Defining Requirements for a Repository to Meet the Needs of K-12 Computer Science Educators, Researchers, and Evaluators

Monica M. McGill
Department of Computer Science
Knox College
Galesburg, IL 61401, USA
mmmcgill@knox.edu

Adrienne Decker
School of Interactive Games and Media
Rochester Institute of Technology
Rochester, NY 14623, USA
adrienne.decker@rit.edu

***Abstract*—This Research Full Paper describes the early research conducted to design a searchable repository of peer-reviewed research related to pre-college computing activities. This repository is part of a larger project to enable the computer science education community to gather and analyze data related to the effectiveness of these activities. To ensure that the repository met the needs of the community, we convened a virtual focus group of experienced and expert researchers and educators to discuss the repository's value, what it should contain, and how it should be presented. This paper presents 1) an analysis of these discussions, which shows that participants were equally interested in the repository's content and quality, 2) an initial list of variables that can affect the outcomes of these activities, and 3) an initial set of questions researchers should ask when authoring computer science education research involving pre-college computing activities. We also consider these results in light of the larger challenge raised by others of how to improve quality research in computing education.**

***Keywords*–education, research, K-12, primary, secondary, impact, pre-college computing activities**

I. INTRODUCTION

There is a large body of evidence that the computing community values and invests time and resources in computing interventions that occur before students graduate from high school [1, 2, 3, 4, 5, 6, 7]. However, there is very little research investigating the long-term impact of pre-college computing activities. In a 2016 U.S. centric study, only 80 of 3,949 recent ACM and IEEE computing education journal articles and conference proceedings discussed the results of pre-college computing activities, and only 7 of those reported longitudinal data [8]. In an expanded literature review that included non-US journals and conferences (7,265 articles), only 9 of the 98 K-12 computing activity articles reported any longitudinal data [9]. Further complicating matters is that there is not a tradition or set of expectations regarding the style, structure, and reporting of the results of these types of interventions, with many using weakly designed studies and anecdotal evidence [10, 11, 12, 13, 14, 15]. This makes it challenging, if not impossible, to have an empirically-based understanding of the research results [12].

Craig (2016) further defines the need to state clear objectives in research, including the specific issues being addressed, target demographic, and circumstances surrounding the activity in order to evaluate its effectiveness. Rigorous research will help the community identify which activities "...should be replicated and which should be abandoned, soundly based on a theory of the problem" [16, p. 588].

Though there are models for exemplary research investigating pre-college computing activities, there is a lack of quality research that investigates such activities as a whole. With this in mind, we developed a website, *https://csedresearch.org* [17], to provide researchers, practitioners, and evaluators with a large-scale repository for published research, evaluation instruments, and research processes.

The value of such a repository has been previously established. Sanders, et al., note that a searchable repository has potential to enable growth through valuable data sharing throughout the community [18], and the U.S. National Science Board notes that digital data collections "enable analysis at unprecedented levels of accuracy and sophistication and provide novel insights through innovative information integration" [19, p. 9] . Other repositories have been established and are actively used throughout the computing education community, such as the National Center for Women & Information Technology's (NCWIT) EngageCSEdu (*https://www.engage-csedu.org/*), Project Quantum (*https://diagnosticquestions.com/Quantum*) from United Kingdom's Computing at School (CAS) organization, and ICPSR from University of Michigan (*https://www.icpsr.umich.edu/icpsrweb/*).

With the evolution of data mining tools, such a repository could enable researchers to discover patterns, relationships, and correlations to improve our understanding of computer science education. To build a repository, we needed to first design it. We defined our objectives to do this as:

- Identify the content to be included in the repository,
- Establish requirements for the design of the repository to effectively present and search for its contents, and
- Identify the type of data analysis that would be most beneficial for the community in determining best practices for teaching to various demographic groups.

As a result, the overarching research question for this part of our study was:

> *What type of data might be usable and useful for educational researchers, evaluators, and practitioners when measuring the impact of pre-college computing activities?*

To answer this question, we convened a representative virtual focus group at the earliest stages of our design process. The group was comprised of experts in the fields of computer science education research and evaluation. We present the findings of this qualitative study here, including feedback related to the usability of the site where the repository will be housed, the general value of the repository, and the challenges to be mitigated as we continue its development.

