



Findings From a Two-Year Effectiveness Trial of the Science Notebook in a Universal Design for Learning Environment

Jennifer Yu^{1*}, Xin Wei¹, Tracey E. Hall², Agatha Oehlkers², Kate Ferguson¹, Kristin H. Robinson² and Jose Blackorby³

¹Center for Learning and Development, SRI Education, Menlo Park, CA, United States, ²Research and Development, CAST, Wakefield, MA, United States, ³Learner Variability and Impact, WestEd, San Francisco, CA, United States

OPEN ACCESS

Edited by:

Antonio P. Gutierrez de Blume,
Georgia Southern University,
United States

Reviewed by:

Lidia Scifo,
Libera Università Maria SS. Assunta,
Italy

Muhammet Usak,
Kazan Federal University, Russia

*Correspondence:

Jennifer Yu
Jennifer.yu@sri.com

Specialty section:

This article was submitted to
Educational Psychology,
a section of the journal
Frontiers in Education

Received: 02 June 2021

Accepted: 12 October 2021

Published: 04 November 2021

Citation:

Yu J, Wei X, Hall TE, Oehlkers A,
Ferguson K, Robinson KH and
Blackorby J (2021) Findings From a
Two-Year Effectiveness Trial of the
Science Notebook in a Universal
Design for Learning Environment.
Front. Educ. 6:719672.
doi: 10.3389/feduc.2021.719672

This cluster randomized control trial examines the effects of the Science Notebook in a Universal Design for Learning Environment (SNUdle) on elementary school student science academic achievement and motivation outcomes. Multilevel analyses examined the impact of SNUdle for all students and important student subgroups. Overall, students who received the SNUdle intervention had similar motivation and academic achievement in science to their peers who did not receive the SNUdle intervention. However, relative to students with disabilities in the comparison group, students with disabilities who used SNUdle scored significantly higher on motivation in science and science academic achievement, with effect sizes (ES) ranging from 0.82 to 1.01. Furthermore, SNUdle appeared to have a small but statistically significant positive impact on science academic performance among students whose home language is other than English or Spanish with an ES of 0.35. Fidelity of implementation analysis shows sufficient teacher training but fidelity of teacher and student usage of SNUdle needs to be improved. The qualitative analysis of teacher interviews suggests that teachers perceived benefits of SNUdle in support language acquisition and science writing skills. Both quantitative and qualitative findings suggest that SNUdle holds promise for improving academic performance in science and confidence and motivation among some of the most vulnerable student populations.

Keywords: universal design for learning, science notebook, elementary education, academic outcomes, motivation for science, students with disabilities, english language learners, mixed methods study

INTRODUCTION

Research on effective science learning shows that conducting experiments and recording data take up most of the allocated time in today's elementary school science classrooms (Fairbanks, 2013). After an experiment, the teacher may share with the class a quick explanation of the connection with a bigger science concept, or may move on to the next subject. Yet this research also points to the critical importance of building sense-making skills and connections with real-life experiences to improving science comprehension and motivation (National Research Council(NRC), 2011). Likewise, national science standards make building sense-making skills and connections an imperative (Next Generation Science Standards (NGSS), 2018).

Participation in these scientific practices builds the habits of mind that drive deeper understanding and motivation toward science learning.

Over the last 10 years in the United States there has been a renewed effort to establish national science standards for teaching and learning (National Research Council, 2012). Amid calls for students to develop an understanding of science beyond rote memorization of facts and procedures, practice-based inquiry approaches have gained popularity in both the classroom and research realms. The Next Generation Science Standards (NGSS) call for a three-dimensional approach to science instruction, around the pillars of 1) science practices, 2) disciplinary core ideas, and 3) crosscutting concepts (Next Generation Science Standards (NGSS), 2018). Beyond the NGSS standards, a long line of experimental evidence in the learning sciences has shown the importance of explanation in developing well-integrated and transferable knowledge in science (ex., Chi et al., 1989; Chi and Wylie, 2014; McNamara, 2017). Yet few teachers are supported to engage their students in authentic sense-making and inquiry (Wee et al., 2007) and thus fall short of realizing the vision of students thinking and acting like scientists.

There are a variety of possible reasons why students spend little classroom time engaging in sense-making. Low teacher confidence and knowledge in science content and pedagogy are significant obstacles (Crawford and Capps, 2016). In addition, science notebooks, a key student tool in scientific inquiry, exhibit construct-irrelevant barriers that make the inquiry and sense-making process less accessible for many students. These barriers hinder students' ability to clearly express understanding of science events and concepts. Writing skills such as spelling, fluency, recording or transcribing of data, and composing text may interfere with expression (Graham and Hebert, 2010). Additionally, there is little experimental research addressing these barriers.

SNUdle Overview

The Science Notebook in a Universal Design for Learning Environment (SNUdle) is a digital science notebook created to help students, particularly those with identified learning disabilities, as well as those who are at risk, struggling, and unmotivated, better realize the benefits of science notebooks used in the scientific inquiry process (Figure 1 and Figure 2). Universal Design for Learning (UDL) was chosen as the SNUdle design framework to minimize construct-irrelevant barriers to learning and provide just-in-time supports for active science learning and effective science notebook use (Rose and Meyer, 2002; Meyer et al., 2014). The research literature indicates that science notebooks can be used to support active science learning and the development of scientific literacy (Hargrove and Nesbit, 2003; Klentschy, 2005). However, teachers typically use science notebooks primarily in a mechanical way—to record data, procedures, or definitions—and rarely to support the development of deep understanding through the active science learning process (Baxter, et al., 2001; Ruiz-Primo et al., 2004). Given these challenges, SNUdle was designed to help teachers use

evidence-based tools and strategies to provide all students with access to the general science curriculum and meet the high academic standards set forth in the Next Generation Science Standards (Thurlow and Wiley, 2004).

Like traditional science notebooks, SNUdle provides students a structured and supported space to collect, organize, and display observations and data in science; space to reflect and make sense of inquiry experiences; and multiple opportunities to demonstrate understanding at every stage of the investigation through text answers and data tables. However, with UDL as the design framework (CAST, 2018) and digital technology as the platform, SNUdle differs from traditional science notebooks in several key ways.

First, SNUdle was developed according to accessibility guidelines from the World Wide Web Consortium (W3C Web Accessibility Initiative (WAI), 2018), Section 508 of the Rehabilitation Act (29 U.S.C. 794d), and the National Center on Instructional Media (2006). Text-to-speech technology is built into the notebook interface with real-time highlighting to support simultaneous access to auditory and visual processing, as well as word-by-word English-to-preferred language translation (in this study Spanish, Vietnamese and Arabic language translations were available), keyboard-accessible actions, and a multimedia glossary to provide just-in-time support for vocabulary use and development. These features remove barriers faced by many students with learning disabilities whose literacy skills would interfere with the efficacy of materials that depend on proficiency in reading and writing. They also help support students for whom proficiency in English is a barrier, and others who would more effectively learn through use of built-in accessibility features.

Second, SNUdle leverages contextual support to develop and reinforce effective science learning behaviors. Pedagogy is built into the interface design itself, guiding students and teachers in the process of active science learning and the effective use of science notebooks. For instance, students are prompted to think about making direct reference to their data and observations in support of their conclusions and to use relevant vocabulary from their inquiry experiences throughout their notebook entries.

