Considerations for Conducting Research in Digital Learning Platforms

Co-Authors: Adam Schellinger, Jenna Zacamy, Jeremy Roschelle, Avery Closser, Cristina Zepeda

April 2024

Digital Promise
Accelerating Innovation in Education
Empirical Education
IES Institute of Education Sciences

Abstract

The five SEERNet digital learning platforms (DLPs) present unique opportunities for researchers by offering tools, processes, and infrastructure to make research more efficient, scalable, and relevant. However, conducting research within a DLP may require a shift in a researcher's orientation or mindset in how they think about potential research questions, study samples, available data, and planning study logistics. We encourage researchers to think flexibly about how they might adapt their ideas to the specific learning platform and consider iterating toward a more comprehensive understanding of their hypotheses. We illustrate some of the potential mindset shifts and offer key considerations at stages of the research process.

Acknowledgements

We would like to thank members of the SEERNet research teams and digital learning platforms for their contributions and review of this paper, as well as the Digital Promise and Empirical Education staff who provided valuable feedback and support.

Suggested Citation:

Schellinger, A., Zacamy, J., Roschelle, J., Closser, A., & Zepeda, C. D. (2024, April). *Considerations for Conducting Research in Digital Learning Platforms.* Digital Promise. https://doi.org/10.51388/20.500.12265/210

This project is supported by the Institute of Education Sciences, U.S. Department of Education through Grant R305N210034 to Digital Promise. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Table of Contents

Executive Summary

As a network, SEERNet connects platform developers, researchers, and practitioners to leverage existing digital learning platforms (DLPs) as research infrastructure. Conducting studies within existing learning platforms—where so much learning is happening—is an important advancement in education research, with the potential to reduce the chasm between research insights and widespread improvements in learning experiences. Within a DLP, researchers have the ability to study their research ideas in authentic educational environments, thus improving external validity and providing more opportunities for replication and scalability. The infrastructures are engineered to protect students' privacy, enable varying methodological approaches (e.g., student-level random assignment), and provide built-in efficiencies to conduct studies on large scales (e.g., data collection or built-in informed consent process). Importantly, the findings from these studies can influence broader applications and thereby achieve direct impacts on everyday teaching and learning.

In late 2023, SEERNet was thrilled to welcome the first IES funded research team to join the network. Drs. Avery Closser and David Purpura from Purdue University will explore whether exposure to perceptual cues in mathematics notation (e.g., using color to highlight key terms such as the inverse operators in an expression) may disrupt in-the-moment fluency and lead learners to pause and notice structural patterns, then ultimately practice more flexible and efficient problem solving. Joining SEERNet in early 2024, Drs. Cristina Zepeda and Kelley Durkin from Vanderbilt University will explore the effects of evidence-based metacognitive enhancements implemented during mathematics problem solving. They will examine impacts on middle school students' metacognitive skills, mathematics knowledge, and motivation. These studies will be conducted in two separate SEERNet DLPs. We are excited to learn from their insights to improve mathematics teaching and learning, as well as the DLP-as-research-infrastructure process.

We seek to encourage more studies, and these could emerge via IES funding, other research funding, and possibly with no funding at all. Yet we recognize that conducting a study within an existing digital learning platform may be different from traditional research settings. Thus, an orientation for researchers considering doing this type of work may be helpful, especially for those less familiar with DLPs. This paper will offer that orientation and illustrate some mindset shifts to prepare researchers for conducting research in DLPs. This paper is not intended to be a step-by-step guide or DLP-by-DLP walkthrough for researchers as they plan studies, but more of a thematic resource for researchers, and others in the field, about the differences in conducting research within a DLP. Through conversations and

informal interviews with members of the SEERNet research teams, and others who have successfully conducted research in a DLP, we have identified several considerations for researchers at the various stages of the research process. We elaborate on the following considerations:

1. Take the time to understand the learning opportunities and learning environment of the DLP and how the prospective research interests fit within those bounds.
2. Explore existing data on DLP users to inform decisions about research questions, design, and sample selection.
3. Collaborate with the DLP to understand the platform capabilities and intent to inform decisions about planned comparisons using experimental designs.
4. Carefully consider how to use data that is readily accessible within the DLP and weigh the tradeoffs and level of effort that may be required to access and integrate additional data, especially if that necessitates further data sharing agreements with educational institutions.
5. Leverage the robust process data on user behavior that DLPs often provide to address key analytic questions.
6. Be prepared that some of the study logistics will require careful attention and planning and that this may require continued interaction with the DLP.
7. Utilize the efficiencies of the DLPs and consider conducting a smaller sequence of studies to gain useful insights through incremental exploration of smaller, rapid-cycle changes.

