Merit or Marketing?: Evidence and Quality of Efficacy Research in Educational Technology Companies

Chris S. Hulleman, Ryan A. Burke, and Maggie May, University of Virginia

David B. Daniel, James Madison University

Mahnaz Charania, Fulton County Schools

Last Revised: May 31, 2017

This white paper was produced by Working Group D as part of the EdTech Academic Efficacy Symposium, held on May 3-4, 2017, in Washington, DC., and sponsored by Jefferson Education, Digital Promise, and the Curry School of Education at the University of Virginia. Please cite this report as:

Hulleman, C. S., Burke, R. A., May, M., Charania, M., & Daniel, D. B. (2017). *Merit or marketing?: Evidence and quality of efficacy research in educational technology companies*. White paper produced by Working Group D for the EdTech Academic Efficacy Symposium. University of Virginia: Charlottesville, VA.

INDEX

INDEX	2
MEMBERS OF WORKING GROUP D	3
PURPOSE OF THE RESEARCH	4
EXECUTIVE SUMMARY	5
INTRODUCTION	7
THE ROLE OF RESEARCH IN EDTECH PRODUCT DEVELOPMENT	7
CURRENT STUDY	8
METHODOLOGY & PARTICIPANTS	9
FINDINGS & INSIGHTS	12
DISCUSSION	22
LIMITATIONS	25
NEXT STEPS	26
REFERENCES	27
APPENDIXES	28

MEMBERS OF WORKING GROUP D

WORKING GROUP CO-LEADERS

David B. Daniel, Professor of Psychology, James Madison University
Mahnaz Charania, Director of Strategic Planning, Research, & Evaluation, Fulton County Schools

RESEARCH LEAD

Chris Hulleman, Research Associate Professor of Education and Psychology, University of Virginia

LOGISTICS LEAD

Isabelle Callaway, Duke University

WORKING GROUP MEMBERS

Researcher/Think Tank	David B. Daniel (Lead)
Researcher/Think Tank	Britte Cheng
Researcher/Think Tank	Andy Rotherham
Researcher/Think Tank	Jill Buban
K12 Entrepreneur	Mark Schneiderman
K12 Entrepreneur	Jason Lange
K12 Entrepreneur	Julie Young
K12 Entrepreneur	Eva Dundas
Investors/Philanthropists	Michael Staton
Postsec Entrepreneur	David Wiley
Postsec Entrepreneur	Pete Wheelan
District Lead	Mahnaz Charania (co-Lead)
Assigned Research Lead	Chris Hulleman
Assigned Logistics Lead	Isabel Callaway

PURPOSE OF THE RESEARCH

The original purpose of our research was outlined as follows by Jefferson Education in June of 2016:

> Of the efficacy research being conducted on EdTech products, who is doing the research, what kind of research is it and what is the quality? For example, how many EdTech companies are doing research of any type? What type of research are they doing and what type of research approaches and criteria/metrics should they be using to evaluate product efficacy? How reliable, unbiased and rigorous are these efficacy studies and how can we better capture the context in which Ed Tech products are tested?

Following several meetings of our working group during the early fall, we settled on the following more focused purpose of our group:

> Our group aimed to understand the extent and quality of efficacy research conducted on edtech tools by K-12 Ed Tech companies who are geared toward instruction and curricula that directly affects student-level outcomes. Our investigation focused on three specific questions:
>
> 1. What kind of research is being conducted by Ed Tech companies on their products (e.g., efficacy, effectiveness, scale-up)?
> 2. Of the efficacy research being conducted on Ed Tech products, who is doing the research and with what quality (e.g., rigorous, reliable, unbiased)?
> 3. To what extent do claims about product efficacy by Ed Tech companies match the actual research?

By understanding the extent and quality of research conducted by EdTech companies, we hope to identify strategies to improve the support and accountability for reliable, research-based development of education technology tools by developers.

EXECUTIVE SUMMARY

Our group aimed to understand the extent and quality of efficacy research conducted on edtech tools by K-12 Ed Tech companies who are geared toward instruction and curricula that directly affects student-level outcomes. Our investigation focused on three specific questions:

1. What kind of research is being conducted by Ed Tech companies on their products (e.g., efficacy, effectiveness, scale-up)?
2. Of the efficacy research being conducted on Ed Tech products, who is doing the research and with what quality (e.g., rigorous, reliable, unbiased)?
3. To what extent do claims about product efficacy by Ed Tech companies match the actual research?

A mixed-methods online survey was developed for EdTech companies to respond to regarding their product's design, development, impact, and usage. Between 3/23-4/28, we received responses from 40 companies from our original list of 142 (28%). We also conducted an extensive investigation of the 59 (41%) of the 142 companies that made some claim of product efficacy on their websites or promotional materials. First, we used the scoring rubric produced by Digital Promise to score the quality of each of the research studies on a scale from 1 (low quality) to 18 (high quality). Second, we compared the degree of alignment between the efficacy claims made by the company and the actual research.

Based on the survey results, although Ed Tech developers value research to inform their products, they are not conducting rigorous research. For example:

- Most companies reported utilizing user feedback and engagement and do not conduct efficacy research. A minority of companies who spoke about efficacy studies being a priority stated that they infuse their product development with useful data on lesson completions and create multiple pathways to improve on the product overtime.
- Generally, 50% of companies conduct their research internally, 11% leverage exclusively external partners to collect data, and 39% use both external and internal sources.