We also report on an equally significant finding–the evolution of a set of questions for computer science education researchers evaluating pre-college computing activities. These questions are not necessarily limited to pre-college activities and can serve as a handy checklist for those authoring or peer-reviewing studies.

It is a non-trivial to create such repositories and to make them useful to the wider community [18]. The remainder of this paper includes the following sections: a brief summary of the development of the *https://csedresearch.org* website, the methodology used in the forming and convening of the virtual focus group, the results of the virtual focus group (including the set of questions for researchers and reviewers), and a conclusion section.

II. *https://csedresearch.org*

To add context to this study, we provide here a brief time line of our work on this project to date that has culminated in the creation of *https://csedresearch.org*:

- Dec 2016-Feb 2017: Conducted virtual focus group (described in this paper)
- February 2017-May 2017: Developed design concept of the website/repository based on previous research and the results of the virtual focus group
- May 2017: Conducted concept testing with eight potential users (K-12 CS Educators, CS Education Researchers, K-12 CS Program Evaluators)
- June-July 2017: Developed alpha version of the website/repository
- July 2017: Conducted alpha testing with 15 potential users (K-12 CS Educators, CS Education Researchers, K-12 CS Program Evaluators)
- August-December 2017: Developed beta version of the website/repository
- January 2018: Conducted open beta testing within the computer science education researcher community with over 150 respondents
- February-August 2018: Revise the website/repository based on feedback and release the final version website/repository
- Beyond Fall 2018: Continue to populate the repository with data

The study described in this paper focuses on the early research in which we conducted the virtual focus group to explore the three objectives described in the introduction. We engaged six experienced researchers, educators and evaluators in valuable discussion to help shape the foundation of the project design.

III. Methodology

To explore our research question, we first formally defined three objectives:

1) Determine what content should be included in the repository
2) Determine how to enable effective presentation of and search mechanisms for its content, and
3) Determine what data analysis and results would be most beneficial for the community the community in order to identify best practices for teaching to various demographic groups.

We then conducted a qualitative study through a virtual focus group of potential users within the computing education research community and report this study using Plakhotnik and Rocco's (2009) framework [11, 20]. This framework is similar to the framework used for the U.S. Institute of Education Sciences' *What Works Clearinghouse* but with a wider scope [21].

As computing educators and education researchers, we have been conducting computing education research for approximately 15 years each, with experience in qualitative, quantitative, and mixed methods studies. The first author served as the primary moderator and has had past experience conducting research with and moderating focus groups. Both of us have analyzed and coded qualitative data using grounded theory research. In addition, an external researcher well-versed in qualitative studies and focus groups reviewed our methodology and data analysis techniques and provided constructive feedback during the study design, the analysis of the results, and the final derivation of this paper.

A. Participants

The study was developed using methods from Liamputtong, Patton, and Onwuegbuzie, et al, and was reviewed and approved by the Institutional Review Boards at both of our institutions [11, 22, 23]. The virtual focus group had six participants, all educational researchers involved in computing education. Participants were recruited based on their intermediate and advanced experiences within the educational research community. Four were computing education researchers employed as faculty at geographically diverse locations in the U.S., with two at public post-secondary institutions, one at a private post-secondary institution, and one at a private, for-profit company. At least one has designed and implemented middle-school computing outreach activities. The remaining two participants were evaluators, both at non-profit institutions in the U.S. Each participant received a $500 stipend for full participation in the study.

Participants were provided with a letter of consent and a description of the study, detailing how the focus group will operate, who will facilitate the group, and what their role would be. Since each participant would know the identities of other participants and have digital access to the focus group discussions, participants were asked to treat the discussions as confidential.

TABLE I
TOPIC AREAS BY WEEK

Week	Topic Area
1-2	Defined "pre-college computing activity" and discussed variables important for evaluating activities
3-4	Sought input on the website design and repository content
5-6	Discussed gaps in data collection and reporting in relevant research, including variables
7-8	Explored a system for rating research and discussed how to enable the creation of quality research
9-10	Explored the value and feasibility of quality standards, including what basic criteria might be universally acceptable as well as objectively attained

B. Data Collection

We formed a Google Group to conduct the focus group discussions and we served as its moderators. Throughout the discussions, we worked to create a welcoming, non-threatening environment where participants felt comfortable sharing a range of responses by stating our goals and guidelines for the discussions [11]. We carefully posed prompts and follow-up questions, were available to answer questions, and followed the format as initially defined.