Additionally, one overarching mindset shift has emerged. SEERNet encourages researchers to be prepared to think flexibly about how their research interests can be addressed through the DLPs. We caution that it may be hard to conduct a strongly preconceived study without a more thorough understanding of the platforms because they may have bounds on the type of data available, the comparisons that are supported during experiments, or the features that can be modified. We encourage researchers to see these platforms not as a malleable mechanism to research their own ideas but as a process that will require collaboration between the researcher and the DLP that could influence what hypotheses researchers focus on and how to (re)frame, implement, and test those hypotheses.

Introduction

The SEERNet movement to support research using digital learning platforms (DLPs) builds upon a rich tradition of learning engineering, data analytics, and related disciplines that have been doing this work for years. Collecting data is costly, and steps such as recruiting participants and building a technological infrastructure for an experiment can be time consuming. DLPs can standardize some of these steps and make research easier.

Precedents in the research literature include shared data sets from technology platforms, experiments conducted by external researchers in an available technology platform, and creation of scientific communities (such as the International Educational Data Mining Society, Society for Learning Analytics Research, and ACM Learning @ Scale) that address the availability of large-scale data. Building on these precedents, SEERNet and related programs like the AIMS Collaboratory (supported by the Gates Foundation) are transforming DLPs into research infrastructure and recruiting researchers to more widely use that infrastructure.

Infrastructure can make research "easier" but will not make research trivially "easy." Through the first several years of the SEERNet effort, we've seen that conducting research on an existing platform requires a mindset shift from the researcher, as well as an awareness of specific areas of challenge they will need to contend with. Through conversations and informal interviews with members of the SEERNet research teams, and others who have successfully conducted research in a DLP, we have identified several considerations for researchers throughout the research process. This paper is intended to help researchers anticipate what may be easy and what may still require new ways of thinking and working. Although the considerations may generalize to all sorts of educational platforms, here we tailor our suggestions to working with SEERNet DLPs as we build and support this research community.

Getting to know the learning environment of the DLP

Consideration #1: Take the time to understand the learning opportunities and learning environment of the DLP and how the prospective research interests fit within those bounds.

Researchers considering conducting research within a DLP should be prepared to spend time getting to know the utility and capabilities of the available platforms. While some researchers may be more experienced navigating an existing research infrastructure, working within a DLP will require researchers to conduct due diligence to understand *both* the DLP as the application or environment that delivers content and learning opportunities (within SEERNet, this is OpenStax, Mathia, ASSISTments, Canvas, and ASU) *and* the respective infrastructure that enables research on the platform (within SEERNet, this is Kinetic, UpGrade, E-TRIALS, Terracotta, and ASU's Learning@Scale). The SEERNet DLPs vary in the content covered, users targeted, and research capabilities. For example, ASSISTments is a K-12 mathematics tutoring system for teachers and students, and E-TRIALS enables research using ASSISTments content, whereas Terracotta is a research platform that facilitates experiments on learning activities within Canvas. SEERNet offers a Guide to Doing Research on DLPs as a central resource for researchers looking to understand more generally what each platform offers.

To make this potential research infrastructure more visible to researchers, one of SEERNet's partners, AIMS Collaboratory, is creating a searchable database that will catalog DLPs that are "researcher ready" (which could include DLPs within and beyond SEERNet). The catalog will include some of the key areas for researchers to understand when considering potential projects they could conduct with the platforms that are both abstract (e.g., potential research questions) and very specific (e.g., grade level, content area, and type of research supported). Particularly for researchers who have not worked within DLPs before, the time investment in this "information gathering" stage will be well worth it to determine alignment between their research interests and the learning opportunities and environment of the DLP. For example, in addition to basic information like the grade and content area covered, a researcher might inquire about the following:

- What is the learning environment in which the users interact with the content (e.g., classroom setting [in-person or virtual], at home/homework, or hybrid)?
- What is the platform's theory of action or application concept and content, if applicable?
- What does typical implementation/user interaction look like and how might this vary?
- What types of interactions can teachers and students have with each other within the platform?
- What kinds of media (e.g., video and audio) and interactions can be embedded?
- Do teachers develop and/or assign content, or is content embedded and determined by the user in the platform, or both?
- Is there a pre-existing content library or demo of what students experience (e.g., through different modules or units or using different application components) that researchers can review to determine if there is alignment with their research interests?