Furthermore, the only a minority of companies have high quality research that matches their claims about product efficacy. For example:

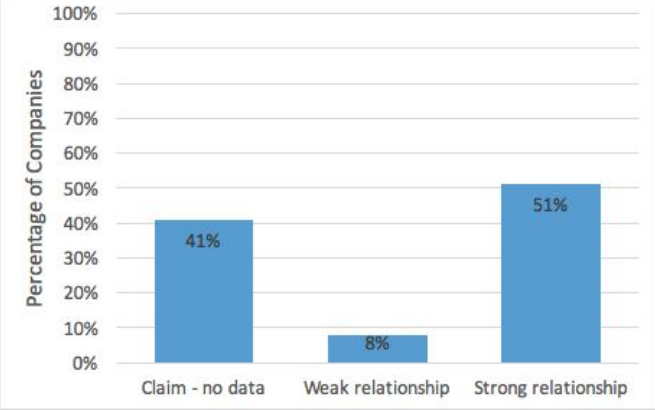
- Of the 59 companies that made efficacy claims in the online materials, only 35 (59%) had actual research to potentially support those claims, whereas we could not find research evidence for the remaining 41% of companies.
- Of the 35 companies who had research evidence, only 4 (7%) could be considered high quality research according to the Digital Promise scoring rubric.

In sum, the availability and quality of efficacy research is at odds with companies' confidence about their product's efficacy. For example:

- The vast majority of respondents (88%) reported that they were very or extremely confident that their product has the intended impact.
- However, the vast minority of companies (7%) who made efficacy claims of their product's impact had high quality research to support these claims.

Figure 1. Survey respondents' confidence that their products impact student learning outcomes.

Figure 3. Relationship between efficacy claim and research evidence.



INTRODUCTION

Digital tools are increasingly utilized in classrooms across the nation in an effort to create learning environments that blend face-to-face and digital learning. As emphasized by Horn and Staker (2014), such learning environments can be particularly impactful on student learning outcomes when schools are able to strategically use targeted digital tools and resources to support student learning. Over the past five years, researchers and education leaders have been particularly interested in the value-added of education technology tools to student learning outcomes. In a survey conducted by Project Tomorrow (2014), 98% of school and district administrators say that effective use of technology within instruction is important for achieving their core mission of education and preparation of students. A central question that remains critical to informing this effective use is the degree of research evidence available to school and district administrators regarding the impact of the technology on student learning outcomes, and guidelines to ensure effective implementation. This review focuses on the degree of research evidence provided by education technology (EdTech) companies on products designed to directly promote learning outcomes for K-12 students.

THE ROLE OF RESEARCH IN EDUCATIONAL TECHNOLOGY PRODUCT DEVELOPMENT

EdTech companies use a multitude of avenues in their research to assist them in their decision-making process on whether to continue developing or marketing their products. This research takes various forms such as case studies (examines how the product was implemented in a specific environment), pilot studies (examines the product for a set period of time within a school/classroom or setting), randomized controlled trials (examines the outcomes between groups given the product and groups that do not), or short cycle evaluations (examining a product by matching companies with NYC school teams to pilot, refine, and evaluate the edtech) (Digital Promise, 2016; “Short cycle evaluation challenge”, 2015). The research design and evaluation of the company’s product efficacy may be influenced by the stage in which the company is in their development. A goal of efficacy research in the EdTech sector is to help schools make informed decisions on which product is appropriate for their classrooms. Observable trends in this field show that more developed companies that have been in existence for years are more likely to conduct a full-scale evaluation research (Digital Promise, 2016).

Digital Promise (2016) conducted an investigation of efficacy research and its relation with schools' decision making processes in choosing products, research spending toward educational technologies, and the approaches dedicated toward the evidence and quality of products. This examination found areas in which efficacy research needs improvement. A common limitation in product efficacy studies is the lack of objective information for educators. This introduces a deficiency in providing teachers with the tools needed to implement the products in their classrooms. Another limitation is the inadequate amount of peer-reviewed research; this conveys that there is not a sufficient process in which the research can be critically examined. A further gap within the research that needs to be addressed is the translation of the research. Districts or administrators are typically making the decisions for the implementation of these products, yet they are not the primary populations that are using these tools. A final limitation is the lack of consistency in communicating research findings to potential users. For example, there is no method of accountability for companies who make claims about product efficacy but have little or no research to support those claims.

Our brief review of the literature reinforces the conventional wisdom that there are many gaps in how companies approach research that could provide direct efficacy evidence that their products promote student learning outcomes. To further explore these gaps, we investigated the extent to which EdTech companies utilized research, how they conducted the research, and whether their efficacy claims for their products matched the available research.

CURRENT STUDY

Our investigation focused on three specific questions:

1. What kind of research is being conducted by Ed Tech companies on their products (e.g., efficacy, effectiveness, scale-up)?
2. Of the efficacy research being conducted on EdTech products, who is doing the research and with what quality (e.g., rigorous, reliable, unbiased)?
3. To what extent do claims about product efficacy by EdTech companies match the actual research evidence?

To conduct this investigation, we surveyed K-12 EdTech developers and evaluated the efficacy claims from promotional materials obtained from company websites.

