimal	4.45	5.18	0	0	1	0
1	1	6	anxiety	4.31	4.84	0	1	0	0
1	1	7	optimal	3.32	5.76	0	1	0	0
1	1	8	anxiety	6.66	3.12	0	0	0	1
1	2	9	optimal	3.17	4.12	0	1	0	1
1	2	10	boring	2.47	5.94	0	1	0	0
1	2	11	anxiety	4.16	4.33	0	1	0	0
1	2	12	boring	1.68	4.59	0	1	0	1