Within the SEERNet structure, researchers will need to be prepared to explore, modify, and study their ideas within learning environments that already exist. This may be a mindset shift for some learning scientists who are used to building new tools or working within a tool that they crafted from the ground up. When researchers craft their own platforms, making modifications to those tools or customizing content components may be easier and unconstrained. With SEERNet, researchers save the expense and effort of crafting a complete learning environment at high enough quality to deliver in real educational environments. But they may need to be more innovative or creative in conceptualizing how their research interests might fit within an existing environment. They may need to be flexible in how they leverage the capabilities that the SEERNet DLPs offer.

Planning stages: research questions, study design, and sample identification

Consideration #2: Explore existing data to inform decisions about research questions, design, and sample selection.

At the earliest planning stage, researchers may be contemplating the type of research questions they could answer using a DLP. The SEERNet DLPs support exploratory research using existing datasets and/or planned comparisons of learning approaches using an experimental design. One of the unique affordances of the DLP research infrastructure paradigm is scale of the user base and potential to access existing data. Several platforms offer access to data on large numbers of users, or even the full user population, in their typical education routines that can be used for secondary data analysis or to answer exploratory research questions. Other platforms may have available data from similar prior studies that can provide context on sample sizes, statistics on available dependent measures built into the platform, or other information needed for power calculations. Exploring these datasets can allow researchers to understand available data structures and user behavior that can be used to refine their hypotheses. Even if the intent is to conduct an experiment, exploring existing data can better relate the planned experimental contrast to what is happening in the platform now and what kinds of data can be obtained to study the contrast. The advantage of a DLP can be the depth of existing data about how students learn using that platform; the consideration for the research is taking time to explore this available data before fully specifying a research plan.

Consideration #3: Collaborate with the DLP to understand the platform capabilities and intent to inform decisions about planned comparisons using experimental designs.

Researchers who use a DLP gain the benefit of platform capabilities that support the student learning experience they will be studying. For example, DLPs can provide ways to present content to students and collect data from interactions with students as they learn; using these existing mechanisms can save researchers time and money.

Yet researchers need to consider how their interventions or new learning opportunities fit within the framework of existing content or structure of the learning platform. Researchers should be prepared to have ongoing conversations with the DLP about this topic and may need to refine their interventions/planned comparisons based on DLP capabilities and other limitations (e.g., what the DLP is willing to vary for their users).

Frequent collaboration between the researcher and the DLP team will likely need to occur throughout the study, and likely more intensely during these initial planning stages. For example, the researcher will need to understand if there are limits on the type of content that can be created or modified for experimental designs within the learning platform or if there are certain components of the platform content that cannot be modified or manipulated and what the reasons are for those limitations.

Researchers will also need to consider how their particular research questions and planned comparisons will interact with the pool of potential participants. DLPs may provide access to their large user base to serve as study participants if the planned comparison aligns well with existing platform content. However, studies with higher amounts of customization in the content may require that the researcher actively recruit participants or use a further selection/consent process from the larger user base to identify their analytic sample.

Exploring the available input, process, and outcome data

Consideration #4: Carefully consider how to use data that is readily accessible within the DLP and weigh the tradeoffs and level of effort that may be required to access and integrate additional data, especially if that necessitates further data sharing agreements with educational institutions.

Researchers using a DLP will find it easier to collect data via the platform. A key consideration, however, is what data is available.

Within the five SEERNet DLPs, there may be differences from conventional types of research in the kinds of data readily accessible to external researchers. For example, a DLP may not have access to individual student or teacher demographic data (that is, without connecting to other data sources or embedding a survey) but may collect school-level covariates, including locale type (rural, suburban, or urban settings) or Title I status, for the majority of their users. Some platforms may offer the ability to collect new data or otherwise integrate existing demographic or contextual data from other sources. This may be more likely in platforms that serve adult learners who have the ability to agree to participate in research activities and provide additional data through a consent process and survey embedded in the platform. Even if this is not an existing feature of a DLP, a consent process may be accomplished more efficiently through an adaptation within the DLP, as compared to a separate consent process outside of the DLP.

If some types of data of interest are not automatically collected within the DLP and cannot be easily embedded, it is important to understand if it is due to current technical constraints that could be relaxed in the future or if it is so by design. Conversations with the DLPs will clarify additional options. For example, some platforms offer opportunities to connect researchers with participating districts and schools and will help coordinate data privacy processes (e.g., de-identifying demographic data provided by the district and schools such that the researcher never sees personally identifiable information). Some platforms may populate student rosters from existing learning management systems (LMS) or student information systems (SIS) in districts or institutions, which could afford possibilities for linking with other sources of data given appropriate permissions. A researcher might decide to start with limited data at first and later link to additional data sets to achieve greater specificity of findings. This could be something that a researcher iterates toward through multiple small-scale studies, building relationships with the educational institution along the way.

