Title: Supporting content-area learning in Biology and U.S. History: A randomized control trial of *Enhanced Units* in California and Virginia

Authors and email addresses

Hannah D'Apice, hdapice@empiricaleducation.com
Adam M. Schellinger, aschellinger@empiricaleducation.com
Jenna Zacamy, jzacamy@empiricaleducation.com
Xin Wei, xin.wei@sri.com
Andrew P. Jaciw, ajaciw@empiricaleducation.com

Conference Section:
Academic Learning in Education Settings
Teachers and School Leaders

Keywords:
Educational evaluation
Multi-site randomized trial
Special populations

Background

Advancing educational standards (e.g., Common Core State Standards, Next Generation Science Standards), emphasis on academic language, and demands to improve students' higher-order reasoning skills have raised expectations for all students, and created challenges for students with disabilities. In response, and through an Investing in Innovation (i3) development grant, SRI International utilized a researcher-practitioner partnership to develop *Enhanced Units (EU). EU* combines research-based content enhancement routines, collaboration strategies and technology components for secondary history and biology classes. The goal is to improve student content learning and higher order reasoning, especially for students with disabilities (Table 1). *EU* was developed during a two-year design-based learning implementation process (Penuel & Martin, 2015) with teachers and administrators co-designing the units with developers.

This paper presents findings from a randomized control trial (RCT) measuring the impact of *EU* on student learning in three districts in Virginia and California. To support this development grant, researchers focused on exploratory analysis to help developers unpack results and find ways to improve the program.

Research Questions

The confirmatory research questions addressed the impact of *EU* on student learning in (1) biology, (2) history, and (3) biology and history combined. We also addressed if there was a differential impact of *EU* for students with disabilities. Additionally, we addressed several exploratory questions to better understand the results, including: if the impact differed for students and teachers based on background characteristics; if the impact differed by biology content area (Evolution/Ecology) or history content area (World War II/Cold War); and if there was evidence of impact on instructional practices that were posited to mediate student learning.

Setting, Research Design, and Participants

This study was implemented in five high schools across three districts in two states during the 2018 spring semester. To maximize power, the unit of random assignment was the classroom level within teachers. Overall, the study involved 13 teachers, 14 randomized blocks, and 30 classes (15 in each condition, with 18 in biology and 12 in history) (Table 2 and 3).

Data Collection and Measures, Analysis, and Findings

The unit tests in biology and history were written following general curriculum standards in Virginia and California. Each test demonstrated internal consistency of 0.75 or above. Fidelity of implementation and classroom practices were measured by teacher surveys and logs.

Two-level hierarchical linear modeling (HLM) was used to estimate the impact of the intervention (Raudenbush & Bryk, 2002). The first level was student and the second level was classrooms. HLM controlled for student characteristics (e.g., gender, race, ELL, special education status) and school dummy variable to reduce residual error.

The study had low overall and differential attrition (Table 4). We found a positive impact of *EU* on student learning in history, but not on biology or across the two domains combined (Table 5). Within biology, we found that students experienced greater impact on the Evolution unit than the

Ecology unit. The difference was .171 standardized effect size units (t=2.00) and was marginally statistically significant (p=.063). We also found a positive differential effect favoring students with disabilities, which is an encouraging result given the goals of the grant. We did not find any other differential impacts. We observed very little difference between *EU* and control classes in use of 17 mediating instructional practices, with values very close to zero (Table 6).

Conclusions

Based on the results, we posed several questions about implementation and impact.

Were conditions to support seeing an impact present? We found that certain conditions for impact were satisfied. The treatment-control contrast was strong, with limited spillover in the use of *EU* by teachers in their control classes. However, other results made it harder to explain the mechanism behind the observed impact. Implementation did not reach thresholds for levels of fidelity system-wide. Furthermore, though the study was too underpowered to conduct a formal mediation analysis, we saw little difference between *EU* and control in instructional practices identified as mediators of impact on student learning.