In preparation for the first prompt, we asked each researcher to introduce themselves to the group and explain why they agreed to participate in the study. In addition to providing the group with a means of getting to know each other, the prep questions allowed each participant to gain familiarity with the discussion format.

Over the following 10-week period, each participant spent approximately 6 to 8 hours reviewing and commenting on the prompts. In weeks 1, 3, 5, 7, and 9, we posted one or more prompts for the week and asked participants to respond by week's end (see Table I). In weeks 2, 4, 6, and 8, we asked follow-up questions based on the previous week's responses. In week 10, we posted a debriefing of the progress of the project, including a summary of how their responses will shape it.

C. Data Analysis

To review the data, we adapted a systematic design approach for grounded theory, emphasizing the "...use of data analysis steps of open, axial, and selective coding, and the development of a logic paradigm..." [24, p. 343]. We centered our questions around issues pertaining to the repository content and the website in which it would be housed. We used techniques similar to grounded theory research to categorize the feedback on the prompts in a structured manner [23, 25]. We describe our methodological approach using the same terminology as in grounded theory.

The primary moderator for the virtual focus group also served as the primary interpreter of the data. The primary interpreter sought input twice from the other researcher and an independent researcher familiar with qualitative techniques. These two phases were at the start of the data collection process, to determine the soundness of the methods of evaluating the data, and after three passes of coding. In preparation for coding, we transferred each statement into a spreadsheet. We separated and entered distinct clauses into separate fields in order of how they were made during each discussion, so that we could review each statement on a continuum and within context. To analyze the data, we made four passes of the statements over a one-week time frame:

1) Created open codes and reviewed them for consistency,
2) Developed Axial codes based on open codes and sorted statements by these codes,
3) Refined open codes further and created four selective codes (themes) based on these, and
4) Made minor revisions to axial codes to assimilate and combine some codes and separate others.

IV. RESULTS AND DISCUSSION

Early in the discussions, we asked participants to define "pre-college computing activity". We initially proposed a broadly structured definition, and participants proposed revisions and added examples for clarification. The definition evolved into the following:

> A pre-college computing activity is an activity or process that teaches computing or computational thinking and is experienced outside of a K-12 school's formal in-class curriculum. The term *formal in-class curriculum* refers to curricula that fulfill state/national education requirements and/or content that is offered as part of a required or elective course or module. Activities or processes that supplement formal in-class curricula, however, are included.
> For example, pre-college computing activities that will be considered are: after-school or out-of-school computer clubs, summer camps, robotics leagues, activities associated with CS education week, or a computer scientist coming into class to speak on career day. Alternatively, activities that will not be considered as pre-college computing are, for example, course materials developed for Exploring Computer Science or the CS AP exams.

After the four passes of the data were finished, four categories evolved:

- Repository Design and Content,
- Content Quality Measures,
- Challenges of the Repository, and
- Functional Requirements of the Website.

One of the overarching themes throughout the discussions related to quality of the educational research in computing. Participants' statements about research quality indicated a desire to see the general quality of research within the computing educational community improve. We note this prior to discussing specific feedback, so the reader can approach each category keeping a broader vision of quality in mind.

A. Repository Design and Content

Participants actively engaged in discussing the repository design and content, which were grouped into two subcategories: Functionality/Content and Variable Identification.

1) Functionality/Content: Many of the comments regarding the design of the repository centered around the search and filtering mechanism. Participants suggested that the search capability be powerful and the filtering mechanism be granular through the ability to conduct multi-level searchers and then refine the search on additional variables through a dynamic checklist.

Results should show a summary page for each study and activity, including its abstract, links to the publication, a summary of data collected (if any), and the activity's website, if it has one. Participants suggested using machine learning to improve the search process and suggested a public API so meta-researchers can analyze the corpus as a whole.