Consideration #5: Leverage the robust process data on user behavior that DLPs often provide to address key analytic questions.

Some DLPs have the ability to track student behavior, such as which tasks students begin and how long they work on them, as well as measures of their detailed progress on the tasks. Researchers working in conventional classroom settings (outside of education technology environments) may not be familiar with what can be done with these types of log data or may not have experience collecting or analyzing it. Working with the platform team to understand prior research that utilized this data can be helpful, including review of resources from prior studies or codebooks and data dictionaries. There can be subtle as well as obvious uses of this data. For example, in some DLPs the researcher is not provided with direct information of attrition or compliance of the study subjects, but following the activity data (e.g., time stamps of activities, such as reading a prompt or attempting to solve a problem) allows identifying students who stopped participating after some point in time or tended to skip certain types of materials.

Moreover, building a deeper understanding of the existing learning platform, specifically with an eye toward how users interact with the system to generate the exact type of log or process data that is of interest, can be crucial. This process data may serve as a measure of important intermediate outcomes, or mediators, that are often very difficult to collect in conventional research settings. DLPs may have built-in outcomes that measure performance in the platform but have varying abilities to create, substitute, or add additional instruments to measure other more distal outcomes. Researchers will need to consider how the specificity of the learning outcomes built into the platforms can be used to answer their key research questions and when it is necessary to collect other outcomes.

Tackling the logistics: What's worked within SEERNet

Consideration #6: Be prepared that some of the study logistics will require careful attention and planning and that this may require continued interaction with the DLP.

Researchers working within a DLP may need to shift how they view several of the usual logistical challenges that arise in education research. One key learning that has emerged from the SEERNet network is the benefit of an extended feasibility process—that is, a development phase involving the research team, the platform team, and study participants or other stakeholders to finalize the research questions, study design, implementation of the experimental contrasts, and data collection, and to address the logistics of each stage. The IES proposal process for SEERNet researchers included an exchange between research and platform teams in the form of a "feasibility letter." There is also an expectation from IES that the funded teams spend the first six months of the award building out and refining their designs and implementation, working closely with the DLP to work through the logistics and to ensure study feasibility. The research teams and DLPs have both benefited from this extended, planned time to collaborate.

Discussions with current SEERNet researchers and advisory teams have also highlighted the importance of understanding and navigating the IRB process. Although platforms may have their own IRB agreements to collect or share data in certain instances, external researchers should be prepared to work with their own IRB institutions and communicate their study designs in appropriate ways to review boards. Reviewing protocols for studies being conducted in a DLP may be new for some institutional IRB offices, which may cause them to be more conservative in their review process. Researchers should emphasize in their IRB applications, where appropriate, that the studies are comparisons of normal education practices, which may ease the review and approval process. Drawing upon language and examples from the DLP about how the innovation aligns with their current "educational practices" may be beneficial to include in the IRB review process.

Researchers should also consider the logistics of accessing study data. With multiple parties involved in implementation and data collection, there may be differences in the timeline or steps needed to access data. For studies that involve data collection on an ongoing basis—such as from participants who experience the content at different times of the school year—researchers and platform developers may have different procedures for monitoring implementation and data collection. Several of the SEERNet DLPs are building data dashboards that will allow researchers to monitor participation and implementation along the way, as well as include data visualizations of key outcomes. Clear communication throughout the data collection process will manage expectations around the structure, availability, management, and sharing of data.

As the network continues to grow, the lessons learned from these discussions can be shared across platforms to increase efficiency and simplify the feasibility and onboarding processes and other key logistics. Researchers who attempt studies with few precedents may need to spend more time in communication with the platform team or participants/implementation partners to specify and refine their processes. The network will also learn more about the process and options related to analysis and reporting as the funded research teams get to that stage. One of the broad goals of SEERNet as a network is to build tools to increase the momentum and efficiency of procedures to tackle these logistical considerations across the platforms and research teams.

Iterating toward improvements

Consideration #7: Utilize the efficiencies of the DLPs and consider conducting a smaller sequence of studies to gain useful insights through incremental exploration of smaller, rapid-cycle changes.

Another important *mindset* shift in conducting research within a DLP is for researchers to consider how they could do a sequence of smaller studies that build upon one another instead of planning from the beginning for a large confirmatory study. Again, this may start with using existing data available by the DLP to answer the research questions (in part) and/or inform the design of a new study.