METHODOLOGY & PARTICIPANTS

To identify EdTech products that targeted student learning outcomes, we utilized several existing lists of educational technology products from EdSurge Curriculum and Products list. We also conducted our own review of popular technology products known to us or available online. Our initial list of over 400 companies was narrowed to 142 who met our criteria of having an educational technology product geared toward instruction and curricula and that directly targeted student learning outcomes. Some examples include Reading Plus, Big History Project, and the ACT Career Curriculum (the complete list is presented in Table 1). The list was curated by two of the co-authors and confirmed by a third author.

Survey

Through group discussion, we developed a survey that focused on the type and quality of EdTech efficacy research that was being conducted on products that had direct links to student learning outcomes. The survey has 4 main sections: Product design, Product impact, Product development, and Product usage; and includes a mix of open-ended and close-ended questions. Example questions include (the complete version can be found here):

- **Product Design:** What informed the creation or initial design of your EdTech product? For example, is the design of your EdTech product based on personal experience, classroom observations, existing research in some way, etc.?
- **Product Impact:** What is the intended impact of your product? For example, is it intended to help students learn faster? Learn more in the same amount of time? Remember what they've learned longer? Learn the same amount while saving money (Compared to the cost of other products)? Save teachers time? Get parents more involved?
- **Product Development:** Do you have a standard development process for demonstrating the impact of your products? What type of information do you collect? Who collects it? How is it collected?
- **Product Usage:** How many districts/schools/students currently use your product?

The survey was developed using the online survey tool, Qualtrics, and pilot tested on several EdTech product developers and revised accordingly before arriving at

our final version. Between March 23 and April 28 , 2017, the survey link was electronically sent to the 87 EdTech companies in which we could identify contact information. Based on the classification of these 87 EdTech companies on EdSurge, 45 (52%) companies created products that are focused on STEM, 30 (34%) companies focused their products on language arts and language learning, 9 (10%) companies emphasized on 21st century skills in their products, and 3 (4%) companies geared their products towards arts and social studies. Of those 87 companies, a total of 40 companies (47% of the original 87 that were contacted, and 28% of the original 142 companies on our list) completed the survey.

Because respondents did not have to answer every question, some of the information presented will have fewer than 40 responses. Some of the questions had the option to choose multiple answers or were free responses, therefore respondents could provide more than one answer per question.

Evaluation of Efficacy Claims

In order to evaluate claims of product efficacy made by EdTech companies, we visited the websites of all 142 companies on our list. We found that 59 (41%) companies made some claim of product efficacy. We then found the research studies from each of the 59 company's websites, or used online search engines to search for the actual studies referenced in the promotional materials. The companies were subsequently evaluated in two ways. First, we used the scoring rubric produced by Digital Promise to score the quality of each of the research studies on a scale from 1 (low quality) to 20 (high quality). Specifically, we used the sections of the rubric to score the study source (0-15 points) and study design of the research (0-5 points). Thus, scores between 1-8 represented low quality research, scores between 9-14 represented moderate quality research, and scores between 15-20 representing high quality research. Each of these quality research scores were consensus coded by two researchers. In the coding of the data, there was a total agreement or inter-rater reliability between the researchers of 81%. Disagreements were resolved through consensus discussions.

Second, we compared the degree of alignment between the efficacy claims made by the company and the actual research. The alignment between efficacy claims and actual research was coded as strong or weak. Scores of "Strong" reflect that the results of the study in which the company conducted was aligned with their subsequent efficacy claims (e.g., if a product claimed it increases student test scores then the research needed to provide some evidence that the product increased test

scores). Weak suggested that the results from their study(ies) did not support their claim. This occurred when the dependent variable was different than the claim (e.g., if a product claimed to increase student test scores but the research only included student satisfaction and engagement with the product).

These two categories then enabled us to place companies into respective categories based upon both their quality of research scores given by the Digital Promise scoring rubric (Low, Moderate, High) and how well their research matched their claims (Weak, Strong). All companies that made efficacy claims were evaluated, but some companies made claims without subsequent research giving them a category of "Claim, but no research", and one company had research, but no efficacy claim, giving it the category of "Research, but no claim". The companies could have fallen within the following categories:

	Low	Moderate	High	Research, but no claim	Claim, but no research
Weak	The Research Design had a Weak Score (1-8) and the results do not match the claim	The Research Design had a Moderate Score (9-14), but the results do not align with the claims	The Research Design Scored High (Above 15) but the results do not align with their claims	The company has research, but did not make an efficacy claim	The company makes an efficacy claim, but has no research
Strong	The Research Design had a Weak Score (1-8) but the results match the claims	The Research Design Had a Moderate Score (9-14) and the results align with the claims	The Research Design Scored High (Above 15) and the results align with their claims		

FINDINGS & INSIGHTS

Surveys - Quantitative Results

It is first important to note that we received 40 responses to our survey, which represents 28% of our original sample of 142 companies. This means that our findings are likely biased for numerous reasons; not the least of which is that companies that who do more research would be more likely to respond, and companies that do less research would be less likely to respond. That means that the companies in our sample likely conduct more research than those who did not respond to our survey.

Product Usage

Of those respondents that answered the demographic questions, 22 (73%) had products that were used by over 10,000 students per year, 20 (69%) had products that were used by over 50 schools per year and serve 10 or more districts. From the companies that answered the demographic questions, 14 (50%) had products that were used by over 1000 teachers.