Third, in addition to the student interface, SNUdle contains a teacher interface which includes features that facilitate active science learning. For instance, teachers are prompted and supported to provide feedback on the students' entries that may include corrective information, alternative strategies, information to clarify ideas, or encouragement to engage in the scientific process.

Research Questions

Using a mixed-methods research approach, this study sought to quantitatively and qualitatively address the impact of SNUdle on student outcomes, and to explore how SNUdle was used by students and teachers in the classroom. Specific research questions (RQs) include:

RQ1 (Overall Impact): Did students who used SNUdle in fourth grade science classrooms achieve greater gains in science

FIGURE 1 | Screenshot of the “Explain” page in SNUdle in which students summarize their analysis of the data and provide their evidence and reasoning as to how the experiment they conducted answers the focus question.

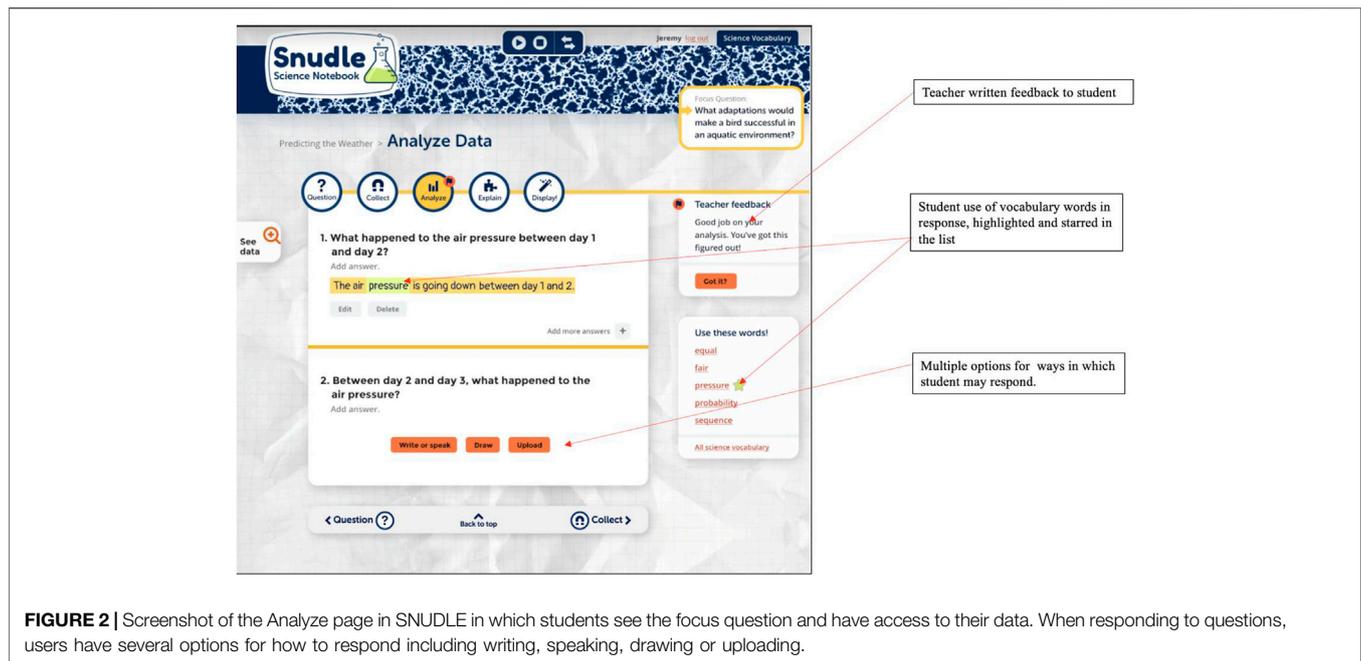


FIGURE 2 | Screenshot of the Analyze page in SNUDLE in which students see the focus question and have access to their data. When responding to questions, users have several options for how to respond including writing, speaking, drawing or uploading.

learning and STEM self-efficacy when compared to students who used traditional paper-based science notebooks?

RQ2 (Differential Impact): Did the impact of SNUDLE vary among students with disabilities and those whose families speak a language other than English or Spanish?

RQ3 (Implementation): Do the usage patterns indicate that SNUDLE was implemented with fidelity by students and teachers? What were the SNUDLE features most commonly used by the different subgroups of students?

RQ4 (Perception): What were teachers' perceptions of the usefulness of SNUDLE in science learning, engagement, and self-efficacy?

MATERIALS AND METHODS

Quantitative Design

A cluster randomized controlled trial was conducted within a large urban school district from August 2017 to January 2019. There were seven participating elementary schools with a total of 36 participating fourth-grade teachers across two cohorts; the first cohort (August 2017 to January 2018) included 29 teachers and the second cohort (August 2018 to January 2019) included seven teachers. Parental consent forms were distributed on the first day of school for each cohort. While all students who enrolled in the study's full inclusion general education science classes were eligible to participate in the study, we received parental consent and student assent to participate for 683 students (372 intervention, 311 comparison) as of August 2017 for cohort 1 and 219 students (97 intervention, 122 comparison) as of August 2018 for cohort 2, for a total of 902 students (469 intervention and 433 comparison) across both cohorts.

Teacher randomization was conducted after parent consent and student assent were completed. Stratifying by cohort and school, blocking schedule, years of teaching experience, and confidence level in teaching science, the 36 teachers were randomized into SNUDLE or business-as-usual (BAU) condition; 20 teachers were randomized to the SNUDLE condition (16 in cohort 1 and 4 in cohort 2) and 16 teachers were randomized to BAU comparison conditions (13 in cohort 1 and 3 in cohort 2). All teachers taught using the district's StemScopes curriculum, which included a traditional paper science notebook or worksheets. Treatment teachers used the SNUDLE science notebook in lieu of the paper notebook, while condition treatment teachers continued use of traditional paper science notebooks/worksheets.

Qualitative Design

Structured interviews were conducted with both treatment and comparison teachers in both years of the study. A random sample of teachers were selected for interviews. In year one, 10 interviews were conducted with treatment teachers and nine with comparison teachers. At the conclusion of the intervention in year two, eight treatment teachers and five comparison condition teachers were interviewed.

SNUDLE Intervention Teacher Training

Training was conducted on two occasions by CAST staff for both cohorts. In the summer, both intervention and comparison teachers participating in the SNUDLE study received a full-day training during which they were introduced to the purpose and goals of the study and received professional development on the principles of UDL. After randomization, teachers in the intervention condition then received an additional 4 hours of training,

during which time CAST staff provided them with SNUdle materials and introduced them to the program and SNUdle's educational philosophy and approaches to pedagogy. During the training, intervention teachers were provided multiple opportunities to practice using both the teacher- and student-facing SNUdle views and role-played how they would use SNUdle in their classrooms. For instance, with student-facing SNUdle, teachers practiced using the multiple modalities of responding, such as speech-to-text, drawing features, and typing or finger-writing responses on the tablet. On the teacher-facing SNUdle, teachers used the dashboard to view student progress and practiced using the comments features to provide just-in-time feedback on student progress. Upon implementation in their classrooms, intervention teachers received ongoing coaching and support from CAST *via* weekly newsletters that provided best practice tips and tricks. They also received individualized support when requested and/or when classroom observations suggested the need for additional implementation support and technical assistance.

