Secure Education and Learning Research at Scale with OpenStax Kinetic

Debshila Basu Mallick†
OpenStax, Rice University
Houston TX USA
debshila@rice.edu

Brittany C. Bradford
OpenStax, Rice University
Houston TX USA
bb25@rice.edu

Richard Baraniuk
Department of Electrical and Computer Engineering,
Rice University
Houston TX USA
richb@rice.edu

ABSTRACT

OpenStax Kinetic is an innovative research infrastructure that aims to transform education and learning research in the digital age. With its access to large sample sizes, authentic learning environments, experimental control, scalability, security and privacy protection, Kinetic provides an unparalleled opportunity for researchers to study the complex interactions between different factors in digital learning environments. This versatile platform utilizes Qualtrics and can support various research designs, including correlational, longitudinal, and interventional studies. Kinetic's unique privacy-by-design implementation via secure enclaves ensures that researchers can analyze fully-identified data without compromising data security and privacy as well as affords greater analytical reproducibility. The findings from Kinetic can inform educational interventions and strategies to enhance student success in digital learning environments. Kinetic has the potential to significantly advance education and learning research by improving pedagogies, practices, and policies in education and learning sciences. In this demo of Kinetic, researchers will be able to interact with the test instance of the Kinetic system online and view the learner experience. All researchers will be able to engage in the experience of creating a study, releasing a study, and interacting with our implementation of secure enclaves for data analysis.

CCS CONCEPTS

• Applied Computing → Education → Interactive Learning Environments

KEYWORDS: Education R&D, Digital learning platform, Digital learning research, Research Methods at Scale, Adult Learning

†Corresponding author

L@S '23, July 20–22, 2023, Copenhagen, Denmark

ACM ISBN 979-8-4007-0025-5/23/07...$15.00
https://doi.org/10.1145/3573051.3596187

ACM Reference format:

Debshila Basu Mallick, Brittany C. Bradford, and Richard Baraniuk. 2023. Secure Education and Learning Research at Scale with OpenStax Kinetic. In *Proceedings of the Tenth ACM Conference on Learning @ Scale (L@S '23), July 20–22, 2023, Copenhagen, Denmark.* ACM, New York, NY, USA. 3 pages. https://doi.org/10.1145/3573051.3596187

1 INTRODUCTION

Digital learning, including online learning, has become a key part of education, yet much remains unknown about how to facilitate student success in these learning environments. Prior learning science research has revealed that how an individual student learns and succeeds in particular learning environments (e.g., digital or in-classroom) depends on many diverse factors and their interactions [1], [2], yet the nature of these interactions and consequent impact remain underexplored. Among these factors, students' individual difference factors have been studied extensively [3], [4]. Other important factors that impact student success include contextual factors such as instructional design [5]–[7], learning strategies employed [8] and various affordances enabled by the rapid development of digital learning technologies [9]. Here we demonstrate OpenStax Kinetic, a research infrastructure integrated with the OpenStax platform, that creates an opportunity to investigate what works for whom and in which context [1] while simultaneously meeting seven key requirements for innovation in learning research. Kinetic provides education and learning researchers access to large sample sizes, multi-dimensional data, authentic learning environments, experimental control, scalability, security and privacy-protection, integrated fairness, ethics, accountability, and transparency (FEAT), and scale, fostering their ability to operate most efficiently and most deeply investigate key issues of education and learning sciences. At the intersection of edtech, textbook publishing, and learning sciences research, OpenStax Kinetic is uniquely positioned to impact this line of investigation as one of the largest publishers of openly-licensed high-quality free textbooks in the world with a growing library of 60+ textbooks serving over 6.5 million learners annually since 2021.

Researchers on Kinetic can deploy a wide range of learning research studies at scale by fully leveraging OpenStax's large higher-ed adult learnerbase within the US in an authentic digital learning environment. Kinetic is intended to enable a rapid research-and-deployment cycle, creating a golden opportunity to significantly advance education and learning research by studying the complex interactions among different factors in digital learning environments, broaden the reach of learning science research findings through their integration into OpenStax products to benefit millions of students, accommodate a rapid iteration and testing ("fast fail") research and integration approach [10].

2 RESEARCH CAPABILITIES OF OPENSTAX KINETIC

Kinetic utilizes Qualtrics, a tool commonly used in the education research space thereby lowering the barrier of entry for researchers. Built in an agile manner, later versions of Kinetic will be able to accommodate progressively more sophisticated tasks and study designs than those currently available (e.g., posthoc analyses, A/B/N tests, and Randomized Control Trials). Researchers are being recruited in small numbers as Kinetic undergoes development and will be released to education and learning researchers across the US over time.