Product Design

In coming up with the design of the product, there are several things that help inform the creation of the initial design of EdTech products. Personal experience was one of the factors, which was mentioned by 28 (70%) respondents. Other factors include classroom observations, mentioned by 24 (60%) respondents; company internal research, mentioned by 29 (73%) respondents; and external research, mentioned by 33 (83%) of the respondents. Thirteen (33% of respondents) companies mentioned other reasons such as teacher input, policy, demand for the product, and past research or analyses as influences for the initial design of their product.

When examining the kinds of research that are being conducted in the field, the survey responses shed some light into what is currently going on. Three (14% of respondents) of the responses mention conducting efficacy research in the laboratory, whereas 30 (73%) of the responses mention that the research is conducted in the field (randomized field experiment was indicated in 12 (57% of respondents) responses, quasi-experimental field study was indicated in 9 (43% of respondents) responses, and pre-post test with no control group was indicated in 9 (43% of respondents) responses). After the research has been conducted, most of the results were reported in published academic research, indicated in 20 (95% of respondents) responses. Aside from these 20 responses, 9 (43% of respondents) responses indicate that the results are disseminated in an unpublished academic

research (e.g., conference presentation), 10 (48% of respondents) responses show that research results are being communicated in unpublished technical report from inside the company, and 4 (19% of respondents) responses disclose that results are reported in unpublished technical report from outside the company.

Product Impact

Respondents were further asked to rate their confidence on the product's ability to meet its intended impact. A majority of respondents (88%) were very confident (50%) or extremely confident (38%) that their product meets its intended impact. Only three (9%) companies described their product as having moderate impact and only one (3%) company was slightly confident. There were no companies that were not at all confident (see Figure 1).

When examining whether the respondent's company collects any information, data, or evidence used to determine whether or not the product is making the difference the company intends it to make, 33 (94%) of respondents said yes, whereas only one (3%) responded with no and one (3%) more responded with unsure. More specifically, when examining the type of data companies collect in order to make these determinations, 26 (81%) respondents mentioned they collect data on student performance, 26 (81%) mentioned student satisfaction, 29 (91%) mentioned student engagement, 27 (84%) mentioned teacher satisfaction, and 23 (72%) mentioned teacher engagement. There were 10 (31%) other responses which did not fit these criteria which include, partner satisfaction, qualitative studies, benchmark surveys, predictive proficiency studies, proficiency improvement over-time, real world evidence of success, school satisfaction with the product, comparing enrollment data, most is internal to our system, classroom observations, administrator feedback, student progress and in system behavior.

Respondents were also asked to reflect on how often customers request further research evidence. The companies were mixed in how often they reported that customers asked for evidence of product impact, with 1 company responding that customers never ask for evidence of product efficacy, 9 (26%) companies responding that customers rarely ask for it, 7 (21%) companies asserting that customers sometimes ask for evidence, 15 (44%) companies indicating that customers often ask for evidence that the product does what it says it will do, and 2 (6%) companies revealing that customers always ask for evidence (see Figure 2).

Product Development

There are different types of information that EdTech companies collect to inform their research. Out of the 18 companies that responded to this question, most if not all of them indicated more than one type of information being collected. Eleven (61%) responses indicate using student outcomes data, 9 (50%) responses show that companies collect recorded time that consumers use the product, 2 (11%) responses

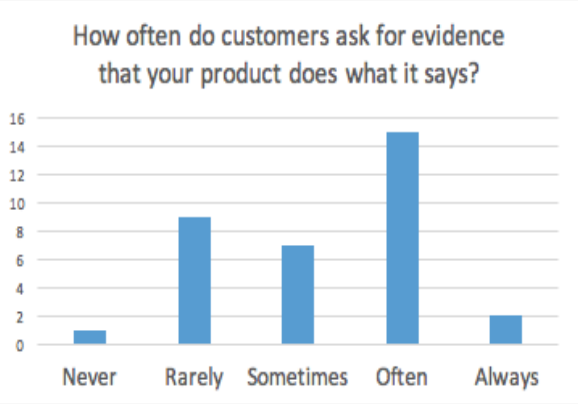
indicate that they collect demographics data, 16 (89%) responses disclose that companies collect survey data from teachers and students, 4 (22%) responses show that companies collect product performance data, and 3 (17%) responses indicate that companies collect data from past research. These different types of data were collected mostly through the company platforms, as indicated by 14 (78%) of the respondents. The data is also collected through surveys as reported by 6 (33%) respondents, and as teachers and students use the products, indicated by 2 (11%) of the respondents. In addition, 4 (22%) respondents indicate that data were collected through the use of school or state administered assessment.

In terms of who conducts the research, 9 (50%) companies reported that they conduct their research internally, 2 (11%) of the companies only leverage external partners to conduct the research, and 7 (39%) companies conduct their research both internally and in partnership with external research agency. The research data is collected internally for 15 (83%) out of the 18 companies (83%) who indicate that they do some kind of research, and the other 3 (17%) companies reported that the data is collected both internally from their company platform and from also from outside sources.

In addition, EdTech companies were asked to identify whether they have a standard development process for demonstrating the impact of their products. From those who responded, 15 (47%) companies responded that they have no standard development process, followed by 9 (28%) companies that reported that they have a standard process for demonstrating the impact of their products, and 8 (25%) companies that were unsure.

Figure 1.

Figure 2.