Intervention teachers in both cohorts implemented the intervention from September to January of the school year in which they participated (2017–18 or 2018–19), during which the SNUdle tablets were integrated into 18 investigations across nine curriculum units. Teachers in both the intervention and comparison group were required to use the district-mandated science curriculum and adhered to the district's pacing guide for administering each lesson in a prescribed timeline. The only difference between the SNUdle intervention group and the BAU comparison group was the use of the SNUdle digital science notebook rather than a paper-based notebook when completing the investigations.

Measures

Student Demographics

From the student's school records, we obtained sociodemographic data on gender, race/ethnicity, free or reduced-price lunch status, dual language learner status, and language spoken at home. The most common languages spoken at home among participating students were English, Spanish, Vietnamese, and Arabic. Disability status was identified from administrative data indicating the student has an individualized education program (IEP).

Curriculum-Based Unit Tests (i.e., Quiz Scores)

Assessment items from *STEMscopes*, the school district's curriculum, were used as academic achievement measures closely aligned with the curriculum content. The curriculum developers categorized the items by the four levels of Bloom's Taxonomy: Understand, Apply, Analyze, and Evaluate. Because SNUdle seeks to provide opportunities to improve higher level science thinking, the items we selected predominantly focused on Analyze and Evaluate questions. One of the nine unit tests or quizzes was dropped from analysis because a natural disaster caused school closure at the beginning of the study that

interrupted teaching and quiz administration. The sum of correct responses across the remaining eight end-of-unit quizzes served as a proximal outcome measure. The standardized Cronbach coefficient alpha for the *STEMscopes* unit tests was 0.88.

District Common Assessment in Science

The District Common Assessment (DCA) in Science was used as a pretest measure of academic performance in science and administered to both intervention and comparison students. The school district developed the DCA as a measure of intermediate-term goals and objectives and it is administered at the end of the first and second semesters each school year. The DCA was designed to assess concepts from the Texas Essential Knowledge and Skills (TEKS), with many of its items based on the validated State of Texas Assessment of Academic Readiness (STAAR®) program (Human Resources Research Organization, 2016). DCA items that are not directly aligned with TEKS were derived from tests published by several commercial publishing companies.

Measures of Academic Progress

For a broader measure of science knowledge, we administered the Northwest Evaluation Association's MAP test of science at the end of both Year 1 and Year 2 data collections. The MAP science test is a formative measure that covers domains of Earth, life, and physical sciences. It is a computerized adaptive assessment consisting of 50 multiple-choice items with four or five options. In the Northwest Evaluation Association's item development, all items match the assessable sections of a set of academic content standards both in breadth of content and depth of knowledge. MAP tests have been validated to link to content standards in all 50 states and have excellent technical characteristics (Northwest Evaluation Association, 2011). Third-grade state standardized achievement scores in English Language Arts measured by the STAAR program were collected as a baseline measure of academic performance for the fourth grade participating students in this study.

Motivation for Science (MFS)

A key outcome for SNUdle is its ability to increase not just students' knowledge of science practices, but also their motivation to learn science. The MFS is an 18-item survey intended to measure the latter. The MFS consists of subscales for the following four constructs: self-efficacy, interest, desire for challenge, and comfort using computers. Reliability of the MFS is 0.85; for the experimental sample, it was 0.89 (Rappolt-Schlichtmann et al., 2013). The MFS was administered to all participating and consented students at pretest after teacher randomization, and again at posttest after completion of the intervention.

Implementation Measure

Implementation fidelity was measured for the two components of the SNUdle intervention: 1) the training and ongoing coaching support provided to the teachers in the SNUdle intervention group and 2) implementation of SNUdle by the teacher and

students for each of the 15 investigations included in the analysis. For the teacher training, implementation fidelity involved full attendance at two training sessions (total of 8 hours of training) and participation in observation and ongoing coaching from the SNUDLE development team and district-level science specialists trained in SNUDLE. To meet fidelity thresholds, teachers were expected to attend both trainings in their entirety and participate in at least one observation and coaching session with their specialist or SNUDLE researcher.

For documentation of SNUDLE implementation in the classroom, we relied on the SNUDLE software usage data collected while the teacher and students interacted with the SNUDLE online notebook during each of their science lesson's experimental investigations. SNUDLE records user data and provides a dataset describing instances of actions such as logins, which pages and investigations students visit, when and on which pages students create content, which features of SNUDLE students use (e.g., text to speech, language translation), and when teacher users provide written feedback to student users in SNUDLE. From this dataset, we measured implementation fidelity based on the quantity or dosage of SNUDLE use, the quality or depth of use, and the frequency with which accessibility features afforded by SNUDLE were accessed by the students.

Dosage. The study data collection period covered the first semester of the district's science curriculum for each of the two cohorts. Dosage was measured in two ways: 1) Number of investigations accessed in the classroom. Pairs or groups of students used SNUDLE together after logging into a single student's account. Therefore, SNUDLE access during investigations was identified at the classroom rather than individual student level. To ensure that student access was intentional rather than accidental log-ins or used for purposes other than the investigation, at least 15% of students in the classroom were required to log onto SNUDLE at the same time to be identified as SNUDLE used for an investigation; 2) Teacher usage. To measure direct teacher use, we calculated the number of days the teacher logged into SNUDLE.

Quality of usage. There are three different steps students worked through on an investigation in SNUDLE: 1) collect data, 2) analyze data, and 3) explain findings. Each step in the investigation process is a separate webpage of the digital notebook. Accessing the "Analyze data" and "Explain findings" pages suggest a higher level of SNUDLE usage as these pages were developed with the purpose of engaging students' deeper scientific thinking. To measure usage of these pages, we calculated the percentage of days in which students created or edited contents on Analyze or Explain pages out of the total number of days students created or edited content on any of three steps/pages.

Accessibility features. In addition to looking at implementation of SNUDLE via dosage and quality of usage, we also observed the backend usage data to understand what types of SNUDLE features were used by the students with and without disabilities. Specifically, we counted how often students used each of a variety of features designed to make the SNUDLE notebook more accessible. These features include things like draw

tool, text-to-speech functionality, translation into three different languages, glossary functions, and sentence starters.

Qualitative Teacher Interviews

A twenty-one question interview was designed to address research question four (RQ4) to better understand teacher perceptions of student performance and use of science notebooks. The research team designed a scripted protocol and conducted the teacher interviews conducted by phone at the conclusion of the intervention period each year. Treatment and comparison teachers were asked the same questions, with references to SNUDLE or to the traditional science notebook adapted to match the condition of the interviewee. Some questions were closed-ended, with response options formatted using a Likert-like scale, while other questions solicited open-ended responses. Each interview required approximately 30–45 min to complete, and teachers and interviewers engaged in dialogue during the process. Interviews were recorded and transcribed to allow for comprehensive analysis.

Data Analysis

Quantitative Data Analysis

This study conducted descriptive analysis of baseline and outcome variables for the whole sample, students with disabilities, and students whose home language was a language other than English or Spanish.

