STUDENT RESULTS FROM AN EFFICACY TRAIL OF A NEW NGSS EVOLUTION UNIT THAT INTEGRATES HEREDITY

The Problem

An understanding of evolution is fundamental to biology (NRC, 2012), yet students have a poor grasp of this essential science idea (reviewed in Gregory, 2009). Students misunderstand common ancestry (e.g., Flanagan & Roseman, 2011), misinterpret evolutionary trees (Meir, Perry, Herron, & Kingsolver, 2007), and poorly understand evolutionary time (Catley, Novick, & Shade, 2010). The most common student-held alternative conceptions about natural selection are rooted in misunderstandings about heredity (Nehm & Schonfeld, 2008). The genetic mechanisms of mutation and random variation—key to understanding evolution—are particularly difficult for students to grasp (Morabito, Catley, & Novick, 2010). Therefore, researchers have called for a stronger genetics connection in students' study of evolution (e.g., Catley et al., 2010; Mead, Hejmadi, & Hurst, 2017). However, few widely produced curriculum materials foster this integration, preventing students from easily making the essential conceptual connections between heredity and evolution (e.g., Postlethwait & Hopson, 2009). Researchers also advocate for integrating science practices that can foster student learning of evolution such as analyzing and interpret data (e.g., Catley, Lehrer, & Reiser, 2004) and arguing from evidence (e.g., Asterhan & Schwarz, 2007). Students' content understanding increases when argumentation is an explicit part of instruction (e.g., Zohar & Nemet, 2002).

In response to the calls for new curriculum materials that integrate genetics and evolution along with the three dimensions of the *Next Generation Science Standards* (NGSS) (NGSS Lead States, 2013), the project team has developed and rigorously tested the *Evolution: DNA and the Unity of Life* curriculum unit for introductory high school biology. Here we briefly describe the unit, its multi-year development process, and present results from a randomized controlled trial (RCT) of the unit. We focus our presentation on the central research question: Compared to a business-as-usual control curriculum unit, what is the effect of the *Evolution: DNA and the Unity of Life* curriculum unit on students' knowledge of evolution and ability to argue from evidence?

Theoretical Framing of the Unit

The freely-available *Evolution* unit is based on the assumption that students will better understand the disciplinary core ideas (DCIs) about biological evolution (LS.4A, B, C & D) when curriculum materials and instruction: (1) integrate the DCIs about LS.3 heredity (inheritance and genetic variation) that are essential for understanding evolution; (2) build students' abilities in the science practice of analyzing and interpreting skill-level-appropriate data about phenomena from published scientific research; (3) engage students in the construction of evidence-based arguments; and (4) frame ideas through crosscutting concepts. Our unit development framework drew on constructivist and conceptual change theories of learning (e.g., Driver, 1995; Strike & Posner, 1992) and was informed by multiple learning progressions (e.g., Berland & McNeill, 2010).

Through online and paper-based lessons, the five-module unit engages students in 3D learning. The unit explores DCIs around the shared biochemistry of life, common ancestry, heredity, natural selection, and speciation in non-human examples, focusing on DNA as a blueprint for all living things. Students engage in the practices of using models, analyzing and interpreting real data, and constructing arguments from evidence. To learn the latter practice, students use a heavily scaffolded CER framework to develop and evaluate arguments over the course of the modules. The unit incorporates the crosscutting concepts of patterns and cause and effect. For details on the unit's theoretical framework, curriculum descriptions, and pilot testing, see Authors (2019).

Efficacy Trial Research Design

Development and testing of the unit followed an iterative, multi-step, multi-year process. Both the student and teacher materials were refined through multiple rounds of testing in introductory biology classrooms, with revisions during and following each round. This presentation will focus on the final round of testing, the efficacy trial.

Student Assessment Development and Data Analysis

Student pre/post assessment items were developed by a project partner who nationally pilot tested and revised the items according to established procedures (Authors, 2014) and administered the test. The assessment tasks were aligned with the learning goals for the unit and associated dimensions of *A Framework for K-12 Science Education* (NRC, 2012) and NGSS. They incorporated published scientific data but did not use the same phenomena as the lessons. Multiple choice (MC) and constructed response (CR) items were pilot tested nationally with 200 to 4,588 students, depending on the testing stage.

In the efficacy trial, students completed different test forms for pre and posttesting. Items were distributed across four forms for both pretest and posttest. Each form contained 26 MC items (the average test difficulty was similar and each test contained approximately the same number of items on each topic) and two CR items that assessed students' ability to write an argument. Students received the same CR item in pre and posttesting that used common ancestry as the content area; and received different CR items that used natural selection as the content area. At each testing stage, Rasch analysis was conducted using WINSTEPS (Linacre, 2016) to obtain measures of item difficulty and student performance level, and to determine reliability of the measures.