Participants commented on the need for a clearly defined data dictionary for each of the search criteria that would be made available to users. One participant suggested thorough usability testing of the search mechanism, the ability to track analytics (by tracking what variables are searched), and ensure that the search provided a positive user experience. Specifically, participants suggested providing the capability for users to search on each of the following study variables:

- Study Framework (including multiple forms of quantitative, qualitative, and mixed method studies),
- Ethics Considerations (IRB Approval, Participant Protection and Privacy),
- Stated Outcomes (Learning Outcomes, Attitudes/Perceptions),
- Types of instruments used in the study,
- Venue publication type (journal, conference, self-published),
- Type of peer review, if any,
- Multiple presentations of activities under different times/locations taught by different people, and
- Whether or not the study was a replication of previous work or the study has been replicated.

Participants had mixed thoughts about what should be included in the repository, with two stating that all relevant papers should be included without regard to quality. One participant stated that it would be beneficial to include a wide variety of studies, rather than focusing on only quantitative studies. Two participants declared a need to ensure that qualitative studies were accommodated, since they are often overlooked. Two others stated that only papers with empirical data should be included and papers should be excluded based on quality. Section 3.3 provides more detail into the feedback on quality measures.

These frameworks and more formal approaches to conducting studies address some of the ongoing challenges in computing education research, such as CS content knowledge, social science research methods, and statistical analysis [1, 10, 26, 27, 28]. The emphasis on qualitative research supports previous research in the computing education community that found that qualitative methods are often overlooked, but they have the power to provide significant value, particularly in evaluating the ”why and how” behind the success or failure of various practices [1, 29, 30]. Although qualitative studies are conducted less frequently, their rigor is critical to the success of fortifying the computing education research landscape [31]. In has also been noted that grounded theory has been used extensively throughout the social sciences, though it is rarely seen in computing education research [32]. Given the participant feedback and the noteworthy absence of qualitative studies thus far, the value and importance of qualitative research must be reflected considered in the development of the repository.

2) Variable Identification: Since one of the goals of the project is to identify variables that should be collected in studies, participants offered several ways to classify the variables. One proposed a structure of participant demographics, participant prior experience, participant attitudes, instructor structural implementation, and instructor self-efficacy. Another suggested separating the variables "...into non-cognitive (engagement, self-efficacy, motivation, etc.) and cognitive (e.g., learning, problem-solving abilities, etc.) variables." Another suggested labeling "...variables as fixed or measured (gender is fixed, attitude is measured)."

Considering their comments and the variables that were suggested, we grouped the variables by activity structural components, demographics, and knowledge, skills, and dispositions (Table II). Participants noted that not all studies must collect and analyze each of the variables identified. It was also noted that study participants can be anyone in the study (e.g. student, instructor, or administrator) and these variables may apply to one or both groups.

B. Measuring Content Quality

Participants confirmed what has been stated previously about challenges in computing education research–that evidence, rigor, and formal methodologies, in general, have been lacking [1, 13, 14, 31, 33, 34, 35]. The need for high quality research through a shift towards more rigorous research methods still remains [36].

Specifics on how to measure quality and with what metrics were discussed by nearly each participant during each discussion period. We review here the primary discussion points, including the quality rating methods, specific metrics that should be considered, and the challenges associated with measuring quality.

Nearly all participants were in favor of having a method for rating the quality of papers and being able to use this rating system when searching for related articles and activities, with one noting that a "rating system would be very useful for practitioners" as well as researchers. The group suggested several methods for achieving a useful rating system, explored over all 5 discussion periods. Methods suggested included:

- A "criteria rating system ordinal (1, 2, 3, 4)" where an article must meet a specific set of criteria to achieve

TABLE II
VARIABLE IDENTIFICATION

Participant Demographics	Activity Components	Participant Knowledge/Skills/Dispositions
• Age • Backgrounds of learners (homogeneous, disparate) • Ethnicity • Gender • Grade levels • Income • Personality • Prior CS education (formal in-school and out-of-school) • Prior CS experience (programming, robotics, etc.) • Prior Experience • Race • Socio-economic status	• Accommodations for Learners with Disabilities • Date of activity • Duration of activity • Facilitation method (Instructor? Students? Media) • Follow-up activities • Frequency of activity • Grouping strategies (Peer instruction, etc.) • Instructor (Activity designer, instructor only, etc.) • Location of activity • Materials needed • Number of participants • Omission of components in activity • Physical space needed • Preparation time • Programming language used (if applicable) • Resources needed to conduct activity (human, money) • Supplementation of components in activity • Time needed for the activity • Type of activity (summer camp, after-school club, etc.)	Cognitive • Basic computer knowledge • Knowledge of careers in computing • Knowledge of ubiquity in computing • Skill measures • Technical knowledge • Understanding of current misconceptions about computing • User knowledge Non-cognitive • Attitude (Instructors, peers) • Desire to participate in more pre-college CS courses and/or activities • Engagement (Interaction with instructors, peers) • Interest • Motivation • See themselves doing something in computing • Self-efficacy • Sense of belonging