*Why did we observe a positive impact in history but not Biology***?** A theory developed by the program developers posits that *EU* works especially well with content that progresses in a sequential and linear way, as history does. We were able to further test this hypothesis by examining whether, within biology, we would see a greater impact for Evolution unit than for the Ecology unit, as the content and routines in the former were structured in a more-sequenced way. Our findings support this hypothesis.

Where can program improvement efforts be focused? If *EU* works better with logically sequenced material, then the obvious place to focus improvement is with less-structured content. We should seek program development to support impact for material with wide-ranging structures, and possibly introductory elements that systematically link content to previously learned content.

Interviews with teachers revealed operational challenges with the technology component of the intervention, which had acute effects on students' experiences. Teachers had several conditions under which they felt the program would be successful: (1) teachers receiving adequate training in the routines, (2) classroom access to computer devices, (3) teacher discretion over pairing routines with topics taught, (4) routines being done on paper instead of through the technology component, and (5) *EU* being used in AP or upper-level classes.

Future improvements to *EU* should focus on: *"What are the best ways for teachers to present the routines to their students?"* While the first two points listed above can be addressed through adequate implementation, the remaining points invite further exploration by the developers. Assuming teachers have greater discretion over the choice of routines, developers should investigate how the routines can be applied to a greater range of topics. In regards to the routines being done on paper instead of through the technology component, developers should consider how introducing devices to the routines potentially presents steeper learning curves and difficulty with buy-in. Finally, with respect to the suggestion that *EU* be used in upper-level classes, developers should consider how the suggestion can be balanced with the priority of the grant: improving academic outcomes for students with disabilities.

References

Penuel, W. R., & Martin, C. (2015, April). DBIR as a Strategy for Expanding Opportunity to Learn. Paper presented NCTM, Boston, MA.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

TABLE 1. LOGIC MODEL FOR I3 *EU* STUDY

INPUT	PROXIMAL OUTPUTS	LONGER TERM OUTPUTS	TEACHER OUTCOMES	STUDENT OUTCOMES
<u>Classroom level</u> Biology and U.S. History teachers receive curricular materials for i3 *EU* **Key Component 1:** Biology and U.S. History teachers receive sufficient support: In-Person PD: Biology and U.S. History teachers receive sufficient support to use i3 *EU* materials by attending 3 days of PD Ongoing coaching: Biology and U.S. History teachers receive sufficient support by receiving at least 8 hours of coaching from SIM professional developers	→ <u>Teacher</u> **Key Component 2:** Teacher use of *EU*: Biology and U.S. History teachers use *EU*. Biology and U.S. History teachers implement one practice *EU* and two study *EU*s as per study design. Teachers deliver quality instruction, adhere to dosage, and report on likely effectiveness of the intervention on student performance. ↓ <u>Student</u> Students understand the purpose and application of the *EU*s in their biology and U.S. History classes	→ <u>Teacher</u> Improved implementation of, adherence to, and quality of *EU* instructional practices; improved effectiveness of *EU*	→ <u>Teacher</u> Improved/ increased implementation of SIM strategies (particularly the content enhancement routines specified in *EU*)	<u>Student</u> Improved achievement on end-of-unit content assessment measures

TABLE 2. CONFIGURATION OF BLOCKS AND CLASSES IN THE EXPERIMENT: BIOLOGY SAMPLE

District	School	Biology teacher	Block	Condition 0 = Control; 1 = *EU*	Class
1	1	1	1	0	1
				0	2
				1	3
	2	2	2	0	4
				1	5
			3	0	6
				1	7
2	3	3	4	0	8
				1	9
	4	4	5	0	10
				1	11
		5	6	0	12
				1	13
3	5	6	7	0	14
				1	15
		7		1	16
			8	0	17
				1	18