Primary estimates of the intervention effect were derived from intent-to-treat (ITT) analyses in which all students remained in the group they were originally assigned to for analyses, regardless of attrition or movement across groups. Regardless of the level of implementation, these analyses compared all students in treatment teachers' classrooms to their peers in comparison teachers' classrooms. Two-level Hierarchical Linear Modeling (HLM; Raudenbush and Bryk, 2002) was performed to estimate the impact of SNUDLE after taking into account students were nested in teachers. Level 1 is the student level and level 2 is the teacher level. Dependent variables were the MAP, DCA, MFS, and total unit quiz score. Independent variables included a constant, a pretest score on the same outcome measure or STAAR (the state standardized reading test) score when pretest on the same outcome measure is not available, demographic characteristics, and treatment indicator. Treatment indicator variable is at level 2. Covariates in this study were derived from the extensive literature on predictors and correlates of students' academic achievement (Hair et al., 2006; Zhai et al., 2011). Specifically, we included the following student-level covariates in the two-level ITT HLM because previous studies have shown that these background characteristics are related to achievement: gender, race, language spoken at home, IEP status, low-income status, and dual language learner status. $Y_{is}^{posttest} = \beta_0 + \beta_1 Pretest + \beta_2 Treatment + \beta_3 COV_{is} + \gamma_{is} + \mu_{os}$, where i is the student; s is teachers; $Y_{is}^{posttest}$ is posttest scores on MAP, DAC, or quiz; $Pretest$ is the baseline test score; $Treatment = 1$ for intervention teachers and 0 for comparison teachers; COV_{is} is student-level covariates. γ_{is} and μ_{os} are individual and teacher random effects. Hedges' g effect sizes for the treatment impact are

TABLE 1 | Baseline demographic characteristics of students.

Student characteristics	Treatment (n = 449)	Control (n = 401)
Female	52(%)	53(%)
Race/ethnicity		
White	4(%)	4(%)
African American	47(%)	29(%)
Asian/Native Hawaiian/Pacific	12(%)	12(%)
Islander		
American Indian/Alaskan	1(%)	3(%)
Hispanic	37(%)	52(%)
Free or reduced-price lunch status	81(%)	84(%)
Dual language learner status	26(%)	23(%)
Home language		
English	54(%)	37(%)
Spanish	29(%)	47(%)
Other	17(%)	16(%)
Individualized Education Program	7(%)	6(%)
504 plan	4(%)	3(%)

Treatment sample size is 469 and comparison sample size is 433, 449 out of the 469 treatment students have no missing data on any demographic variables. 401 out of the 433 comparison students have no missing data on any demographic variables.

calculated as dividing the HLM coefficient for the intervention's effect by the pooled treatment and comparison group standard deviation (What Works Clearinghouse, 2017). This study did not impute missing data. Missing data were listwise deleted from the HLM. Treatment impact for subgroup of students was estimated using HLM restricting data to either students with disabilities subgroups or students whose families spoke a language other than English or Spanish.

Qualitative Analysis

The research team adopted a consensus scoring structure in which we used a method of collaborative qualitative analysis involving six phases: 1) preliminary organization and planning, 2) open discussion coding, 3) development of coding labels and structures, 4) initial testing of the coding structure, 5) agreement and finalizing the coding, and 6) reviewing and identifying themes (Richards and Hemphill, 2018).

The coding team used the qualitative analysis software program MAXQDA to code and analyze teacher interview results (VERBI Software, 2019). Initially all interviews were coded for condition. Researchers then identified "effects" by themes which included: 1) motivation, 2) engagement, 3) independence, 4) collaboration, 5) confidence, 6) building understanding, and 7) teacher self efficacy/teaching practices. The third level of coding included treatment teachers' perceptions of student use of supports and scaffolds in SNUDLE. The next coding level identified student subgroups by demographic characteristics: English Language Learners (ELLs), students who are "struggling" (defined by teachers as students who generally had difficulty in reading, math and writing, (generally in the lower quartile of the class), and students with disabilities. Additionally, the research team assigned coded segments a positive, neutral, or negative rating to evaluate where impact

of the SNUDLE digital notebook was evident for treatment condition teachers.

RESULTS

Attrition Analysis

Although randomizing teachers and their students to conditions should result in statistically equivalent groups, higher overall level of attrition and differential attrition between treatment and comparison groups may jeopardize the initial balance and impact estimate may be biased (What Works Clearinghouse, 2017). Our data analysis began with an attrition analysis. Across seven outcomes at posttest, treatment group attrition rate ranged from 1 to 16%, comparison group attrition rate ranged from 7 to 15%, and the differential attrition rate ranged from 1 to 6%. According to the WWC standards (2017), the overall and differential attrition rate is low for this study.

Descriptive Analysis

After the attrition analysis, a descriptive analysis was conducted for SNUDLE students and comparison students. **Table 1** presents the student background characteristics by condition. **Table 2** describes baseline and posttest scores by condition for the whole sample, students with disabilities subsample, or students whose home language was not English or Spanish subsample. Statistical significance of the difference between the SNUDLE and comparison groups at baseline was determined from HLM analysis. SNUDLE participants were not significantly different from comparison students on baseline assessment scores for the whole sample and two subsamples.

Intent-To-Treat Analysis Results (RQ1)

Primary estimates of the SNUDLE impacts were derived from the ITT analyses. **Table 3** demonstrates that no significant differences were detected between SNUDLE and comparison fourth-grade students among the overall sample on any academic or motivation outcomes at the end of the 5-month intervention.

Subgroup Analysis Results (RQ2)

Our subgroup analysis showed that the effect of SNUDLE was significant and large among students with disabilities. For example, among the students with disabilities, the SNUDLE group scored significantly higher on motivation in science (Efficacy: $ES = 0.88, p < 0.05$; Interest: $ES = 0.82, p < 0.05$; Desire for challenge: $ES = 1.01, p < 0.05$) and science academic achievement (Total quiz score: $ES = 0.82, p < 0.01$). Furthermore, SNUDLE appeared to have a small but statistically significant positive impact on science academic performance among students whose home language was not English or Spanish ($ES = 0.35, p < 0.01$).

Fidelity of Implementation Descriptive Results (RQ3)

Our implementation analysis includes descriptive analysis of three aspects of implementation: teacher training and ongoing support, dosage and quality of usage by teacher/classroom, and student use of SNUDLE features. First, all SNUDLE teachers

TABLE 2 | Pretest and posttest scores by treatment assignment for the whole sample, students with disabilities subsample, students whose home language was not English or Spanish.

