Title: Use of Design Based Implementation Research to Redesign Secondary Courses to Improve Higher Order Thinking and Content Learning in Biology and US History for Underperforming Adolescents

Authors and email addresses

Ellen Schiller (ellen.schiller@sri.com)
Jose Blackorby (jblackorby@cast.org)
Xin Wei (xin.wei@sri.com)
Jan Bulgren (jbulgren@ku.com)
Kathryn Morrison (Kathryn.morrison@sri.com)
Tejaswini Tiruke (tejaswini.tiruke@sri.com)

Abstract

Background

Whether states are implementing their own new standards, or the NGSS or the CCSS, the current landscape increases the demands on students to develop higher-order reasoning skills. The NGSS expect students to learn science concepts and investigate the natural world through scientific inquiry. In parallel, the CCSS expect students to integrate knowledge and ideas using primary sources and causal reasoning and to delineate claims and evaluate arguments. Ultimately, a goal for all new standards is for students to learn to apply higher-order thinking, a challenge for all students, but especially underperforming students.

SRI, University of Kansas, CAST, Culpepper (VA) and Alameda (CA) schools developed an intervention targeting higher order skills in social studies and science for underperforming adolescents. The strategy was to combine the KU's Content Enhancement Routines (CER) with technologies to support both academic content learning and higher order thinking skills (Bulgren & Ellis, 2012). Over 2-years, the research and school based team employed a Design Based Implementation Research (DBIR) methodology (Penuel & Martin, 2012) to benefit from the expertise of teachers, and students to select and refine CERs, academic content, and technology features, called Enhanced Units (EUs).

Research Question

1. What are the impacts of intervention on improving high school students' end of unit test scores in biology and U.S. History?

Setting

This study was implemented in two large high schools (rural and urban) in VA and CA during the 2017 spring semester. District demographics show a high level of student need, with an average 50% minority students, 37% socio-economically disadvantaged, 13% English Language Learners, and 10% SWD.

Participants

Fifteen science or social studies teachers and their 650 high school students in grades 9 to 12.

Intervention

The intervention leverages the CERs with affordances of modern technologies to support both content learning and higher order thinking skills in U.S. history and biology courses, focusing on underperforming students. Over 2-years, the research team followed a DBIR process including an iterative series of design meetings, focus groups, and technology tryouts. Specifically, teachers and students co-created the curricular units to incorporate the CERs and technology for specific history and biology units by: participating in 15 full or half day researcher-practitioner design meetings; 5 student focus groups; 6 technology pilots; and 54 class periods in which teachers piloted the EUs at multiple stages of development. The resulting three-part intervention is comprised of: (1) four science units (cells, ecology, evolution, genetics) and four social studies units (the cold war, the depression, the roaring 20s, world war II), each enhanced with CERs including Cause and Effect, Question Exploration, Compare and Contrast, and a Unit Organizer. (2) A Google App, called CORGI, to support instructional delivery of the enhanced units. CORGI supports familiar Google functions including shared authoring and commenting;

employs the same graphic designs as the original paper-based CERs, and embedded videos about how to use the routines, models of expert examples, text to speech and speech to text, and support for vocabulary and translation. (3) Professional development modules were developed to train teachers how to implement the EUs with CORGI in whole class and small groups.

Research Design

Blocked by school and subject, teachers were randomly assigned to either treatment or business-as-usual conditions. Treatment teachers implemented the EUs using CORGI. Control teachers were provided the goals for each unit and taught using their typical instruction.

Data Collection and Measures, Analysis, and Findings

Students completed DBIR team developed end of unit tests. All *unit test items* in biology and history demonstrated internal consistency of 0.70 or above. Student achievement on state standardized accountability tests were used as *pretest* measures, which had a correlation with unit test close to 0.60. *Fidelity of Implementation* was measured by teacher implementation logs and surveys. Implementation logs recorded teacher usage of intervention by components. Treatment and control teachers answered survey questions about intervention specific instructional practices, technology use, and student grouping and collaboration.

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

The attrition analysis revealed a low overall attrition and differential attrition. The preliminary HLM analyses suggest a significant difference favoring treatment over control groups on two of the four science unit tests ($\beta = 0.15$, *s.e.* = 0.04, $p < 0.001$; $\beta = 0.05$, *s.e.* = 0.02, $p < 0.05$) with hedge's g effect sizes of 0.95 and 0.25. This study also finds a significant treatment effect on improving one of four social studies unit test ($\beta = 0.25$, *s.e.* = 0.03, $p < 0.001$) with an effect size of 1.45.