Hierarchical linear modeling (HLM), *t* tests, and ANOVAs were used to analyze the efficacy trial data. To analyze the significance of the treatment on students' gains, the data were fit to the a mixed-effect HLM model. Demographic variables (gender, ethnicity, whether English was their primary language) and treatment with the unit were treated as fixed effects, while teachers were treated as a second level random effect. Rubrics for the CR items were created on the CER framework. Students were awarded up to 12 points per item based on the specificity, sophistication, and clarity of their answers. Three graders independently scored a subset of student responses to establish .80 or higher interrater reliability and to make rubric revisions.

Early Unit Development and Testing

After conceptualizing the goals, content, and flow of a full evolution unit, the project team first developed one module on a central topic in evolution, Natural Selection, which became the fourth module in the unit. The module underwent four rounds of local and national classroom alpha testing and revision in 2013 followed by a national pilot test with 461 students. The remaining modules were developed and sections were classroom tested and revised. An external reviewer assessed the unit for alignment to the EQuIP rubric (Achieve Inc., 2016), and the unit was refined based on these results. During the 2016-2017 school year, the full unit was nationally pilot tested with 20 introductory biology teachers and their 943 students who completed both the pre and posttest, with the goal of gathering information to refine and improve the unit. See Authors (2019) for a full description of the pilot test procedures and results.

Unit Efficacy Testing (Randomized Controlled Trial)

The final phase of curriculum testing was a national efficacy field test conducted during the 2017-2018 school year that used a randomized controlled trial (RCT) design. The field test was designed to compare pre/post learning gains made by students whose teachers were randomly assigned to either the treatment (new unit) or control (business-as-usual in NGSS-adopting

classrooms, schools, or districts) condition. The units used in both conditions targeted the same NGSS DCIs for evolution (HS-LS4A: Evidence of Common Ancestry; HS-LS4B: Natural Selection; and HS-LS4C: Adaptation).

Participant recruitment.

Forty-six teachers (22 in treatment, 24 in control), representing 23 states and diverse teaching contexts, diverse student demographics (socioeconomic, linguistic, ethnic, and racial diversity), and teaching schedules (alternating vs. daily) were recruited via the curriculum developer's email list of over 20,000 educators nationwide. Inclusion criteria used for teacher recruitment and selection included: teaching in an NGSS-oriented classroom, school, and/or district; teaching at least two introductory biology sections; access to internet-enabled technology; and a willingness to teach their heredity unit after evolution if assigned to the control condition.

Participation and procedures.

Teachers were randomly assigned to a condition by the project's external evaluator. Notifications were emailed to the teachers by the project team, and communications between the teachers and project staff occurred regularly throughout the year. Only the student scores of teachers who completed the participation requirements are reported in this analysis. University IRB and school district approval (when applicable) were received prior to conducting research. Teacher procedures were: Engage in a 1 - 1.5-hour webinar training on the materials in summer 2017; teach the entire new unit in each biology section using no external materials (treatment condition only); teach their regular evolution unit in each biology section (control condition only); administer online student pre and post assessments; complete daily teacher logs and additional surveys to evaluate for fidelity of implementation.

Student Assessment Results

Data Inclusion Criteria

The results represent data from 38 teachers (n=19 treatment, n=19 control) who completed the research requirements and remained in the analytic sample and their 2,269 students (n=1,165 treatment, n=1,094 control) who completed both the pre- and posttest. Student demographics were: 50% female, 9% English not primary language, 36% F/R lunch, and 36% from underrepresented ethnic or racial groups. No significant differences in the demographic categories were found between conditions. The overall study attrition was 17%, with 14% attrition from the treatment condition and 21% attrition from the control. The differential attrition rate of 7% compared to the overall attrition of 17% meets the optimistic boundary for attrition by the What Works Clearinghouse (WWC, 2018), and is lower than the maximum allowable attrition applied to past reviews of science interventions (WWC, 2012).

Results

Results from the MC items, reported here using Classical test theory (and corroborated with Rasch and HLM modeling), indicated that the treatment students had significantly higher gains (M=18.3%, SD=17.5%) over those in the control group (M=8.4%, SD=16.1%) on topics in evolution, $t(2,267) = 13.9, p < .001$, with a Cohen d effect size of .58. The treatment effect on gains was found to be approximately 7% of the increase in a student's gains. Gender, ethnicity, and English as primary language were not found to have significant effects on student gains at the $p < .05$ threshold. Variance in student gains across teachers only explained about 8% of the variance in the data, a relatively small amount.

Overall, factor analysis and Rasch analysis do not provide strong evidence that the assessment items are measuring multiple distinct evolution topics (natural selection, common ancestry, and speciation). Therefore, while we will provide results per topic at the conference,

these results are less reliable than results reported unidimensionally (when the topics are combined).

CR item analysis of student skills in constructing arguments on the natural selection items revealed that students in the treatment group earned statistically significantly more points on the posttest than the control students on all 12 rubric elements except two that required citing data. When comparing percentage point gains on these rubric elements, however, the treatment group had greater gains than the control group: for the two citing data elements, the treatment group improved by 4 percentage point for each; while the control group increased by 1 percentage point for each. For this item, pre/post percentage gains could not reliably be measured statistically because the items are not identical.