Student assessments	Whole sample						Students with Disabilities						Home language other than English or Spanish					
	Treatment			Comparison			Treatment			Comparison			Treatment			Comparison		
	n	Mean	SD	n	Mean	SD	n	Mean	SD	n	Mean	SD	n	Mean	SD	n	Mean	SD
STAAR reading used as pretest for MAP posttest	422	1,423.84	150.48	375	1,455.69	151.58	29	1,303.86	232.27	22	1,334.86	120.60	75	1,460.63	145.44	59	1,473.81	154.05
MAP RIT posttest	422	197.83	11.37	375	198.65	10.72	29	190.90	13.98	22	185.86	10.08	75	203.24	11.64	59	200.17	11.08
DCA May 2018	390	65.4	17.12	350	66.66	17.44	26	53.88	20.76	22	50.73	14.69	67	69.57	17.10	58	68.97	17.58
DCA Dec 2019	390	68.7	17.13	350	70.97	17.89	26	56.19	18.98	22	48.41	21.34	67	73.42	18.93	58	69.10	18.89
MFS-Efficacy pretest	373	3.35	0.61	337	3.33	0.54	23	3.02	0.74	18	3.10	0.63	63	3.25	0.71	51	3.33	0.57
MFS-Efficacy posttest	373	3.42	0.57	337	3.35	0.57	23	3.33	0.73	18	3.00	0.93	63	3.34	0.56	51	3.37	0.49
MFS-Interest pretest	372	3.43	0.58	334	3.49	0.48	22	3.16	0.72	18	3.36	0.56	63	3.39	0.64	50	3.48	0.44
MFS-Interest posttest	372	3.44	0.63	334	3.49	0.5	22	3.50	0.59	18	3.19	0.59	63	3.39	0.65	50	3.44	0.52
MFS-Desire for challenge pretest	371	3.16	0.62	335	3.19	0.57	22	2.76	0.66	18	2.93	0.62	63	3.12	0.64	50	3.25	0.54
MFS-Desire for challenge posttest	371	3.25	0.64	335	3.31	0.55	22	3.08	0.85	18	2.99	0.63	63	3.27	0.68	50	3.30	0.54
MFS-Comfort using computer pretest	372	3.16	0.59	337	3.14	0.56	23	3.21	0.52	18	2.96	0.79	63	3.22	0.50	50	3.10	0.51
MFS-Comfort using computer posttest	372	3.32	0.59	337	3.25	0.56	23	3.32	0.56	18	2.93	0.71	63	3.44	0.50	50	3.28	0.55
STAAR reading used as pretest for total quiz score	430	1,425.44	150.22	385	1,455.15	151.16	28	1,305.75	236.31	22	1,341.73	115.75	75	1,460.63	145.44	63	1,462.73	156.36
Total quiz scores	430	29.58	10.69	385	30.98	11.58	28	25.50	8.50	22	21.95	11.20	75	34.33	11.51	63	32.38	13.85

The sample size is the number of students who had both pretest and posttest for a particular outcome. STAAR, state of Texas assessments of academic readiness; MFS, motivation for science.

across two cohorts attended the two training sessions in their entirety (total of 8 h of training). Additionally, they were observed by and received post-observation coaching from district-level science specialists trained in SNUDLE or the SNUDLE research team at least one time during the research study. Therefore, there was 100% implementation fidelity as pertains to the teacher training and professional development component of the SNUDLE intervention.

Second, the SNUDLE usage data confirmed that all SNUDLE teachers implemented the intervention. As shown in **Table 4**, the dosage descriptive data shows that the mean number of investigations accessed on SNUDLE was 6.53 investigations, which is a little less than half of the total number of investigations SNUDLE offered. The number of days teachers accessed SNUDLE was 9.63, which suggests that teachers did not access SNUDLE for each investigation. We expected higher quality of the SNUDLE usage as presented by using SNUDLE to analyze data and explain findings instead of to collect data, as this indicates that students were more actively engaged with SNUDLE as they created or edited the content of analyze or explain pages. Students in SNUDLE classrooms on average spent 46.65% of the time on analyze or explain pages. Overall, the backend usage data suggest an insufficient level of implementation within SNUDLE classrooms.

Third, when we examine SNUDLE features that were used by students, we found that the top five most frequently used SNUDLE features were the draw tool, sentence starters, data table, glossary, and text-to-speech (**Table 5**). There was no difference in SNUDLE feature usage by disability status except that students with disabilities used the draw tool functionality less often than their peers without disabilities ($t = 2.22, p < 0.05$). Although we expected students with disabilities might have used SNUDLE features as often as their peers without disabilities, the t-test results indicated that students with disabilities were willing to use and benefit from SNUDLE. This study did not find differences in feature usage by home language except that students whose home language was a language other than English or Spanish used the “set language” functionality more often than their peers whose home language was English ($t = 2.70, p < 0.01$). The “set language” function assigns the language used by the translation feature when translating from English to a target language.

Qualitative Findings (RQ4)

During qualitative data analysis, researchers identified several frequently occurring codes in the interview dataset. These codes were broadly categorized as 1) subgroups of student users, 2) SNUDLE features, and 3) impact on student learning and performance. While teachers in the comparison condition (traditional science notebook) observed students in their classrooms responding positively to use of a science notebook to support the inquiry process, more teachers in the SNUDLE condition described their students as engaged, collaborative, and motivated by science learning. Several teachers using SNUDLE in their classroom attributed these

TABLE 3 | Estimated treatment impact among the whole sample and among students with disabilities.

Outcomes	Whole sample			Students with Disabilities			Home language other than English or Spanish		
	β	SE	Effect size	β	SE	Effect size	β	SE	Effect size
MAP	0.71	1.10	0.06	3.41	2.88	0.27	4.03**	1.54	0.35
DCA	-1.05	1.78	-0.06	10.06 [†]	5.07	0.50	3.97	2.87	0.21
MFS-Efficacy	0.05	0.04	0.10	0.73*	0.28	0.88	-0.01	0.09	-0.03
MFS-Interest	-0.02	0.04	-0.04	0.49*	0.22	0.82	0.02	0.09	0.03
MFS-Desire for challenge	-0.03	0.05	-0.05	0.77*	0.28	1.01	0.04	0.10	0.06
MFS-Comfort using computer	0.05	0.04	0.09	0.19	0.23	0.31	0.07	0.08	0.14
Total quiz	1.29	2.02	0.12	8.06**	2.70	0.82	2.48	2.41	0.20

No need to run Benjamin-Hochberg multiple comparison adjustment because treatment impact was not significant. The HLM controls for pretest, gender, race, free or reduced lunch status, dual language learner status, home language, and IEP status.

[†]p < 0.1, *p < 0.05, **p < 0.01.

TABLE 4 | Description of fidelity of implementation measures aggregated to the teacher/classroom level for cohort 1 and 2.

Variables	Mean	SD	n	Min	Max
Number of investigations accessed	6.53	3.22	19	0	11
Days teacher used SNUDLE	9.63	8.25	19	0	25
Higher level of usage (%)	45.32	20.01	19	0	77.6

Number of investigations accessed by more than 15% of students. Teacher usage is measured by the number of days teacher used SNUDLE with any usage data. High level of usage is measured by the percentage of student-days in which students created or edited their content on either the analyze or explain pages (higher-order thinking) over the total days they created or edited their content on collect, analyze, or explain pages.

student behaviors to specific features unique to the digital notebook environment.

Teachers in both conditions across the 2 years consistently noted positive impact from both SNUDLE and traditional science notebooks on students who struggle, including those with disabilities, in contrast to students not identified with learning challenges. Although both SNUDLE and comparison condition teachers noted the positive impact of science notebooks on this subgroup, a greater number of treatment condition teachers (16 out of 17) noted this impact, in comparison to control teachers (8 out of 14). Treatment teachers also observed that students were using sense-making skills in inquiry science: “SNUDLE guided (students) more (versus) those who struggled with the lab and concept learning without SNUDLE. They had high engagement with SNUDLE. . .now they can connect the lab to the questions and the intent of the lab they did.” A smaller but meaningful impact was described for students for whom English is not their primary home language. Many treatment teachers (8 out of 17) noted positive outcomes for these students, whereas fewer comparison teachers (2 out of 14) noted a positive impact. As one teacher using SNUDLE described in their interview, “ESL students are writing more, elaborating more. This is a great celebration.”