One benefit of a more iterative approach is that it will allow researchers to build in improvements in study design, which can involve things like adding outcome measures or collecting additional demographic data. For example, an initial study may use a built-in outcome measure (e.g., time on task, student performance on next problem set, or completion of assignments) to see if results are trending in the expected direction. If those trends are positive, a researcher can consider what would be added by including a more specific outcome measure to fine-tune the dependent variable to fit within the logic of the study. Likewise, if it is outside of the current scope of the DLP to collect and integrate demographic data on a large scale, a researcher could start with a pilot study with a smaller sample to collect/integrate important demographic variables. This would allow the researcher to examine, preliminarily, if there are differences by important subgroups and plan to modify or improve the intervention prior to deployment in a larger trial. Researchers who have used DLPs for experiments have acknowledged that once an intervention (or content modification) is built into the DLP, making smaller tweaks toward improvement, and then deploying those improvements to a new sample, is a relatively efficient process.

Starting with a smaller sequence of studies may also allow researchers to confirm the planned comparisons (e.g., language modifications) are appropriate for the study sample, that the demands of the content are developmentally appropriate, and that the intervention or comparisons are received by students as expected. This would be especially beneficial to pilot tests if, for example, the intervention had been previously tested in a lab or classroom setting outside of a DLP and the transfer into a DLP was sufficiently different. The smaller sequence of studies would also allow the DLP to test for any technical issues (on the platform end) before being tested with a larger sample.

Discussion: Highlighting the benefits of research within a DLP

While we have established several key considerations and possible mindset shifts needed to conduct research with DLPs, the benefits of this work are numerous. For example, research can become easier to do, better grounded in the details of learning processes, and more applicable to real-world settings. A central goal of SEERNet is to grow the community of platform developers, researchers, practitioners, and funders, and through these efforts we continue to see the excitement from the field in building infrastructures to allow research to occur within DLPs. A prior paper articulating the network's visions describes four advances in educational research that are emerging from efforts of SEERNet participants—*better community, better practitioner engagement, better engineering, and better science*—and we build upon those ideas here.

One benefit we are particularly excited to see, and that early efforts within the network illustrate, is that the opportunity to conduct research using a DLP may be especially well suited to early career researchers, thus providing opportunities for more diverse and innovative ideas to be explored. Those with training in experimental design can readily take advantage of the efficient recruitment and data collection processes within a DLP to implement studies that may be costlier and more time consuming to do in traditional classroom settings, particularly absent the resources and relationships of more experienced researchers. Likewise, a willingness to adopt a flexible approach in working with platform developers can be essential to achieving research goals in this paradigm. The flexibility may come with more enthusiasm from an early career researcher who is in the early stages of pursuing their research interests. Moreover, harnessing DLPs to more quickly produce impactful research in pursuit of career growth supports the practical concerns of these researchers. We also encourage more senior researchers with well-established research agendas to consider working with early career researchers to advance aspects of that agenda with the DLP research infrastructure. This dynamic can support a *better community* of researchers.

Another key benefit is that studies within the DLPs will contribute content, product, or process improvements to existing platforms that are already in schools and have a large user base. The key learnings from these studies can support immediate product improvements and contribute to materials that reach more students. The results are thus more likely to scale from research to practice and have the potential to greatly improve *practitioner engagement* in the research process. Some of the platforms are open source, further creating an opportunity to share the benefits with a larger community beyond the study, and efforts across all platforms to increase the generalizability of study results is central to SEERNet and its goal of *better science*.

For studies that take advantage of large numbers of users for secondary data analysis or for planned comparisons of interventions that fit within existing learning platform parameters, the advantage of DLP research infrastructure is immense and cannot be overstated, potentially delivering hundreds of thousands of users. Facilitating access to larger, more detailed data sets for more researchers will further SEERNet's mission of creating *better engineering*. Even for customized comparison studies that utilize more traditional recruitment methods for smaller samples, randomizing students or classes within their natural learning environments affords a scale and scope not often available in lab-based or classroom-based (outside of an edtech environment) research. Exposing a larger pool of classically trained researchers to the types of log data that are generated in DLPs can have benefits for the types of questions they seek to answer when using the systems and further understanding for the field of how these outcomes can be measured and manipulated.

As more research teams join the network, SEERNet will continue to gather, synthesize, and disseminate information about the benefits and real applications of DLP research studies to researchers, practitioners, and funders to further develop and grow this community.

For researchers interested in learning more about the SEERNet DLPs or opportunities for research, please visit www.seernet.org.