Longitudinal Growth Across Content, SEL, and Future-Ready Competencies

immersionED Simulation Platform · Multi-Domain Growth Study · April 2026

About this study. immersionED's adaptive simulations let students practice reasoning, communication, and decision-making in narrative scenarios that change based on what they say and do. This study examines whether students show measurable growth across three independent measurement frameworks — academic content, social-emotional learning, and future-ready competencies — through repeated simulation use. The goal is to give educators evidence that immersive practice produces real, measurable improvement in the skills that matter most, so they can invest classroom time with confidence. These findings are observational and preliminary; immersionED is extending this work through formal, university-led research.

The more students play, the more they grow — across academic content, SEL, and future-ready competencies.

Statistically significant longitudinal growth on three independent measurement frameworks (all p<.001)

+7.0
AP-RUBRIC GROWTH (US HISTORY)
pts · 61.7→68.6

+0.31
SEL GROWTH
0–4 scale

+0.35
COMPETENCY GROWTH
0–4 scale

~4
SIMS TO SUBSTANTIAL
growth

+0.63
PERSUASIVE COMMUNICATION
1.87→2.50 | d=0.48 | p<.001

+0.56
LOGICAL THINKING
2.50→3.06 | d=0.36 | p<.001

+10.1
AP CONTEXTUALIZATION (US HISTORY)
64.1→74.2 | d=0.39 | p<.001

Study Sample

Metric	Value
Total simulations analyzed	5,586
Unique students	2,227

Students with 3+ simulations (trait sample)	**790**
Distinct academic courses	5
Distinct academic subjects	4 (History, Economics, World History, Geography)
Frameworks evaluated	Course-specific content, an SEL framework, and a durable-skills competency framework
Scoring period	Jan 28 – Apr 28, 2026

Headline Findings

Framework	n	First	Last	Growth	Cohen's d	p
Academic content — AP-rubric overall (US History)	329	61.7	68.6	+7.0 pts	0.28	p<.001
SEL aggregate	771	2.79	3.10	+0.31	0.35	p<.001
Competency aggregate	788	2.03	2.38	+0.35	0.38	p<.001
Supplementary — durable-skill measures (US History)	390	75.1	80.7	+5.5 pts	0.24	p<.001

Growth is highly significant across all three measurement frameworks — academic content, SEL, and competencies (p<.001), with aggregate effect sizes of d = 0.24 to 0.38. Within US History, immersionED's own supplementary durable-skill measures also improve significantly (+5.5). For context: these results come from organic classroom use, not controlled research conditions, and under real classroom conditions educational interventions typically achieve effect sizes around d = 0.20 (Kingston & Nash, 2011). The AP-rubric and durable-skill measures are reported separately (see below) and are not directly comparable in absolute terms.

Per-Course Content Growth

Course	n	First → Last	Growth	d	p
US History (AP rubric)	329	61.7 → 68.6	+7.0 pts	0.28	p<.001
E4E Microeconomics	205	3.67 → 3.85	+0.17	0.34	p<.001
WWI History (C3-aligned)	134	2.09 → 2.36	+0.27	0.20	p=.024
C3 World History 1200CE	45	1.83 → 2.17	+0.34	0.40	p=.019
Geography (course rubric)	26	2.95 → 3.15	+0.21	0.14	p=.48 (n.s.)

4 of 5 academic courses show statistically significant content growth. Geography did not reach significance (p=.48) and its sample (n=26) is underpowered. US History is scored 0–100 on the AP rubric; the other four use their own 0–4 course rubrics, so absolute scores are not comparable across rows.

US History — AP U.S. History Rubric Skills

Transcripts re-scored by immersionED's AI scoring engine against the AP U.S. History framework (9 skills), first vs last occurrence per student (skill rated in 3+ sims). Small-n skills are footnoted rather than featured.