Evaluation of Online Efficacy Claim Data

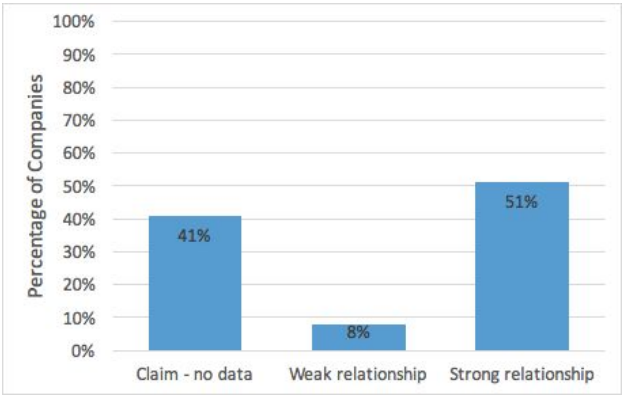
Our review of online efficacy claim data evaluated companies on whether or not they made efficacy claims of their products, and which claims matched the company's subsequent research. Of the 59 companies that made efficacy claims in the online materials, only 35 (59%) had actual research to potentially support those claims, whereas we could not find research evidence for the remaining 41% of companies.

In terms of research quality, of those 35 companies for whom we found research evidence, their quality scores ranged from 6 to 20, with an average score of 11.6 (*SD* = 3.42). A score of a 11.6 is classified as moderate quality by Digital Promise with scores from 1 to 8 categorized as low quality, 9 to 14 as moderate quality, and 14 to 20 as high quality. As presented in Table 2, the research of eight products was categorized as low quality, the research of 23 products was categorized as moderate quality, and the research of four products was categorized as high quality.

In terms of how well the actual research supports the efficacy claims, we found that 30 (51%) companies showed a strong relationship between their research and their claim, and five (8%) companies showed a weak relationship between their research and their claim.

Of the 35 companies that offered efficacy claims, these companies were further broken down based upon their research quality (Low, Moderate, High) and how well the research supported the company's efficacy claims (Weak, Strong) (See Table 3). Three (9%) companies were identified as Low Weak, five (14%) companies were identified as Low Strong, two (6%) companies were identified as Moderate Weak, 21 (60%) companies were identified as Moderate Strong, and four (11%) companies were identified as High Strong.

Figure 3. Relationship between efficacy claim and research evidence.



Surveys - Qualitative Data

Our coding of the open-ended survey responses data indicated that most companies do not use efficacy research in the product development phase. Instead, a majority of companies utilize user feedback and engagement. Here are a few representative quotes from respondents when asked how they used efficacy data in product development:

- *"Future product improvements are all about improving engagement for both students and teachers. Efficacy does not play much of a role in our plans."*
- *"We rely mostly on user feedback and classroom observations."*
- *"Data on efficacy is mostly unrelated to our research on product improvement. Figuring out if our product achieves outcomes is a totally different process and set of data from figuring out how our product can be improved."*

However, some respondents did indicate that efficacy data is an important aspect of product development. Several companies identified that this research is important in analyzing the needs of students, improving their product, informing teachers how to use the product, and research is a means to demonstrate efficacy to stakeholders and customers. Some examples of the responses in which companies indicated the importance of efficacy data include:

- *"Our ongoing efficacy research agenda include multiple milestone check-ins where results learned are provided to our production team to help inform subsequent releases of our online product. For example, if evidence suggests certain instructional lessons are not producing expected results, this information is provided to our curriculum team who refines the lessons, and refined lessons are released in updates to the system."*
- *"We collect data on completion and pass rates of individual lessons. Based on*

benchmarking, we improve these lessons to make sure they achieve above a certain baseline."
- *"We write up results to share with our teachers and prospective customers. We use this report as the kick off to our annual planning process each year."*

Many companies indicated several different sources of information that their customers count as evidence in the product's efficacy. Trends in the responses found that companies have a perception that their consumers primarily focus on student performance after engaging with the product (43% of respondents), followed by academic research (40% of respondents), testimonials or case studies (31% of respondents), and recommendations and reviews (22% of respondents). Some examples of what companies think customers count as evidence include:

- *"Engagement and retention of students on our platform and completion rates and grades they get on our online course automatic assessment."*
- *"Ha. In this market, they will count ONE study with a brand name on it as sufficient evidence the program works!"*
- *"I think the testimonials from students and teachers is all I have to offer until I can find a larger audience and collect a meaningful amount of data."*

Further investigation revealed that personal experience, research (internal or external), and classroom observations influenced the design and creation of the product. In terms of personal experience, the respondents reflected upon either positive or negative experiences as students or employees of schools which subsequently led to the development of these different products. Some examples of personal experiences that influenced product development include:

- *"I was a private Algebra tutor. The software replicates key things I did: check each step as soon as it is done, offer hints when asked, demonstrate solutions to problems similar to the one with which they are struggling, and offer moral support."*
- *"Both co-founders are speech language pathologists. The current vocabulary tools in their toolboxes were not working and felt that using language to teach language to students who struggle with language was not working. 65% of students are below proficient in reading comprehension and vocabulary is a building block for comprehension. It was time for a change in how we teach students vocabulary."*
- *"As a computer science teacher, i found that students had difficulty with the abstract nature of the programming process. I was interested in introductory*

tools that made the abstract concrete."

