Incorporating REDCap for Data Collection of K-12 CS Education Research

Monica M. McGill
CSEdResearch.org
Knox College
monica@csedresearch.org

ABSTRACT

Over the course of the last two years, the resource center for the computer science educational community at csedresearch.org has been collecting evaluation instruments for measuring the impact of CS education primarily in primary and secondary schools. The CSEdResearch.org Resource Center provides over 100 surveys and protocols for use in researching and evaluating computer science education. Work has begun to expand the capability of the site so that researchers and evaluators can distribute the surveys directly on the site. This enhancement is a packaged plug-in called Research Electronic Data Capture (REDCap), a product of Vanderbilt University built under support from the National Institutes for Health (NIH) [2][3]. In this extended abstract, we present a summary of REDCap, the benefits it can provide to our community, and how we intend to vet the tool before releasing it to the public.

Keywords

Educational research, resources, evaluation, assessment, constructs, K-12, primary, secondary, REDCap, data collection, open science

1. INTRODUCTION

The CSEdResearch.org Resource Center is designed to support the computer science (CS) education community, with a particular focus on K-12 researchers, evaluators, teachers, administrators, curriculum designers, and policymakers [8, 9]. The site currently houses data curated from over 600 primary and secondary CS K-12 education articles across ten venues(2012-2019) and currently houses over over 110 computing education evaluation instruments and protocols.

We are currently in the process of vetting a data collection tool, the Resource Electronic Data Capture (REDCap) tool, for integration [2][3]. REDCap (`https://projectredcap.org`) is a well-established data collection tool that was originally built for clinical researchers to collect data that met HIPAA guidelines and that could be used in single- or multisite research projects [3]. Since 2004, the project has grown to be used across more than 4,000 institutions with over 1 million users, having evolved from its original clinical roots into a general data collection platform. Institutions that use REDCap include Harvard affiliated institutions (e.g., Boston Children's Hospital, Harvard T.H. Chan School of Public Health, and Joslin Diabetes Center) [4], Weill Cornell Medicine Clinical and Translational Science Center [14], the University of Minnesota Clinical and Translational Science Center [12], and even the prestigious Mayo Clinic [6]. Closer to education and social sciences, REDCap is used in institutions such as the PennState Social Science Research Institute [11], Deakin University [1], and University of Nebraska-Lincoln Methodology and Evaluation Research Core [13].

Key features of the tool includes the capability to use existing surveys, easily managed contact lists, scheduling of surveys, and creating multiple arms of surveys in longitudinal studies. Users can download their data sets to Excel, SAS, Stata, R, or SPSS for data analysis. It can also means important data security and privacy protocols, such as removing identifiers from datasets prior to exporting and logging and audit trails on all data interactions.

2. DATA COLLECTION USING REDCAP

REDCap is maintained and supported by consortium partners' local hosts. Prior to releasing REDCap, three educational researchers created an evaluation plan that consists of four primary categories: Usability, Functionality, Data Integrity, and Security and Privacy. We are in the process of evaluating the tool against the criteria (summer 2020). Initially, three undergraduate students and the primary researcher will use the evaluation plan to vet the tool. To perform this evaluation, each tester will have a copy of the testing template to complete. We anticipate that going through each of the criteria in the evaluation plan will take about 40-60 hours for each team member to properly learn the tool, compare it to at least one other tool (e.g., Survey Monkey, Qualtrics), and gain maximum understanding of its capabilities. For each item to be evaluated, each team member will rate the tool on a scale of Pass (Excellent), Pass (Very Good), Pass (Good), Pass (Fair), and Fail. Each team member will also provide notes about their examination process and how they made a decision for each item.

Once the team has completed their evaluation, we will review the results for a better understanding of the tool's pros

and cons. We will then decide if REDCap should be abandoned as potential platform or if alpha testing should begin. For alpha testing, we will identify 2-4 users in the computer science education research community to test the tool using real surveys. Alpha testers will be interviewed to ensure that their needs are met. We will then review the feedback from the alpha testers and, in conjunction with them, will again decide if REDCap should be abandoned as a potential platform or if beta testing should begin.

For beta testing, we will recruit 15-20 potential users. Similar to alpha, we will train the beta users and ask them to use the tool for surveying for at least one project. We will ask them for their feedback to determine if this tool would provide the necessary support for their data collection needs in their projects. If beta testing is successful, then we will release REDCap to the public.