TABLE 3. CONFIGURATION OF BLOCKS AND CLASSES IN THE EXPERIMENT: U.S. HISTORY SAMPLE

<table>
<tr><th>District</th><th>School</th><th>U.S. History teacher</th><th>Block</th><th>Condition
0 = Control; 1 = EU</th><th>Class</th></tr>
<tr><td rowspan="8">2</td><td rowspan="4">3</td><td rowspan="3">8</td><td rowspan="2">9</td><td>0</td><td>19</td></tr>
<tr><td>1</td><td>20</td></tr>
<tr><td rowspan="2">10</td><td>0</td><td>21</td></tr>
<tr><td>9</td><td>1</td><td>22</td></tr>
<tr><td rowspan="4">4</td><td>10</td><td rowspan="2">11</td><td>1</td><td>23</td></tr>
<tr><td rowspan="3">11</td><td>0</td><td>24</td></tr>
<tr><td rowspan="2">12</td><td>0</td><td>25</td></tr>
<tr><td>1</td><td>26</td></tr>
<tr><td rowspan="4">3</td><td rowspan="4">5</td><td rowspan="2">12</td><td rowspan="2">13</td><td>0</td><td>27</td></tr>
<tr><td>1</td><td>28</td></tr>
<tr><td rowspan="2">13</td><td rowspan="2">14</td><td>0</td><td>29</td></tr>
<tr><td>1</td><td>30</td></tr>
</table>

TABLE 4. ATTRITION COUNT OF POSTTEST IN BIOLOGY AND U.S. HISTORY SUBJECTS

	Count of students at baseline	Students with Unit 2 posttest	Attrition	Students with Unit 3 posttest	Attrition	Students with both Unit 2&3 posttest	Attrition (both outcomes)	Students with Unit 2 posttest only (and not Unit 3)	Students with Unit 3 posttest only (and not Unit 2)	Students with either Unit 2 OR Unit 3 posttest	Attrition (either outcome)
Biology											
***EU* (N)**	194	170	12.4%	174	10.3%	163	16.0%	7	11	181	6.7%
Control (N)	219	198	9.6%	198	9.6%	186	15.1%	12	12	210	4.1%
Total N	413	368		372		349		19	23	391	
Overall attrition			10.9%		9.9%		15.5%				5.3%
Differential attrition			2.8%		0.7%		0.9%				2.6%
Potential for bias			low		low		low				low
U.S. History											
***EU* (N)**	111	105	5.4%	107	3.6%	103	7.2%	2	4	109	1.8%
Control (N)	128	122	4.7%	120	6.3%	115	10.2%	7	5	127	0.8%
Total N	239	227		227		218		9	9	236	
Overall attrition			5.0%		5.0%		8.8%				1.3%
Differential attrition			0.7%		2.6%		2.9%				1.0%
Potential for bias			low		low		low				low

TABLE 4. ATTRITION COUNT OF POSTTEST IN BIOLOGY AND U.S. HISTORY SUBJECTS

	Count of students at baseline	Students with Unit 2 posttest	Attrition	Students with Unit 3 posttest	Attrition	Students with both Unit 2&3 posttest	Attrition (both outcomes)	Students with Unit 2 posttest only (and not Unit 3)	Students with Unit 3 posttest only (and not Unit 2)	Students with either Unit 2 OR Unit 3 posttest	Attrition (either outcome)
Biology and U.S. History											
***EU* (N)**	305	275	9.8%	281	7.9%	266	12.8%	9	15	290	4.9%
Control (N)	347	320	7.8%	318	8.4%	301	13.3%	19	17	337	2.9%
Total N	652	595		599		567		28	32	627	
Overall attrition			8.7%		8.1%		13.0%				3.8%
Differential attrition			2.1%		0.5%		0.5%				2.0%
Potential for bias			low		low		low				low