Launched in early 2022, Kinetic has since accrued responses from over 7,000 learners across 36 activities created by researchers from within OpenStax and the Department of Psychological Sciences at Rice University. These research activities range from self-report survey measures of learner characteristics that impact academic and learning outcomes, all the way to interventional research studies. These learner characteristics are available for all Kinetic researchers to use within their own research, to diagnose inequities in educational opportunities and outcomes [11], and generating insights into how to achieve equitable learner success that ultimately have the potential to be translated into improved pedagogies, practices, and policies.

3 PRIVACY BY DESIGN ON OPENSTAX KINETIC

OpenStax has huge volumes of student data to share with researchers, but significant barriers exist. Most notably, it is technically difficult to export massive-scale data, and even small exports increase the risk of exposing sensitive student data. Typical data-sharing efforts address these issues by reducing the size of the shared data and/or passing it through some kind of de-identification process. These approaches attempt to strike an impossible balance between removing identifying information and leaving enough information with which to prove research hypotheses, perform interventions that work well for diverse students with different prior experiences, and validate that interventions are fair and equitable. Critical student and contextual factor relationships may be lost in data that is scrubbed prior to export.

To address these barriers, a unique feature of Kinetic is its privacy-by-design implementation via secure enclaves. These enclaves enable researchers to use their preferred analytic toolkit starting with the R statistical suite of tools to analyze fully-identified data without direct access to it and only return aggregate information. This approach to secure large-scale data analysis contrasts to opening the data to researchers [12]. The benefit of enclaves are that they mitigate any risk for data re-identification without compromising on the analytic value of the data by masking crucial student information. In the secure data enclave approach, researchers generally do not see real data. Instead, they see documentation on the data in the enclave, as well as training and test data with which to build and test their research software. Building on prior work with Massive Online Open Course (MOOC) Replication Framework (MORF) [13], the research software lives in containers for portability and flexibility, and members provide example containers that researchers can use as starting points. Containers are like little computers saved in a file that can be run on any other computer. Researchers pass their analysis code within containers to the enclave where Kinetic-provided infrastructure runs the code with access to fully-identified data and passes back the knowledge it generates, avoiding the need for the researcher to ever see the fully-identified data. Moreover, an affordance of the enclaves is that they enable analytic reproducibility that can help address some of the reproducibility crisis [14] in learning and educational sciences.

4. CONCLUSION

OpenStax Kinetic is an ambitious new research infrastructure that has the potential to transform research by connecting researchers to large populations of diverse learners distributed across the United States. It has built in privacy by design that fosters secure ways to account for learner characteristics as well as diagnose and address issues of inequity in educational opportunities and outcomes. Additionally, its use of secure enclaves for data analysis effectively supports researchers in maintaining analytic reproducibility of research.

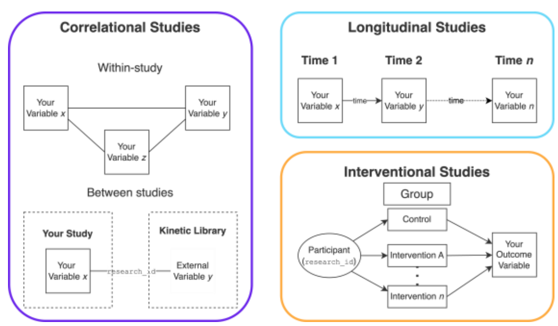

Figure 1: OpenStax Kinetic is able to support the correlational, longitudinal, interventional, and combinations of these research designs with its current capabilities. The features are being incrementally expanded and iteratively refined. Anonymized `research_ids` connect individuals across different studies.

ACKNOWLEDGMENTS

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant

R305N210064 to Rice University, NSF Award 1842378, as well as philanthropic foundations such as Schmidt Futures, Valhalla Foundation. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

REFERENCES

[1] R. F. Kizilcec *et al.*, "Scaling up behavioral science interventions in online education," *PNAS*, vol. 117, no. 26, pp. 14900–14905, Jun. 2020, doi: 10/gg2c6t.

[2] D. McNamara *et al.*, "Situating AI (and Big Data) in the Learning Sciences: Moving toward Large-Scale Learning Sciences," in *Artificial Intelligence in STEM Education*, CRC Press, 2022, pp. 289–308.

[3] M. E. Beier and P. L. Ackerman, "Age, Ability, and the Role of Prior Knowledge on the Acquisition of New Domain Knowledge: Promising Results in a Real-World Learning Environment.," *Psychology and Aging*, vol. 20, no. 2, pp. 341–355, Jun. 2005, doi: 10.1037/0882-7974.20.2.341.

[4] X. Wang, "Why Students Choose STEM Majors: Motivation, High School Learning, and Postsecondary Context of Support," *American Educational Research Journal*, vol. 50, no. 5, pp. 1081–1121, Oct. 2013, doi: 10.3102/0002831213488622.