a particular rating. The "bottom rating should be very basic (so papers aren't arbitrarily excluded)" and "Level 1 really is about basic components of a study." Another suggested that ordinal levels may be ineffective and using a scale such as "was informative, was interesting, was inspiring, was inspiring and changed behavior" may be more intuitive.

- Metrics on the number of downloads to demonstrate a level of community engagement and potentially how many times the paper was cited.
- A scoring rubric ("use rubric if checklist not doable"), with possible "multi-dimensional rubrics compared to checklists"
- A binary checklist with an up/down voting system to keep reviewing in the hands of the user community.

Participants suggested that authors could suggest corrections and potentially post additional information about the study directly on the website. Participants also noted that even if a paper listed its author contact information, number of activity participants, gender and race of participants, and self-efficacy scores, this does not necessarily make it a quality study. However, meeting basic technical components in the report will allow for replication and assurance that basic educational research standards were met.

Additional challenges discussed related to sustainability, completeness, and validity. One participant wanted to know how "experience reports will be classified," since their quality measures will be different than those that use experimental design. Likewise, another participant wanted to ensure that "all methodologies are valued in any rating system."

One participant acknowledged the issue of "too much bias in these [rating] systems", with others noting that the experience level and qualifications of the reviewers will be critical for any rating system to be effective. To actually review and rate each published study for its research quality would require trained and qualified reviewers on an ongoing basis. Additionally, the "complexity of the system may be overwhelming (completeness/thoroughness vs usability)", as one participant noted. Since this is essentially out of scope for our particular project, participants circled back to the idea of a simplified checklist as an achievable, basic addition to the website. This simplified version was discussed as a way for users to rate the quality of the papers as well. Users could rate each based on quality and usefulness as they considered papers in the repository.

Participants suggested several frameworks to consider when assessing research quality. Correctness, Rigorousness, Appropriate, and Pedagogical Soundness encapsulated many of the thoughts about how to consider quality. Another framework proposed was Bias, Methodology, Rigor, Generalizability. A third embraced the concept of examining studies in light of the Child, the Content, and the Context. These frameworks and what methods for assessing quality of the research generated a list of questions and criteria. We then found a natural grouping for these questions and provide them in Table III.

C. Repository Value and Related Challenges

With respect to challenges related to the repository, four subcategories were generated: Data Management, Project Feasibility, Project Value, and Project Sustainability. Each partic-

TABLE III
POTENTIAL QUALIFYING QUESTIONS FOR ARTICLES INCLUDED IN THE REPOSITORY

Purpose, Goals, Intent, Clarity: Do the authors...
- Make a case for why the reader should care about the problem?
- Provide their contact information for the activity/study organizer/instructor/designer?
- Clearly and explicitly state the research question(s) and hypothesis?
- Clearly state the study's objectives, including articulating any learning outcomes?
- Use correct language related to educational researcher?
- Provide any definitions used that are crucial to the study?
- Specify the research question(s) the study sought to answer?

Study Design: Do the authors...
- Indicate the research methodology used and the rationale for that choice?
- Use an appropriate design related to its type of study?
- Describe the methodology in sufficient detail for another researcher to replicate the study?
- Describe the methodological framework (quantitative, qualitative, mixed methods) in terms of educational research? (Qualitative: case studies, ethnography, longitudinal, etc.; Quantitative: (quasi) experimental designs, survey, etc.)
- Describe any efforts to offset the novelty effect, Hawthorne effect, John Henry effect?
- Use and rigorously apply instruments appropriate to the research question?
- Describe and provide the instruments used within the study?
- Fully describe the setting for the study (location, classrooms, courses, schools)?
- Use an appropriate instrument to measure impact?
- Consider sample size and whether it is sufficient?