3. DATASET

The data for REDCap is stored in a MySQL database that is self-contained across 132 tables, including permissions, folder and project structures, and other tool settings. For individual users, data can be exported for analysis in Excel, R, SPSS, Python, STATA, XML, and SAS.

Discovering best practices in computer science education cannot be achieved by one researcher alone. Large-scale collaborative science can provide a way to "...understand which instructional practices work for whom and under what condition...." [5, abstract]. To conduct large-scale collaboration across computer science education requires tools to gather this data in one place or provide another means of integrating the data for meta-analysis [7]. Meta-analysis can then be performed across projects with heterogeneous properties selected by the user. Building this data lends itself to the empirical evaluation of heterogeneous research studies through quantitative meta-analysis [7, 10]. Additionally, an open science approach to data makes it possible for data analysis to be reproduced easily and systematically.

This type of data collection and analysis is already happening within tools designed to teach computer science and computational thinking. By extending this practice to include multiple studies across multiple sites, we can start to synthesize the data and potentially accelerate the process of determining promising practices for various learners.

4. CONCLUSION

With our work to build a platform for the community to use as a central data collection site, the power of the underlying data becomes open for further analysis. As the tool's usage grows, there becomes the potential for comparing what may work better for some groups of learners than for others, providing information to feedback to researchers and practitioners.

5. ACKNOWLEDGEMENTS

This material is based upon work supported by the U.S. National Science Foundation under Grant Nos. 1745199 and 1757402.

6. REFERENCES

[1] Deakin University Faculty of Arts and Education. Staff news. https://blogs.deakin.edu.au/artsed/2020/04/28/online-sessions-for-redcap-and-qualtrics-on-24th-april-and-4th-may/, 2020.

[2] P. A. Harris, R. Taylor, B. L. Minor, V. Elliott, M. Fernandez, L. O'Neal, L. McLeod, G. Delacqua, F. Delacqua, J. Kirby, et al. The redcap consortium: Building an international community of software platform partners. *Journal of biomedical informatics*, 95:103208, 2019.

[3] P. A. Harris, R. Taylor, R. Thielke, J. Payne, N. Gonzalez, J. G. Conde, et al. A metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*, 42(2):377–81, 2009.

[4] Harvard Catalyst. Redcap. https://catalyst.harvard.edu/services/redcap/, 2020.

[5] F. G. Hillary and J. D. Medaglia. What the replication crisis means for intervention science, 2019.

[6] Mayo Clinic Center for Clinical and Translational Science (CCATS). Workforce development redcap software. https://www.mayo.edu/research/centers-programs/center-clinical-translational-science/education/workforce-development/classroom-courses/redcap-software, 2020.

[7] M. M. McGill. Discovering empirically-based best practices in computing education through replication, reproducibility, and meta-analysis studies. In *Proceedings of the 19th Koli Calling International Conference on Computing Education Research*, pages 1–5, 2019.

[8] M. M. McGill and A. Decker. csedresearch.org, 2017.

[9] M. M. McGill and A. Decker. Defining requirements for a repository to meet the needs of k-12 computer science educators, researchers, and evaluators. In *2018 IEEE Frontiers in Education Conference (FIE)*, pages 1–9. IEEE, 2018.

[10] M. M. McGill, T. McKlin, and E. Kaylor. Defining what empirically works best: Dynamic generation of meta-analysis for computer science education. In *Proceedings of the 2019 ACM Conference on International Computing Education Research*, pages 199–207, 2019.

[11] PennState Social Science Research Institute. Electronic management of data and regulatory documentation. https://ssri.psu.edu/electronic-management-data-and-regulatory-documentation, 2020.

[12] University of Minnesota Clinical & Translational Science Center. Redcap. https://www.ctsi.umn.edu/researcher-resources/tools-and-software/redcap, 2020.

[13] University of Nebraska-Lincoln Methodology and Evaluation Research Core. Redcap. https://merc.unl.edu/redcap, 2020.

[14] Weill Cornell Medicine Clinical & Translational Science Center. Redcap. https://ctscweb.weill.cornell.edu/research-resources/clinical-translational-research-informatics/redcap, 2020.