To analyze student learning performance, researchers coded for effects by recurring themes as listed in the

methods section. Of the seven themed categories, three emerged most frequently in the teacher interviews: collaboration, engagement, and motivation. These student behaviors were noted during hands-on science investigations and use of science notebooks. Collaboration was characterized as student interactions with peers, engagement as student participation and on-task focus in science, and motivation as maintaining engagement and interest. While both treatment and comparison teachers observed positive impact on collaboration, engagement, and motivation, this impact was reported at a higher rate by treatment condition teachers (see Table 6).

Finally, during interviews, many treatment teachers noted specific features designed to support pedagogy, usability, and accessibility within SNUDLE for their students (Table 7). While teachers did not speak to all features available in SNUDLE, teachers did identify eight specific features when speaking of impact of the tool on learners. Those features were: uploading images, drawing tool, use of tables, sentence starters, translation, speech-to-text, text-to-speech, and multimedia glossary. Teachers noted benefits to their students most frequently from the multimedia glossary, drawing tool, sentence starters, speech-to text, and text-to-speech (Table 8).

Overall, teachers who used SNUDLE in their classrooms found the tool to be helpful in supporting their students and themselves. Factors such as structure and organization of the digital notebook were noted as helpful for many students because they enabled students to focus on the new material of the experiment, as opposed to deciding what to write or do next. Teachers also appreciated that having science notebook materials online helped students access those records and responses at any time, from anywhere. Science notebook materials were not lost in student desks or left at home. Additionally, teachers found benefits in SNUDLE providing multiple means of action and expression on all prompts and questions: “SNUDLE guides you from beginning to end. It allows variation in how students respond—drawing, typing, speech-to-text—it allowed students to respond in a variety of ways.”

TABLE 5 | Description of SNUdle features usage at the student level.

Features	Whole sample (n = 394)		Students without disabilities (n = 367)		Students with disabilities (n = 27)		Language-English (n = 216)		Language-Spanish (n = 109)		Language-other	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Draw	26.7	22.78	27.23	23.08	19.52*	16.89	25.97	26.15	29.09	17.3	25.26	18.57
Sentence starter	8.8	13.44	8.81	13.39	8.74	14.42	10.3	15.4	5.48	8.46	9.38	12.49
Table	8.79	7.74	8.88	7.89	7.59	5.24	9.52	8.44	6.48	5.97	10.1	7.21
Glossary	7.19	13.5	7.28	13.48	5.85	13.88	8.27	15.53	5.61	9.85	6.28	11.29
Text to speech	2.18	7.12	2.15	7.25	2.63	5.1	2.25	8.32	2.07	4.68	2.14	6.26
Upload	1.18	3	1.21	3.08	0.78	1.67	1.35	3.54	0.96	1.86	0.99	2.56
Speech to text	1.03	4.04	1.06	4.14	0.63	2.2	0.86	3.75	0.86	3.18	1.84	5.73
Set language	0.25	1.24	0.25	1.28	0.15	0.53	0.11	0.64	0.39	1.85	0.45*	1.46
Translation	0.16	1.15	0.17	1.19	0	0	0.02	0.22	0.51	2.13	0.01	0.12
Set speed	0.15	0.67	0.14	0.62	0.26	1.16	0.19	0.82	0.12	0.5	0.06	0.29

The outcome is the number of times accessibility features were used by each student. Two independent sample t-test were conducted to test whether there are differences in feature usage by disability status or home language status. *p < 0.05.

TABLE 6 | Teacher perceptions of positive impact of science notebook use.

Area of perceived impact	Treatment teachers (SNUdle) (n = 17)	Comparison teachers (traditional science notebook) (n = 14)
Students who struggle	94(%)	57(%)
Student engagement	76(%)	29(%)
Student collaboration	71(%)	57(%)
Student motivation	59(%)	7(%)

Teacher perceptions were measured by the number of teachers who referenced these areas of impact in their interviews. These measurements were then filtered to include only positive impact.

DISCUSSION

Findings indicate that SNUdle holds promise for improving academic performance in science and confidence and motivation among some of the most vulnerable student populations. SNUdle was created to support students, especially those with disabilities, in applying and demonstrating understanding during the scientific inquiry process. Previous research on science notebooks suggests that active science learning is particularly challenging for struggling learners who usually have low motivation for science learning (Englert et al., 1988; Graham, 1990; Swanson, 1999). Students with disabilities struggle not only with understanding science concepts, but also with all aspects of active science learning (Rappolt-Schlichtmann, et al., 2013). When compared to students with disabilities in science classes that used traditional paper-based science notebooks, students with disabilities who received the SNUdle intervention in their science classes had significant positive outcomes in their motivation to learn science and desire for challenge, as well as their ability to demonstrate understanding of science concepts as measured by content area quizzes. Our subgroup analysis for students with disabilities showed that the effect of SNUdle was significant when analyzing student efficacy ($p < 0.05$), interest ($p < 0.05$), desire for challenge ($p < 0.05$), and science academic achievement based on the total quiz scores ($p < 0.01$).

Additionally, SNUdle appeared to have a small but statistically significant positive impact on science academic performance among students whose home language was not English or Spanish ($ES = 0.35$, $p < 0.01$). These students had higher academic test scores in science when compared to their counterparts in comparison classrooms. It should be noted that the academic gains in these subgroups were based on curriculum-based quizzes and district-administered assessments and did not extend to the MAP science test, a national curriculum-based assessment. Thus, it appears that the benefit of SNUdle in encouraging scientific inquiry and science performance might be more easily captured by these two science achievement measures that were more directly aligned with the district

TABLE 7 | Teacher Interview Comments Regarding Use of SNUdle.

Interview code	Teacher comments
Teaching and learning	
Structure	They can't lose their packet. It's (SNUdle) all there on the iPad. They can log in and change something if they want to. The easiest part was the organization of SNUdle, from data to analyze to explain. That makes it easier for students, when one small part is finished, it answers, "What do I do next?" SNUdle makes connections between tasks and instruction. (I see) lots of enthusiasm they become independent learners, they become peer tutors, and they become scientists
Transfer	Students used skills learned in SNUdle in math and reading. There was generalization and transfer of skills from SNUdle to their other work
Teaching practice	SNUdle helped my teaching because it helped me figure out the proper time to ask these deeper questions SNUdle helps me as teacher try to figure out a road map for the students moving from no idea to conceptual understanding
Technology	I thought they would decrease interaction by using technology because they were all working on their own devices, but it actually increased their interaction with each other because they were all interacting with the project Now they are focused in on technology and are excited by it. They feel like they have power and control over their learning now that they have the iPad in their hand
Collaboration and engagement	Students were more collaborative in SNUdle than when working on their own worksheets or doing the packets they typically would have done if they weren't using SNUdle. They (students) do more collaboration compared to previous years, because the step-by-step process helped to understand questions and steps. They were able to answer and respond to questions. They like collaborating, and they actually became peer tutors, shared background knowledge and to relate/share science experiences with one another, and builds excitement about what to do. In short, they were more motivated
Feature use	
Multimedia glossary	With tablet(s), after a couple of times, even bilingual kids were looking at vocabulary words. (I) challenged them: whatever you write, try to use vocab words. Students liked the challenge/reward (getting stars) when they used vocabulary. Vocab was clearer to them because of the pictures and translations for bilingual students
Drawing tool	(SNUdle) had really nice drawings and they could explain what they understood. Not that detailed, but the drawing complimented the idea and I could see what they knew The parts that really excited them was the ability to draw, find pictures, or they could go online and find a picture. It gave students options on what they want to do, which was very helpful for some of them
Sentence starters	Students who struggled now want to show me, "Look, I chose this stem, and here is why." They have more ownership The sentence starters were some of the biggest benefits. ESL students know how to begin
Speech to text (STT)	The accommodations that are embedded (STT) help me see in writing what (students) are thinking. A lot of students struggle in spelling. When they speak it (in SNUdle) they can see the word that they are trying to say. They can verbally express what they think in science but can't connect to the writing
Text to speech (TTS)	For language acquisition/learning language, the TTS was beneficial. This worked out really well For students who struggle, they are able to be independent. They can hear (text) being spoken to them and don't need the teacher to read to them
Tables	It's (SNUdle) already organized for them. The data tables, the worksheet is there. The students don't have to create the data table. They can focus on what they want to put down vs. how to put it down

These interview excerpts were selected to be representative of frequently occurring themes from teacher interviews.