When companies responded that the product has been developed through classroom observations, the respondents stressed the importance of both the needs of the students and the teachers. Some of the examples include:

- *"We have a small group of teachers that we work with before any learning module's general release. Observations from the classroom help us find and fix potential pain points in any lesson."*
- *"Similar to the above and refinements were made during 4 years of use within our school district. Student and teacher feedback on all aspects of the activities helped us streamline the scaffolding and improve collaborative and competitive elements as well as compact the learning activities and the training required to put the program use. In addition, requests for certain wordlists provided direction as we expanded the Tier II words available for learning."*

When companies were specifically asked about the intended impact of their product there was a large consensus that the products were designed to increase learning in students followed by supporting educators. Some examples include:

- *"Our intention is to bring high quality coding instruction into the general middle school curriculum. Long term, this helps to fix the gender and race gaps associated with the study of computer science. We also save teachers time by providing a built out curriculum with mostly auto-graded lessons."*
- *"Our product is intended to help both students and schools. Students gain critical exposure to rigorous online courses that require, in addition to content expertise, clear communication, self-advocacy, self-regulation, and executive function skills. Feedback from students via student survey data and course satisfaction feedback indicates that courses have had their intended result for many students. Schools benefit from access to an extensive catalog of courses taught by certified educators. Our program provides schools with affordable access to nearly 200 additional courses that alleviate individual student scheduling challenges, fill needs that would have required small sections, and/or ensure all courses are taught by credentialed teachers."*

Table 1. Educational technology companies included in our review.

- 3 Digits
- 51 Talk
- ABA English
- Academy of Math
- ACT Career Curriculum
- AdaptedMind
- ALEKS
- Algebra Nation
- Alice
- Amplify ELA
- Apex Learning and Comprehensive Courses
- AppStudio
- Ascend Math
- Avaz Freespeech
- Babbel
- Bankaroo
- BetterChinese
- Big Brainz** (subset of Imagine Learning)
- Big History Project
- Big Word Club
- BioBeyond
- Brainingcamp
- Brainology
- BrainPop
- Busuu
- Callido
- Cell-Ed
- Code Red
- Education Coding Curriculum
- Code School
- Codecademy
- CodeHS
- CodeMonkey
- Codesters
- Core Skills Mastery
- CreatorUp
- DefinedSTEM
- Digital K-12 World
- Language
- Learning
- DragonBox
- Foundations
- Fundamentor
- GamesforLanguage
- GameSpark
- Globaloria
- Hakitzu
- HeyMath!
- i-Ready
- Imagine Learning
- Imagine Learning Espanol
- Immersia
- InferCabulary
- Inquiry
- Integrated Science Program
- Istation Espanol
- Istation Reading
- Jade Robot
- Joytunes Games
- Khan Academy
- Knewton Math Readiness
- Komodo
- LanguageNut
- Learn with Homer
- LearntoMod
- Lentil
- Life Stages
- Lincoln Empowered
- LingoJingo
- Lingokids
- Little Pim
- Logics Academy
- Logtera
- MAD-Learn
- Math Score
- Math Snacks
- Mathseeds
- Mathspace
- Meludia
- Merit Software
- MindSage SEL and MESH Program
- Mindsnacks
- Mission US
- Mobile Art Academy
- Number Worlds
- OnlineFreeSpanish
- Ooka Island
- Pathblazer
- Pup's Quest for Phonics
- Quill
- Reading Eggs
- Reading Plus
- Reasoning Mind
- Edifice
- Reflex
- ReKode Education
- Rithm School
- Rosetta Stone
- S.P.IN.
- SAS Curriculum Pathways
- Science Fusion
- Scratch
- Singspiel PianoPal
- Smart Tutor
- Smarthistory
- Sofatutor
- Sophia
- SpellNow Series
- St Math
- Standard Deviants
- Accelerate
- SuccessMaker
- Summer Academy by Make School
- Symphony Math
The Foos
- The Science Game
- ThotsLab
- Thrive 'n' Shine
- Tilton's Algebra
- Topper Learning
- Tripppin
- Tritone Music Education System
- Tynker
- Universal Zoom
- VHS Online & Blended Learning Solution

• DreamBox • Learning • Edgenuity • EdTechLens • ELL Digital Curriculum • English Attack • Essential Skills • FASTT Math • Flatiron School	• MobyMax • myBlee Math • Native Accent • NCLab • Ninchanese • NovaNet Coursework	• VIPABC • Voxy • Wespeke • Where Does My Money Go • Word Lab Web • Wordplay • WordQuations

Table 2. Educational technology company research classified by Digital Promise scores

Low Quality (scores between 1-8)	Moderate Quality (scores between 9-14)	High Quality (scores between 15-20)
• ACT Career Curriculum • Brainology • DreamBox Learning • Essential Skills • Inquiry • Logtera • Reflex • Symphony Math	• Academy of Math • ALEKS • Alice • Apex Learning Comprehensive Courses • Big History Project • BrainPop • FASTT Math • Flatiron School • Foundations • Globaloria • i-Ready • Learn With Homer • Mathseeds • Merit Software • Science Fusion • Scratch • Smart Tutor • Reading Plus • Rosetta Stone • St Math • Success Maker • VHS Online &	• Algebra Nation • Cell-Ed • Istation Reading • Math Snacks

	Blended Learning Solution • Voxy	

Table 3. Educational technology company classified by Digital Promise score and alignment of efficacy claims to research

Low Weak	Low Strong	Moderate Weak	Moderate Strong	High Weak	High Strong
• Inquiry • Logtera • Reflex	• ACT Career Curriculum • Brainology • DreamBox Learning • Essential Skills • Symphony Math	• Alice • VHS Online & Blended Learning Solution	• Academy of Math • ALEKS • Apex Learning and Comprehensive Courses • Big History Project • BrainPop • FASTT Math • Flatiron School • Foundations • Globaloria • i-Ready • Learn with Homer • Mathseeds • Merit Software • Science Fusion		• Algebra Nation • Cell-Ed • Istation Reading • Math Snacks

			• Scratch • Smart Tutor • Reading Plus • Rosetta Stone • St Math • Success Maker • Voxy		

DISCUSSION

Digital learning tools have gained traction in classrooms across the globe because of their promise towards helping students learn more efficiently and at their own pace. It remains imperative that such tools are grounded in a strong evidence-base and continuously improved upon through the use of data to support student learning. Our research was aimed at understanding the extent to which Ed Tech developers conduct research to inform their product, how that research is conducted, and to propose future strategies for improving upon these important practices.