The fidelity of implementation analysis revealed that treatment teachers reported significantly higher use of intervention practices and other instructional practices than control teachers. Furthermore, teachers' higher use of intervention practices ($\beta = 0.02$, *s.e.* = 0.008, $p < 0.05$) and other instructional practice scores ($\beta = 0.009$, *s.e.* = 0.004, $p < 0.05$) were significantly associated with students' higher end of unit test scores after controlling for student baseline score, demographic characteristics, and treatment status. These mediation analyses showed that the intervention improved student outcomes through increased use of intervention practices. Additional analysis on implementation logs will be presented in the full paper.

Conclusions

The EU intervention, a combination of the CER and the CORGI Google technology improved student achievement on biology and U.S. history unit tests in spring semester 2017. The analysis on the mediating effects of fidelity of implementation emphasized the importance of teacher use of the intervention practices and other instructional practice by following implementation guidelines created during the DBIR process in achieving desirable student reading outcomes. In 2017-18, a larger cluster RCT will test the efficacy of this intervention and provide a more complete understanding of the mechanism of how the intervention works for all students as well as those who are underperforming.

Appendix
References

Bulgren, J., & Ellis J. (2012). Argumentation and evaluation intervention in science classes: Teaching and learning with Toulmin. In *Perspectives on scientific argumentation*, 135–54. New York: Springer.

McGuinn, P. (February, 2016). *ESSA and the shifting landscape at the US Department of Education.* Washington, D.C.: Brookings.

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.

What Works Clearinghouse (2014). *What Works Clearinghouse procedures and standards handbook* (Version 3.0). Retrieved from http://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_v3_0_standards_handbook.pdf

Table 1. Attrition Analysis for Clusters/Teachers

Group	Number of Teacher Randomized	Number of Teacher at the end of the study	Attrition Rate
Treatment	7	7	0
Control	7	7	0
Total	14	14	0

Table 2. Attrition Analysis for Students

Group	Number of students when Teachers were randomized	Number of students in the analytic sample (with both pretest and posttest)	Attrition Rate
Treatment	364	306	16%
Control	285	268	6%
Total	649	574	10%

Table 3. Student Test scores by Condition

Variables	Treatment			Control		
	N	Mean	SD	n	Mean	SD
Pretest	306	3.19	0.88	268	3.00	0.85
Science Unit Test						
Cell	20	0.87	0.13	191	0.68	0.16
Ecology	80	0.56	0.23	207	0.58	0.21
Evolution	108	0.64	0.23	217	0.63	0.22
Genetics	80	0.75	0.2	197	0.69	0.2
History Unit Test						
The cold war	126	0.76	0.12			
The depression	183	0.64	0.16	47	0.47	0.17
The roaring 20s	128	0.7	0.17	51	0.45	0.18
World War II	187	0.63	0.22	46	0.46	0.2

Note. A student can take up to 4 unit tests in a subject. Sample size for treatment students who took at least one unit test is 306. Sample size for control students who took at least one unit test is 268.

Table 4. Student Demographic Characteristics by Condition

Variables	Treatment (n=306)	Control (n=268)
	%	%
Grade 9	30%	60%
Grade 10	8%	18%
Grade 11	60%	21%
Grade 12	2%	1%
Male	56%	53%
Asian	30%	30%
African American	13%	13%
Hispanic	11%	14%
Others	6%	4%
English Learners	6%	7%
Free or reduced priced lunch	25%	29%
Special Education	10%	5%

Table 5. Overall Intent-To-Treat Impact Analysis of Intervention on Student Unit Test Scores

Outcome Measures	Estimated Impact	s.e.	Effect Size
Science Unit Test			
Cell	0.15 ***	0.04	0.95
Ecology	-0.05	0.02	-0.23
Evolution	-0.02	0.02	-0.09
Genetics	0.05*	0.02	0.25
History Unit Test			
The cold war	No estimate	NA	NA
The depression	0.05	0.04	0.31
The roaring 20s	0.25***	0.03	1.45
World War II	-0.03	0.04	-0.14

Note. Estimated impact is the coefficient associated with treatment variable from the HLM model; Effect size = Estimated impact/pooled SD of the treatment and control group. HLM impact models control for pretest and demographic variables using unimputed dataset.

$*p < .05$, $**p < .01$, $***p < .001$.

NA = Not Available because no control students who took this unit test.

Table 5. Fidelity Score by condition

Variables	Treatment			Control			T value
	n	Mean	SD	n	Mean	SD	
SIM instructional Practice	17	2.4171	0.5732	16	1.6023	0.5862	4.04***
Other instructional Practice	17	2.93	0.35	16	3.4792	0.4605	3.87***
Technology Use	17	3.3309	0.484	16	3.2813	0.6793	0.24
Collaboration	17	3.3039	0.4092	16	3.4896	0.6483	0.99

***$p < .001$.