AP Skill	First	Last	Growth	d	p	n
Contextualization	64.1	74.2	+10.1 pts	0.39	p<.001	224
Argumentation	61.4	70.3	+9.0 pts	0.32	p<.001	237
Making Connections	74.4	81.7	+7.3 pts	0.54	p<.001	103
Sourcing & Situation	76.0	83.2	+7.2 pts	0.47	p=.003	46
Causation	66.8	73.9	+7.1 pts	0.33	p<.001	206
Developments & Processes	64.3	71.0	+6.7 pts	0.26	p<.001	308

Footnoted (small sample — rarely elicited by current episodes): Claims and Evidence in Sources +8.6 (d=0.53, n=24); Continuity and Change −0.3 (n=11, n.s.); Comparison +3.1 (n=7, n.s.). ***AP-rubric overall: 61.7 → 68.6, +7.0 (d=0.28, p<.001, n=329)*** *— the mean of whichever AP skills each session elicited, first vs last per student.*

US History — immersionED Durable-Skill Measures (Supplementary)

These are immersionED's own ***supplementary*** *durable-skill measures within US History —* ***not*** *AP skills, and not a separate measurement framework — reported separately so they are never conflated with AP performance.*

Durable Measure	First	Last	Growth	d	p	n
Ethical Reasoning	75.2	81.6	+6.3 pts	0.25	p<.001	388
Perspective	72.3	78.5	+6.2 pts	0.26	p<.001	381
Agency & Resilience	78.4	82.9	+4.6 pts	0.19	p<.001	371

Durable-skills overall: 75.1 → 80.7, +5.5 (d=0.24, p<.001, n=390).

> **AP and durable-skill scores are not directly comparable.** The AP figures come from a strict grade-11 AP re-score (0–100 on the AP rubric); the durable-skill figures use immersionED's separate skill scoring. They use different rubrics and scales — compare growth *within* each group, not absolute scores *between* groups.

US History — Academic Dose-Response (AP composite)

Compare	n	Δ (AP composite)	Cohen's d	p
s1 vs s2	606	+3.1	0.13	p=.002
s1 vs s3	329	+4.1	0.16	p=.005
s1 vs s4	190	+6.9	0.30	p<.001

s1 vs s5	135	+5.2		0.23	p=.007
s1 vs s6	94	+6.3		0.25	p=.019

Academic growth in AP historical-thinking skills is detectable by the 2nd session and substantial by the 4th — mirroring the SEL and competency dose-response below.

Social-Emotional Learning Growth (10 Outcomes)

Measured against a nationally recognized SEL framework with research-based adolescent development benchmarks. Each outcome scored independently per simulation.

SEL Outcome	n	First	Last	Growth	d	p
Teamwork	502	2.57	2.99	+0.42	0.35	p<.001
Concern for Others	420	2.70	3.09	+0.39	0.29	p<.001
Communicating About Problems	557	2.85	3.17	+0.32	0.29	p<.001
Standing Up for Others	361	3.07	3.35	+0.28	0.20	p<.001
Managing Emotions	177	3.40	3.66	+0.27	0.24	p=.002
Speaking Up Kindly	535	3.19	3.45	+0.26	0.24	p<.001
Resolving Solutions	569	3.04	3.30	+0.26	0.20	p<.001
Listening & Responding	800	2.68	2.92	+0.24	0.24	p<.001
Linking Feelings/Thoughts/Actions	485	3.43	3.56	+0.14	0.13	p=.005

9 of 9 (100%) measurable SEL outcomes show statistically significant growth. *Largest gains in interpersonal-cognitive skills (Teamwork, Concern for Others, Communicating About Problems), the skills directly exercised by simulation engagement.*

Durable-Skills Competency Growth (17 Sub-Skills)

Measured against a nationally recognized competency framework with research-based progressions, used under Creative Commons license. Each sub-skill scored independently per simulation.