Activity/Intervention: Do the authors...
- Fully describe the intervention and/or activities?
- Explain how the activity is suitable to the targeted participant group (age/range/experience/etc.)?
- Describe the skill, knowledge, or disposition that was being targeted?
- Describe the length and frequency of the intervention (hours, days, months)?
- Describe who conducted the intervention, including qualifications?

Ethics: Do the authors...
- Disclose their IRB approval process and methods to ensure participant privacy, confidentiality, and protection?
- Disclose any costs/funding sources to conduct any aspect of the research/activity in order to assess possible bias?
- Disclose whether or not participants or researchers receive monetary or gift incentives?
- Include researchers qualifications and how researcher bias has been mitigated?
- Declare any personal, organizational, or institutional biases?

Participants: Do the authors...
- Include participant demographic information, including age, grade range, gender, race/ethnicity, socioeconomic status?
- Include number of participants in the study?
- Include recruitment process for participants (volunteer? required?)?
- Describe sampling technique used?

Data Analysis: Do the authors...
- Indicate the analysis methods and tools used and the rationale for those choices?
- Describe how the data analysis methods were appropriate for the design?
- Fully describe the analysis methods with sufficient detail for replication?
- For quantitative frameworks, describe all statistical tests used and a rationale for non-standard measures used? Include or provide a link to the raw assessment data for others to verify/analyze? Distinguish between correlation and causality?
- For qualitative frameworks, describe how the data was analyzed, how inter-rater reliability was maintained, and provide researcher reflexivity statement?

Results: Do the authors...
- Provide a compelling argument (sample size, quantitative or qualitative analysis, etc.) for the significance of its results?
- Describe the results of the study?
- Explore the implications of the results on research, policy, and practice?
- Describe how this research and/or results fit into the larger context of related research?
- Consider whether the results are appropriate for the scale of the intervention?
- Describe limitations of the study, including issues related to ability to generalize, sample size, confounding variables, whether or not participants were randomized or not representative, with any alternative hypothesis stated?
- Include data (sample size, statistical analysis, etc.) indicating its significance?

ipant mentioned that the project had value over what is currently available for computing educational researchers, noting that the value is in categorizations to help people search data, a user-friendly search and filtering capability, classification of papers based on content and/or quality, and the value if the research repository only published significant findings. One participant noted that the "value is in giving clues about the research in small pieces", referring to the notion that results from different research can provide a broader picture of the research. Another noted that the project will be valuable to "research, policy, and practice."

1) Data Management: Participants noted that managing the data and ensuring privacy and confidentiality of activity participant data in the data repository were significant hurdles to overcome. Another mentioned the accuracy of the data and how reliability and validity would be handled. On the more technical side, two participants questioned how missing data would be handled within the filtering system ("How is 'not applicable' data going to be handled?"). These challenges correspond with the challenges noted previously by Sanders, et al [18], including managing differently formatted data, institutional, legal, and ethical considerations, and building community support. Specifically, privacy and anonymity concerns, context and methodological concerns, and how and who contributes data are all issues that need to be addressed.

2) Feasibility: Project feasibility related to participant feedback that the project could actually be undertaken and completed. Participants noted that the project was doable and valuable, but that boundaries must be clearly set to contain scope. One participant mentioned that to contain scope, papers could be evaluated for quality at a later time. Another noted that to reduce errors, a correction request mechanism could be made available to authors if they found any errors in the categorization of their research.

3) Project Value: Quality measures were important for the participants. However, participants felt that it would be threatening for the researchers/authors to have their previously submitted materials reviewed again, particularly considering that in some cases the review may be more thorough. Reviewing each and every entry for quality also will be time-consuming and beyond the scope of the project. Further discussion with the broader community needs to occur to weigh the pros and cons of such an approach.

4) Project Sustainability: Sustainability was also an important topic mentioned by the participants without prompting from the moderators. One participant noted that the upkeep of the research repository would be labor intensive, but that transitioning to volunteer reviewers would also be labor intensive and create its own set of challenges, including reviewer qualifications and inter-rater consistency. Another participant mentioned the possibility of automating the process of data collection for the research repository by creating a "mechanism for authors to have their papers included".

D. Functional Requirements of the Website

Participants provided requirements for the website and for the repository over three of the five discussion periods. Statements regarding the repository were placed into one of the selective codes discussed in previous sections, and this section focuses on general comments about functionality. Two subcategories were formed from the participant feedback: User Experience and additional researcher content in the form of Guides for Researcher.