TABLE 5. OVERALL INTENT-TO-TREAT IMPACT ANALYSIS ON **BIOLOGY AND U.S. HISTORY OUTCOME**

	Condition	Means	Standard deviations[a]	No. of posttest scores	No. of students	No. of teachers	Effect size	*p* value	Change in percentile ranking
Biology									
Unadjusted effect size[a]	Control	70.77	22.45	396	210	9	0.01	.958	0%
	EU	71.00	22.19	344	181	9			
Adjusted effect size[b]	Control	70.77					0.01	.892	0%
	EU	71.02							
U.S. History									
Unadjusted effect size[a]	Control	49.39	22.16	242	127	6	0.33	.214	12%
	EU	56.51	20.40	212	109	6			
Adjusted effect size[b]	Control	49.39					0.32	.037	12%
	EU	56.18							
Biology and U.S. History combined									
Unadjusted effect size[a]	Control	62.66	21.97	638	337	13	0.14	.516	6%
	EU	65.65	21.53	556	290	13			
Adjusted effect size[b]	Control	62.66					0.14	.067	6%
	EU	65.77							

[a] The unadjusted effect size is the regression-adjusted impact estimate in a model without covariates divided by the pooled standard deviation in outcomes.

[b] The adjusted effect size estimate is the point estimate for impact from the benchmark model divided by the pooled standard deviation of the outcome variable.

Note. Full estimates corresponding to the impact models are reported in Appendix I.

Source. Empirical Education staff calculations

TABLE 6. MEAN AND MEDIAN DIFFERENCES (*EU* – CONTROL) IN ORDINAL RESPONSES TO FREQUENCY OF USE OF EACH OF 17 POTENTIAL MEDIATING PRACTICES

		Mean			Median		
	Description	**Overall N = 13**	**U.S. History N = 6**	**Biology N = 7**	**Overall N = 13**	**U.S. History N = 6**	**Biology N = 7**
1	**Explicit instruction**	.02 (*p*=.625)	.00	.04	.00	.00	.00
2	**Reteach to a few students**	.02 (*p*=.875)	-.25	.25	.00	.00	.00
3	**Identifying similarities/differences (non-SIM)**	-.10 (*p*=.322)	.00	-.18	.00	.00	.00
4	**Explicit strategy for asking clarifying questions (non-SIM)**	.21 (*p*=.781)	.33	.11	.00	.50	.00
5	**Explicit summarizing strategy (non-SIM)**	.10 (*p*=1.00)	.25	-.04	.00	.25	.00
6	**Explicit paraphrasing strategy (non-SIM)**	.37 (*p*=.424)	.17	.54	.50	.25	.50
7	**Explicit vocabulary strategy (non-SIM)**	.25 (*p*=.359)	.17	.32	.00	.25	.00
8	**Graphic organizer (non-SIM)**	.12 (*p*=1.00)	.00	.21	.00	.00	.00
9	**Note-taking technique**	.15 (*p*=.625)	.25	.07	.00	.25	.00
10	**Mnemonic device for remembering information**	-.04 (*p*=.375)	.00	-.07	.00	.00	.00
11	**Rehearsing information aloud**	.35 (*p*=.156)	.50	.21	.00	.25	.00
12	**Teacher laptop or Chromebook**	-.12 (*p*=.375)	.00	-.21	.00	.00	.00
13	**Student laptop or Chromebook**	.17 (*p*=.906)	.25	.11	.00	.00	.00
14	**Student tablet**	.13 (*p*=.750)	-.08	.32	.00	.00	.00
15	**Student collaboration on group and partner assignments**	.31 (*p*=.250)	.17	.43	.00	.00	.00
16	**Teaching higher-order course content**	.25 (*p*=.688)	.33	.18	.00	.00	.00
17	**Support for learners with different abilities**	.17 (*p*=.563)	.00	.32	.00	.00	.00

Note. Teacher responses were on an ordinal scale: Never, Seldom, Sometimes, Often, and Always; *p* values for differences in means are based on Wilcoxon Signed Rank Test. Given the small samples involved, any inferential test will be underpowered. We observe very little difference between EU and control in their practices, with values very close to zero compared to maximum possible differences ranging between -4 and +4.)

Source. Empirical Education staff calculations