For the common ancestry CR item, results from pre and posttest percentage gains indicated that the treatment group had higher gains than the control group for summarizing data, citing data as evidence, and using basic reasoning to link genetic similarity to common ancestry. Statistically meaningful gains were not observed between conditions in students' ability to detail the mechanisms that links genetic similarity to common ancestry.

In all, these findings provide support for the instructional approach used in the unit. Further, they suggest that the unit improves students' understanding of the targeted evolution ideas and in their ability to construct arguments from evidence.

Contribution to the Teaching and Learning of Biology

Our curriculum and research contributes in important ways to biology education by providing: (1) a model of a full high school unit that aligns to the 3D-based and phenomenon-based principles of NGSS; (2) research on how thoughtfully integrating evolution and heredity DCIs can support student understanding; and (3) an example of working with teachers to field test a new curriculum unit using an RCT design.

General Interest to NABT Members

The unit provides an example of an NGSS-aligned high school biology curriculum that incorporates all three dimensions of NGSS, integrates evolution and heredity DCIs, engages students in working with data from research studies about phenomena, and illustrates how students' skills in argumentation can be scaffolded throughout a unit.

For researchers, the presentation will discuss the iterative stages of design, testing, and revision in a rigorous, multi-year, multi-phase evaluation of a new curriculum unit, with a focus on RCT designs. For curriculum developers, the presentation will address the theoretical- and research-based decisions behind the development of a curriculum unit that integrates topics to allow students to make more sophisticated connections in science.

References

Achieve, Inc. (2016). EQuIP Rubric for Lessons & Units: Science, V. 2.0. Retrieved https://www.nextgenscience.org/

Asterhan, C., & Schwarz, B. (2007). The effects of monological and dialogical argumentation on concept learning in evolutionary theory. *Journal of Educational Psychology*, *99*(3), 626-639.

Berland, L., & McNeill, K. L. (2010). A learning progression for scientific argumentation: Understanding student work and designing supportive instructional contexts. *Science Education*, *94* (5), 765-793.

Catley, K. M., Lehrer, R., & Reiser, B. (2004). *Tracing a prospective learning progression for developing understanding of evolution*. National Academies Committee on Test Design for K-12 Science Achievement.

Catley, K. M., Novick, L. R., & Shade, C. K. (2010). Interpreting evolutionary diagrams: When topology and process conflict. *Journal of Research in Science Teaching*, 1-22.

Dougherty, M. J. (2009). Closing the gap: Inverting the genetics curriculum to ensure an informed public. *The American Journal of Human Genetics*, *85*, 6-12.

Driver, R. (1995). Constructivist approaches to science teaching. In L. P. Steffe & J. Gale (Eds.), *Constructivism in education* (pp. 385-500). Hillsdale, NJ: Lawrence Erlbaum Associates.

Flanagan, J. C., & Roseman, J. (2011). Assessing middle and high school students' understanding of evolution with standards-based items. Paper presented at the *National Association for Research in Science Teaching,* Orlando, FL.

Gregory, T. R. (2009). Understanding natural selection: essential concepts and common misconceptions. *Evolution: Education and Outreach, 2*, 156-175.

Linacre, J. M. (2016). Winsteps® Rasch measurement computer program. Beaverton, OR.

Mead R., Hejmadi M., & Hurst L. D. (2017). Teaching genetics prior to teaching evolution improves evolution understanding but not acceptance. *PLoS Biology*, *15* (5), 1-30.

Meir, E., Perry, J., Herron, J. C., & Kingsolver, J. (2007). College student's misconceptions about evolutionary trees. *American Biology Teacher*, *69* (7), 71-76.

Morabito, N. P., Catley, K. M., & Novick, L. R. (2010). Reasoning about evolutionary history: Post-secondary students' knowledge of most recent common ancestry and homoplasy. *Journal of Biological Education*, *44* (4), 166-174.

National Research Council. (2012). A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas. Washington, DC: The National Academies Press.

Nehm, R. H., & Schonfeld, I. S. (2008). Measuring knowledge of natural selection: A comparison of the CINS, an open-response instrument, and an oral interview. *Journal of Research in Science Teaching, 45* (10), 1131-1160.

NGSS Lead States. 2013. *Next Generation Science Standards: For States, By State*s. Washington, DC: The National Academies Press.

Postlethwait, J. H., & Hopson, J. L. (2009). *Modern Biology (Teacher Wraparound ed.)*. Austin, TX: Holt, Rinehart, and Winston.

Strike, K. A., & Posner, G. J. (1992). A revisionist theory of conceptual change. In R. A. Duschl & R. J. Hamilton (Eds.), *Philosophy of science, cognitive psychology and educational theory and practice*. Albany, NY: State University of New York Press.

Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching*, *39* (1), 35-62.