Our survey results - which much be qualified by our 28% response rate - indicate that respondents and their companies integrated some type of research in each of the four aspects of product development that we assessed: product design, impact, development, and usage. Although the focus on efficacy research was usually secondary, it was recognized by many respondents as an important aspect of product development. Our analysis of the qualitative responses suggests that Ed Tech developers value research to inform their products but, for a variety of reasons, are not conducting rigorous research. This was also reflected in the lack of a standard process for collecting efficacy data by many respondents.

Importantly, nearly all respondents (90%) reported that they were very or extremely confident that their product has the intended impact. This confidence is disconcerting given the available data on their products. Companies' claims about product efficacy did not match the available research evidence. Although a majority of companies convey great confidence in their product's ability to impact student outcomes, there is little subsequent research to verify such perceptions. Nearly half of the companies that have efficacy claims for their products produce no research to back up these assertions; and of the companies that do conduct research on their product, nearly all of these companies have research that is considered to be low to moderate quality according to the Digital Promise criteria. This constructs an

interesting portrait in the apparent discrepancies between the perceptions and actions of the companies. This discrepancy between company perceptions and reality regarding product efficacy suggests that many companies are more focused on product utilization, dissemination, and marketing, rather than creating a solid foundation of efficacy and merit.

At issue may be the processes by which these companies can, or think they do, develop useful data to support their claims. Most respondents indicated that they did not have a standard process for evaluating the efficacy of their products. About half of the survey respondents reported that they have no standard development process for demonstrating the impacts of their product. Furthermore, only about one-third of companies reported that they collect any type of student performance or achievement data. This is significant, because the majority of the data being used in the research process for many EdTech companies is collected and analyzed internally. Even though research being conducted internally has the opportunity to more appropriately account for the developer's perspective, it introduces the potential for company biases to play into the evaluation process and does not allow the potential consumer the ability to properly vet the product.

Almost as disconcerting are the sample's uneven utilization, and perhaps valuing, of efficacy research across EdTech companies. This was reflected in respondents' perceptions of whether their consumers are even asking for evidence that their product is able to meet its intended purpose. There is nearly an even split across companies in this sample their beliefs regarding consumer's need for evidence. Half of the respondents perceive that customers rarely or sometimes ask for evidence that demonstrates that the product is fulfilling its purpose, compared to the other half of companies that suggest customers often or always ask for such evidence. This discrepancy is particularly important when considering the motive of companies to conduct research, as companies may be less motivated to invest in perform higher quality research, or research altogether, if their consumer base does not actively seek or value such information. These perceptions may lead to lower incentives and demands for companies to conduct product-based research because it is not beneficial in terms of time and resources.

This discrepancy in perception highlights the important distinction between research designed to improve the product experience, and research deployed to assess potential impact. According to Simone & Gross (2017), user research describes research focused on improving the product design by specifically examining the consumer's or target population's (students, teachers, administrators and parents) interactions with the product. Through this lens of research, studies are conducted by means of interviews, surveys, observations, focus groups, data analytics, pilot study, or A/B testing. Evaluation research is described as the process of determining the effectiveness of the company's product by evaluating its effectiveness in its promotion of learning or solving problems in educations. This

latter category includes efficacy and effective research targeted by our studies, however the distinction is not necessarily used by our respondents.

Respondents to our survey indicated that user feedback and student engagement are the primary source of information for the development of their product. This does not mean that companies are unconcerned with efficacy; instead, it may instead reflect more pragmatic and immediate concerns. Concerns for which user research may be more appropriate. Examining user data allows for the company to identify certain areas for improvement in a quick and efficient manner. While this trend exemplifies the company's focus on both meeting teacher and student needs and increasing ease and frequency of use, it does not have the ability to focus on the impacts of the product over a period of time.

Utilizing research designs for the purpose of measuring product impact over time may not be efficient for companies that want to continually develop and modify their products to fit the needs of the consumers. The time and resources needed for randomized control trials are outweighed by a demand to make adjustments to content and products on a continual basis. Randomized control trials have the opportunity to provide representative information, but it may not be necessary, affordable, or in the best interest of the company to conduct this form of study (Francisco et al., 2017). This does not diminish the need for quality long term research, but it suggests that there are further aspects of research that are essential in the development and implementation of EdTech products.