TABLE 8 | Teachers reporting positive impact of SNUdle Features on Students (*n* = 17).

	All students	Students with disabilities	English language learners
Upload images	3	—	—
Tables	3	1	—
Translation	1	—	3
Speech to text	5	1	2
Sentence starters	6	—	3
Multimedia glossary	13	1	3
Draw tool	10	2	3
Text to speech	2	2	2

Features represented in this table were the ones referred to by teachers during the course of interviews, independent of interview questions.

science curriculum than with a more generalizable science assessment. These findings provide promising evidence that use of a UDL science notebook, designed to support students and their teachers during active science learning, improves science achievement and motivation outcomes for students

with disabilities and students whose home language is not English or Spanish.

Furthermore, by incorporating the principles of UDL, in which a foundational tenant is recognizing and respecting variability in learning and learners (CAST, 2018), SNUdle

was designed to overcome construct-irrelevant barriers and provide contextual supports that promote active science learning for all student users (Rappolt-Schlichtmann et al., 2013). While SNUBLE was originally designed with disabilities in mind, its features can be universally leveraged to support active engagement and learning for all students, and was indeed shown to improve academic outcomes among students whose home language was neither English nor Spanish. SNUBLE levels the playing field for students with disabilities and non-Spanish-speaking ELLs.

One possible explanation for the improved academic performance among these students is that traditional paper-based science notebooks may have inhibited full understanding and expression of science learning for students below a certain level of literacy proficiency (SWD and those learning English). SNUBLE's embedded accessibility features and scaffolded supports enabled the students to overcome these barriers and focus on the learning at hand. Qualitative data portray the extent to which teachers perceived benefits of using SNUBLE with their students who are learning English. Several teachers described seeing positive effects on learning for these students, particularly in SNUBLE's ability to support language acquisition and science writing skills. In fact, SNUBLE usage data indicate that students whose home language was neither English nor Spanish used the Set Language (language preference) feature more often than their peers whose home language was English. SNUBLE provides flexible, interactive learning spaces and options for all students to demonstrate science understanding, including those whose home language is neither English nor Spanish. SNUBLE allows students to draw on diverse language strengths and resources and to move from a limited learning space to a flexible space in which they can express both science and language learning through multiple means, such as drawing, uploading images, text, and tables (Wilmes and Siry, 2020).

The usage data also revealed that all students, regardless of disability or home language status, used some accessibility features, with the most frequently used features being the drawing feature, sentence starter, glossary, and text-to-speech. Teacher interviews provided insights on the value these features provided to students. As one treatment condition teacher observed, "For students who struggle, the drawing is a bonus, and using the sentence stems is a lifesaver." These features, as well as the UDL-supported Collect, Analyze, and Explain pages, may reduce the effects of barriers to science learning, which might be particularly useful for students with disabilities.

While the study found several positive outcomes for students with disabilities and students whose home language was reported as a language other than English or Spanish, with the most common "other" languages being Vietnamese and Arabic, no measurable impact was detected on the aggregate participant student population.

LIMITATIONS AND FUTURE RESEARCH

These results of the impact of SNUBLE observed among students with disabilities and those whose home language is other than English or Spanish are even more impactful when one considers that they occurred despite the fact that fidelity of implementation did not reach ideal levels for two of the three FOI measures. The lack of implementation is a possible reason for the lack of significant findings in the overall student sample, in contrast to the positive effects reported in the previous study conducted by Rappolt-Schlichtmann et al. (2013). There are several reasons why implementation was not ideal. During the first year of the study, a natural disaster occurred, which impacted the whole district and resulted in challenges to consistent implementation, especially during the first few months of the study. In addition, teachers at times had inconsistent access to the tablets, making it challenging to routinely incorporate SNUBLE into their science lessons as planned. Finally, the study represents first-time use of SNUBLE by teachers, which required them to get up to speed quickly and may not have provided sufficient exposure to achieve mastery.

While the findings from this efficacy study are promising, future studies should consider the potential of using SNUBLE under conditions where it is fully integrated and implemented with no systemic disruptions, such as natural disasters, over multiple years. Replication efforts are needed to substantiate current findings and allow a cumulative synthesis of results. Future research may address the limitations of the current work by collecting individual SNUBLE usage data (instead of the mixture of individual and small group usage data collected by the current study) over time, and clarifying the mechanism behind the impact, particularly focusing on critical features of implementation and how they are associated with student science outcomes. Ideally, future studies should use a randomized controlled trials design (for example, by randomizing teachers to different levels of teacher training and ongoing support, or by randomizing students to groups that use different combinations of SNUBLE critical features) and follow the participants over multiple years. Doing so would allow future studies to examine which implantation features impact which student outcomes and whether the impact of SNUBLE fades or intensifies over time. As mentioned by a previous study (Paek and Fulton, 2021), the impact of a digital notebook is limited by students' ability to use the tool. Future studies should also examine how teachers can enhance student learning by helping them understand the value of each feature and how to use each feature in different science inquiry phases.

Given the identified limitations and challenges to implementation fidelity, the current data also lead to additional questions and rich opportunities for ongoing research regarding teacher supports for SNUBLE. The current SNUBLE implementation included two sessions of teacher training, updates by the research team, and coaching by onsite science specialists. Future studies may address questions such as: What impact would further coaching and supports for teachers have on student outcomes? What impact might a varied menu of training, mentoring, and ongoing coaching have on these outcomes? How could providing teachers with access to data

visualizations of student behavior and performance impact teacher data-based decision making for science instruction? We intend to consider these questions in future studies and call upon the field to address these and other questions to support building students' understanding and sense-making skills in science education at the elementary level.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The CAST institutional review board approved this study. Researchers obtained signed consent forms from all adult