Sub-Skill	n	First	Last	Growth	d	p
Persuasive Communication	776	1.87	2.50	+0.63	0.48	p<.001
Logical Thinking	678	2.50	3.06	+0.56	0.36	p<.001
Critical Dialogue	783	1.69	2.18	+0.49	0.41	p<.001
Building Empathy	447	2.30	2.70	+0.41	0.28	p<.001
Interpreting Information	741	2.52	2.89	+0.37	0.25	p<.001
Navigating Power	602	2.07	2.42	+0.35	0.31	p<.001

Negotiating Conflict	464	2.22	2.55	+0.33	0.26	p<.001
Synthesis	759	2.49	2.81	+0.32	0.23	p<.001
Productive Collaboration	371	1.83	2.14	+0.31	0.29	p<.001
Problem Seeking	768	1.97	2.24	+0.27	0.26	p<.001
Problem Solving	803	1.56	1.82	+0.27	0.24	p<.001
Diverse Perspectives	213	2.52	2.74	+0.22	0.15	p=.025
Making Meaning	466	1.86	2.03	+0.18	0.14	p=.002
Recognizing Conflict	303	2.67	2.79	+0.12	0.14	p=.015
Healthy Relationships	306	2.20	2.23	+0.03	0.03	p=.59 (n.s.)

16 of 17 (94%) sub-skills show statistically significant growth. *Communication skills dominate (Persuasive Communication d=0.48, Critical Dialogue d=0.41), the highest-effect-size sub-skills in the study. Healthy Relationships is the lone non-significant result (d=0.03).*

How Quickly Does Growth Show Up?

Compare	SEL Δ	SEL d	Comp Δ	Comp d
s1 vs s2	+0.13	0.15	+0.09	0.12
s1 vs s3	+0.24	0.27	+0.12	0.15
s1 vs s4	+0.32	0.37	+0.25	0.30
s1 vs s5	+0.32	0.36	+0.33	0.39
s1 vs s8	+0.40	0.47	+0.56	0.62
s1 vs s10	+0.63	0.72	+0.62	0.78

Growth detectable from session 2. Substantial (d > 0.30) by session 4. By session 10, effects are large by Cohen's conventions (d > 0.7) on both SEL and competencies. The academic (AP) dose-response above shows the same pattern.

Time to Impact: immersionED vs Typical Interventions

Dimension	immersionED	Typical intervention
Effect size achieved	d = 0.24 to 0.48	d ~ 0.20 under real classroom conditions (Kingston & Nash, 2011)
Active student dose	~1.5 to 2 hours (~5 sessions × ~20 min)	40–100+ hours
Elapsed calendar window	Median 22 days (3.1 weeks)	15–30 weeks

When growth is detectable	From session 2 (~20 min)	Multi-week pre/post window
Frameworks improved	3 simultaneously	Usually 1
Dose for large effects (d>0.7)	~10 sessions (~3 hours)	Rarely achieved

Roughly an order-of-magnitude compression. The median student reaches above-benchmark effect sizes in ~3 weeks and ~2 hours of simulation time, versus 15–30 weeks and 40–100+ hours typically.

Three Findings for Positioning

1. Growth is real and simultaneous. Students don't trade off content learning against SEL or competency development. All three frameworks move together. AP-rubric +7.0 pts, SEL +0.31, competencies +0.35 — plus immersionED's supplementary durable-skill measures (+5.5) within US History — all p<.001.

2. Effect sizes exceed the real-world benchmark. Under real classroom conditions, educational interventions typically achieve d ~ 0.20 (Kingston & Nash, 2011). Every aggregate in this study (d = 0.24 to 0.38) clears that benchmark, and the top sub-skills — Persuasive Communication (d=0.48) and Critical Dialogue (d=0.41) — more than double it, from roughly 1.5 to 2 hours of total simulation time.

3. Dose-response is favorable for adoption. Growth is detectable from session 2 and reaches medium effect size by session 4 — on academic content, SEL, and competencies alike. Schools should expect measurable gains within the first few sessions.

Caveats. *Observational, not causal; students who complete more simulations may differ systematically. The ~2 hours figure is active simulation time only; classroom instruction between sessions may contribute to growth. The same AI scoring system was used across all frameworks (in independent scoring runs); external assessments might show different magnitudes. US History content was re-scored against the AP U.S. History framework; the AP-rubric and durable-skill measures use different rubrics and are reported separately. Findings are preliminary and form the basis for formal university-led research now being designed.*

Reference. *Kingston, N., & Nash, B. (2011). Formative Assessment: A Meta-Analysis and a Call for Research. Educational Measurement: Issues and Practice, 30(4), 28–37.*