Overall, our findings, suggest that the type of research Ed Tech developers are actually conducting to inform the uptake of their products may be influenced by what is expected and being asked of them by consumers. On the gloomy side, most companies that make claims about product efficacy failed to support these claims with even moderate quality research. Given the economic pressures on companies to support their financial bottom lines, it is not surprising that EdTech companies are more likely to favor efficiency over longer term impact when it comes to their products. On the relatively sunnier side, the majority of developers are leveraging some type effectiveness research in lieu of efficacy research. Although most developers are focusing their outcomes on implementation fidelity and ease of use, these important aspects can be viewed as a start towards rigorous research and not the end goal. There are many models of research that include design-based and implementation aspects on the pathway to demonstrating efficacy, such as design-based implementation research (Penuel et al., 2011) and improvement science (Bryk et al., 2015).

LIMITATIONS

It is important to note that this research is limited by the small response rate (only

40 of our original 142 companies responded - a 28% response rate). It is also possible that the Ed Tech companies that chose to respond to our survey may be more likely to engage in efficacy/effectiveness research than those who did not respond. Thus, it is possible that our numbers may be be a bit inflated and not representative of the larger Ed Tech community. Future conversations and research will be required to gain a more reliable understanding of strengths and opportunities for research by Ed Tech companies in general.

NEXT STEPS

Despite the inconsistencies that exist in efficacy claims and research processes in the Edtech sector, EdTech products are widely used by schools and educators across the country. The emphasis on efficacy research is still not very strong, and there are some inconsistencies in the value that different stakeholders put on efficacy research. This fosters subsequent conflicts in how companies conduct research for the development and implementations of their product. In addition, although the EdTech sector places some value on efficacy research, many companies are facing obstacles that make it difficult to engage in such evaluation practices. These obstacles include an array of things such as resources, money, or time. Subsequently, there is an added obstacle in regards to immediate demand. Most EdTech companies are faced with the pressure to cater toward research that has a greater potential to respond to the immediate demand of providing better user experience and increase frequency of use; this thus hinders companies from being able to engage in full-scale efficacy research.

Important next steps must include clarity in the type of research that is minimally necessary to inform product development and impact as well as cultivating the necessity of quality evidence before adopting EdTech by schools and districts. It is also particularly important that all stakeholders recognize the critical value of continuous improvement and identify strategies to embed these improvement practices into their systems. While one study may be informative for product adoption, ongoing research and analyses of user data as well as user impact will be useful in understanding who the product works for and under what conditions - while continuously enhancing product impact.

REFERENCES

Anderson, G. H., & Moore, S. E. (2004). Dietary proteins in the regulation of food intake and body weight in humans. *The Journal of Nutrition, 134(4),* 974S-979S.

Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to Improve: How America's Schools Can Get Better at Getting Better*. Harvard Education Press. Cambridge, MA.

Digital Promise. (2016, February). *Using research in ed-tech: What does research-based product development mean?* Retrieved from: http://digitalpromise.org/wp-content/uploads/2016/02/Using-Research-in-Ed-Tech.pdf

Francisco, A., Epstein, B., Counselman, G., Stevens, K., Chou, L., Charania, M., . . . Griffiths, R. (2017, May 1). *Movement Around Research, Impact in Ed Tech* [Blog post]. Retrieved from The 74 website: https://www.the74million.org/article/showtheevidence-building-a-movement-around-research-impact-in-ed-tech

Penuel, W. R., Fishman, B. J., Haugan Cheng, B., & Sabelli, N. (2011). Organizing research and development at the intersection of learning, implementation, and design. *Educational Researcher, 40(7)*, 331-337.

Project Tomorrow. (2014). *The New Digital Learning Playbook: Advancing College and Career Ready Skill Development in K-12 Schools [PDF]*. Retrieved from http://www.tomorrow.org/speakup/pdfs/su13educatorreport_web.pdf

Short cycle evaluation challenge. (2015, May 26). Retrieved from https://www.edsurge.com/opportunities/short-cycle-evaluation-challenge-on-2015-05-26

Simone, C. M. & Gross, K. (2017, January 30). *Using research in ed-tech: How ed-tech companies use research to make better products*. Retrieved from Digital Promise Website: http://digitalpromise.org/wp-content/uploads/2017/01/dp-using_research_in_ed_tech-1.pdf

APPENDIX A - KEY DEFINITIONS

- **Efficacy**: allows for testing of a strategy or intervention under "Ideal" circumstances, including with a higher level of support or developer involvement than would be the case under normal circumstances. Efficacy research studies may choose to limit the investigation to a single population of interest.

- **Effectiveness:** examines effectiveness of a strategy or intervention under circumstances that would typically prevail in the target context. The importance of "typical" circumstances means that there should not be more substantial developer support than in normal implementation, and there should not be substantial developer involvement in the evaluation of the strategy or intervention.

- **Scale-up:** examines effectiveness in a wide range of populations, contexts and circumstances, without substantial developer involvement in implementation or evaluation. This should be carried out with no more developer involvement than what would be expected under typical implementation.

- **Degree of reliability, bias, and rigor**:

  - Reliability = Have multiple studies been conducted or a study with multiple groups? Are the findings stable / predictable over multiple studies/groups?

  - Bias = Is the study conducted by an external partner or by the ed-tech developer?

  - Rigor = Are the methods systematic and reproducible? Does the research design match the desired inference (e.g., causal claims are not made from correlational results)?

Adapted from: Gottfredson, D. C., Cook, T. D., Gardner, F. E., Gorman-Smith, D., Howe, G. W., Sandler, I. N., & Zafft, K. M. (2015). Standards of evidence for efficacy, effectiveness, and scale-up research in prevention science: Next generation. *Prevention Science, 16(7)*, 893-926.