[5] R. E. Mayer, "Thirty years of research on online learning," *Appl Cognit Psychol*, vol. 33, no. 2, pp. 152–159, Mar. 2019, doi: 10.1002/acp.3482.

[6] R. E. Mayer, "Applying the Science of Learning : Evidence- Based Principles for the Design of Multimedia Instruction," no. November, 2008.

[7] R. E. Mayer, "Advances in applying the science of learning to education: An historical perspective," *Journal of Applied Research in Memory and Cognition*, vol. 1, no. 4, pp. 249–250, 2012, doi: 10.1016/j.jarmac.2012.10.001.

[8] J. Dunlosky, K. A. Rawson, E. J. Marsh, M. J. Nathan, and D. T. Willingham, "Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology," *Psychological Science in the Public Interest*, vol. 14, no. 1, pp. 4–58, 2013, doi: 10/nbw.

[9] M. Aparicio, F. Bacao, and T. Oliveira, "An e-Learning Theoretical Framework," 2023.

[10] "Equity," *STANDARDS FOR EXCELLENCE IN EDUCATION RESEARCH*, Jun. 16, 2022. https://ies.ed.gov/seer/equi[1] R. F. Kizilcec *et al.*, "Scaling up behavioral science interventions in online education," *PNAS*, vol. 117, no. 26, pp. 14900–14905, Jun. 2020, doi: 10/gg2c6t.

[2] D. McNamara *et al.*, "Situating AI (and Big Data) in the Learning Sciences: Moving toward Large-Scale Learning Sciences," in *Artificial Intelligence in STEM Education*, CRC Press, 2022, pp. 289–308.

[3] M. E. Beier and P. L. Ackerman, "Age, Ability, and the Role of Prior Knowledge on the Acquisition of New Domain Knowledge: Promising Results in a Real-World Learning Environment.," *Psychology and Aging*, vol. 20, no. 2, pp. 341–355, Jun. 2005, doi: 10.1037/0882-7974.20.2.341.

[4] X. Wang, "Why Students Choose STEM Majors: Motivation, High School Learning, and Postsecondary Context of Support," *American Educational Research Journal*, vol. 50, no. 5, pp. 1081–1121, Oct. 2013, doi: 10.3102/0002831213488622.

[5] R. E. Mayer, "Thirty years of research on online learning," *Appl Cognit Psychol*, vol. 33, no. 2, pp. 152–159, Mar. 2019, doi: 10.1002/acp.3482.

[6] R. E. Mayer, "Applying the Science of Learning : Evidence- Based Principles for the Design of Multimedia Instruction," no. November, 2008.

[7] R. E. Mayer, "Advances in applying the science of learning to education: An historical perspective," *Journal of Applied Research in Memory and Cognition*, vol. 1, no. 4, pp. 249–250, 2012, doi: 10.1016/j.jarmac.2012.10.001.

[8] J. Dunlosky, K. A. Rawson, E. J. Marsh, M. J. Nathan, and D. T. Willingham, "Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology," *Psychological Science in the Public Interest*, vol. 14, no. 1, pp. 4–58, 2013, doi: 10/nbw.

[9] M. Aparicio, F. Bacao, and T. Oliveira, "An e-Learning Theoretical Framework," 2023.

[10] M. Schneider and K. Garg, "Medical Researchers Find Cures by Conducting Many Studies and Failing Fast. We Need to Do the Same for Education," *The 74 Million*, Feb. 10, 2020. https://www.the74million.org/article/schneider-garg-medical-researchers-find-cures-by-conducting-many-studies-and-failing-fast-we-need-to-do-the-same-for-education/

[11] "Equity," *STANDARDS FOR EXCELLENCE IN EDUCATION RESEARCH*, Jun. 16, 2022. https://ies.ed.gov/seer/equity.asp

[12] "Make findings, methods, and data open," *STANDARDS FOR EXCELLENCE IN EDUCATION RESEARCH*, Jun. 10, 2021. https://ies.ed.gov/seer/open_data.asp

[13] J. Gardner, C. Brooks, J. M. Andres, and R. S. Baker, "MORF: A Framework for Predictive Modeling and Replication At Scale With Privacy-Restricted MOOC Data," in *2018 IEEE International Conference on Big Data (Big Data)*, Seattle, WA, USA: IEEE, Dec. 2018, pp. 3235–3244. doi: 10.1109/BigData.2018.8621874.

[14] R. D. Peng and S. C. Hicks, "Reproducible Research: A Retrospective," *Annu. Rev. Public Health*, vol. 42, no. 1, pp. 79–93, Apr. 2021, doi: 10.1146/annurev-publhealth-012420-105110.