REFERENCES

- Baxter, G. P., Bass, K. M., and Glaser, R. (2001). Notebook Writing in Three Fifth-Grade Science Classrooms. *Elem. Sch. J.* 102, 123–140. doi:10.1086/499696
- CAST (2018). About Universal Design for Learning. Available at: <http://www.cast.org/our-work/about-udl.html>. (Accessed January 15, 2020).
- Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., and Glaser, R. (1989). Self-Explanations: How Students Study and Use Examples in Learning to Solve Problems. *Cogn. Sci.* 13 (2), 145–182. doi:10.1207/s15516709cog1302_1
- Chi, M. T. H., and Wylie, R. (2014). The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educ. Psychol.* 49 (4), 219–243. doi:10.1080/00461520.2014.965823
- Crawford, B., and Capps, D. (2016). *What Knowledge Do Teachers Need for Engaging Children in Science Practices? Cognition, Metacognition, and Culture In STEM Education*. New York: Springer.
- Englert, C. S., Raphael, T. E., Fear, K. L., and Anderson, L. M. (1988). Students' Metacognitive Knowledge about How to Write Informational Texts. *Learn. Disabil. Q.* 11 (1), 18–46. doi:10.2307/1511035
- Fairbanks, A. M. (2013). Changing the Role of K-12 Teachers. *Education Week* 32 (32), S4–S7.
- Graham, S., and Hebert, M. (2010). *Writing to Read: Evidence for How Writing Can Improve Reading*. New York: A report from Carnegie Corporation of.
- Graham, S. (1990). The Role of Production Factors in Learning Disabled Students' Compositions. *J. Educ. Psychol.* 82, 781–791. doi:10.1037/0022-0663.82.4.781
- Hair, E., Halle, T., Terry-Humen, E., Lavelle, B., and Calkins, J. (2006). Children's School Readiness in the ECLS-K: Predictions to Academic, Health, and Social Outcomes in First Grade. *Early Child. Res. Q.* 21, 431–454. doi:10.1016/j.jecresq.2006.09.005
- Hargrove, T. Y., and Nesbit, C. (2003). *Science Notebooks: Tools For Increasing Achievement across the Curriculum (ERIC Document Reproduction Service Number ED 482720)*. Columbus, OH: ERIC Clearinghouse for Science Mathematics and Environmental Education.
- Human Resources Research Organization (2016). Independent Evaluation of the Validity and Reliability of STAAR Grades 3-8 Assessment Scores: Part 2. Available at: https://tea.texas.gov/sites/default/files/Independent%20Evaluation%20of%20the%20Validity%20and%20Reliability%20of%20STAAR%20Grades%203-8%20Assessments_Part2.pdf.
- Klentschy, M. (2005). Science Notebook Essentials. *Sci. Child.* 43, 24–27.
- McNamara, D. S. (2017). Self-Explanation and Reading Strategy Training (SERT) Improves Low-Knowledge Students' Science Course Performance. *Discourse Process.* 54 (7), 479–492. doi:10.1080/0163853x.2015.1101328
- Meyer, A., Rose, D. H., and Gordon, D. (2014). *Universal Design for Learning: Theory and Practice*. Wakefield, MA: CAST Professional Publishing. Available at: <http://udltheorypractice.cast.org>.
- National Center on Instructional Media (2006). *Item Writer Guidelines for Greater Accessibility*. Retrieved on: http://ncamftp.wgbh.org/ncam-old-site/file_download/Item_Writer_Guidelines_for_Greater_Accessibility.pdf.
- National Research Council (2012). "A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas," in *Committee on a Conceptual Framework for New K-12 Science Education Board on Science Education, Division of Behavioral, Social Sciences and Education* (Washington, D.C: The National Academies Press).
- National Research Council (NRC) (2011). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Washington, DC: National Academy Press.
- Next Generation Science Standards (NGSS) (2018). Disciplinary Core Idea (DCI) Arrangements of Standards. Available at: <http://www.nextgenscience.org/search-standards-dci>.
- Northwest Evaluation Association (2011). Technical Manual for Measures of Academic Progress (MAP) and Measures of Academic Progress for Primary Grades (MPG). Portland, OR. Retrieved from: <https://www.richland2.org/RichlandDistrict/media/Richland-District/AdvancED/Standard%205.1/5-1-NWEA-Technical-Manual-for-MAP-and-MPG.pdf>.
- Paek, S., and Fulton, L. (2021). Digital Science Notebooks: A Tool for Supporting Scientific Literacy at the Elementary Level. *TechTrends* 65, 359–370. doi:10.1007/s11528-020-00579-0
- Rappolt-Schlichtmann, G., Daley, S. G., Lim, S., Lapinski, S., Robinson, K. H., and Johnson, M. (2013). Universal Design for Learning and Elementary School Science: Exploring the Efficacy, Use, and Perceptions of a Web-Based Science Notebook. *J. Educ. Psychol.* 105 (4), 1210–1225. doi:10.1037/a0033217
- Raudenbush, S., and Bryk, A. (2002). *Hierarchical Linear Models*. 2nd ed. Newbury Park, CA: Sage.
- Richards, K. A. R., and Hemphill, M. A. (2018). A Practical Guide to Collaborative Qualitative Data Analysis. *J. Teach. Phys. Educ.* 37 (2), 225–231. Available at: https://libres.uncg.edu/ir/uncg/f/M_Hemphill_Practical_2018.pdf. doi:10.1123/jtpe.2017-0084
- Rose, D. H., and Meyer, A. (2002). *Teaching Every Student in the Digital Age: Universal Design for Learning*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Ruiz-Primo, M. A., Li, M., Ayala, C., and Shavelson, R. J. (2004). Evaluating Students' Science Notebooks as an Assessment Tool. *Int. J. Sci. Educ.* 26, 1477–1506. doi:10.1080/0950069042000177299
- Swanson, H. L. (1999). Reading Comprehension and Working Memory in Learning-Disabled Readers: Is the Phonological Loop More Important Than the Executive System. *J. Exp. Child. Psychol.* 72, 1–31. doi:10.1006/jecp.1998.2477
- Thurlow, M., and Wiley, H. I. (2004). Almost There in Public Reporting of Assessment Results for Students with Disabilities. (Technical report 39). University of Minnesota, National Center on Educational Outcomes.

study participants and signed parent/guardian consent forms and student assent forms for all children in the study.

AUTHOR CONTRIBUTIONS

Study conception and design, JB, JY, and TH. Acquisition of data, KR, TH, JY, and KF. Analysis of Data, XW, JY, AO, TH, and JB. Interpretation of data, JY, KF, AO, TH, and XW. Drafting of manuscript, JY, XW, AO, TH, and KF. All authors contributed to the article and approved the submitted version.

FUNDING

This material is based upon work supported by the U.S. Department of Education's Institute of Education Sciences under Grant R324A160008.

- VERBI Software (2019). *MAXQDA 2020 [computer Software]*. Berlin, Germany: VERBI Software. Available at: maxqda.com.
- W3C Web Accessibility Initiative (WAI) (2018). W3C Accessibility Standards. Available at: <https://www.w3.org/WAI/standards-guidelines/>.
- Wee, B., Shepardson, D., Fast, J., and Harbor, J. (2007). Teaching and Learning about Inquiry: Insights and Challenges in Professional Development. *J. Sci. Teach. Educ.* 18 (1), 63–89. doi:10.1007/s10972-006-9031-6
- What Works Clearinghouse (WWC) (2017). *Procedures and standards handbook version 4.0*. Available at: <https://ies.ed.gov/ncee/wwc/handbooks>.
- Wilmes, S. E. D., and Siry, C. (2020). Science Notebooks as Interactional Spaces in a Multilingual Classroom: Not Just Ideas on Paper. *J. Res. Sci. Teach.* 57 (7), 999–1027. doi:10.1002/tea.21615
- Zhai, F., Brooks-Gunn, J., and Waldfogel, J. (2011). Head Start and Urban Children's School Readiness: a Birth Cohort Study in 18 Cities. *Dev. Psychol.* 47, 134–152. doi:10.1037/a0020784

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Yu, Wei, Hall, Oehlkers, Ferguson, Robinson and Blackorby. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.