1) User Experience: With respect to user experience, one participant noted that the user interface of the entire website is more important than any quality indicator for the content of the repository. Three other comments referred to the search system for the repository, with one suggesting a filtering system similar to newegg.com and another to ICPSR from University of Michigan [38]. A final comment provided was a suggestion to have "samples for different datasets for quick reference (like sample size, grade level, students, teachers, admin, etc.)." This feedback aligns with Sanders, et al, who noted that an easy-to-use interface and creating an easily searchable repository with easy to understand results would be paramount to the success of such a project [18]. This is consistent with general design practices for developing websites that promote good user experiences [39, 40].

2) Guides for Researchers: Participants supported the idea of providing guides to researcher, including a general description about what scientific inquiry is and how it pertains to educational research (forming research questions, writing good learning outcomes, assessment, and reporting data in a reliable and valid way). Suggestions to do this ranged from tutorials, links to other websites, "how to" tools, case studies, and checklists for researchers. One participant noted to "keep in mind that grad students interested in CS ed research may not have formal training." This related back to our general discussions about how to measure quality in research papers that may appear in the research repository, and one participant suggested providing a "general list of quality indicators on the website to think about when looking at papers" and providing data dictionaries ("codebook/ coding manual") for each variable used in either repository.

Guides for new researchers may also alleviate other issues in computing education research, including the lack of a strong evidence base at pre-college levels [34], as well as a need for collaborations between researchers and practitioners [41]. Daniels and Pears state that "[w]ithout higher order research frameworks systematic research in [Computing Education Research] will ultimately lack power and credibility" [35, p. 101], while Valentine calls on researchers to "prove that you did what you said that you did!" [14, p. 259]. The use of formal methodologies to improve computing education research has been and continues to be emphasized [10, 13, 31, 37], and providing such a guide to new researchers or educators who may want to contribute to the body of knowledge within the community can be of value [31, 41].

V. LIMITATIONS

Although great care was taken in the development of this virtual focus group, we are aware of several limitations that are worthy of consideration when putting this study into context [42]. The virtual focus group was limited to a one-time 10-week discussion with 6 active participants. Holding several additional focus groups would have provided a broader set of input and would have created more assurance that we had reached data saturation from the participants.

Another limitation was that participants were researchers and evaluators. We may have received broader feedback if we had included others who might this information helpful, including teachers and school administrators. Having designers/developers of existing repositories, like NCWIT's EngageCSEdu as part of the discussion would have also added another perspective to ensure our repository design was strong.

Although the virtual focus group format met our participants' geographic needs, we are aware that an in-person focus group may have yielded greater feedback. An in-person focus group could have been formatted as a one day event, with follow-up on-line discussions to verify the feedback received from such an event. This latter approach would have its own drawbacks and limitations, but may have opened the door to greater freedom of suggesting new ideas and approaches (beyond those that we received from the virtual focus group format).

VI. CONCLUSION

As noted by the U.S. National Science Board, educational repositories can provide a fundamental change to the field, providing previously hidden insights. Such data collections "are a powerful force for inclusion, removing barriers to participation at all ages and levels of education" [19, p. 9]. Many computer science education researchers come from a strong technical background with little educational research training. Providing guides and checklists for creating quality research may help new researchers understand what is important in educational research and why. The ultimate effect can be one where the researchers use formal methodologies, validated evaluation instruments, and formal reporting of results within computing education research. The result of this can be a larger quantity of research that is of higher quality, providing the community with a rich set of results that can be compared and analyzed. From this, empirically-based best practices for students of differing demographics can be more readily defined and be tailored to the experiences and interests of the curriculum designers and instructors.

We have provided here a framework for others to follow when creating resources for the community. The focus group format and analysis of the data provided by the members allows for us to have a degree of confidence that the design and information for this resource will satisfy some of the needs of the community. Together with additional data collection, website development, and user testing, we can produce a product that will be valid, meaningful, and useful to the wider computer science education community.

ACKNOWLEDGMENT

The authors would like to thank those who participated in the virtual focus group. We also extend our thanks to Jeff Xavier of SageFox Consulting for providing guidance and feedback on our study design, implementation, and analysis. This material is based upon work supported by the U.S. National Science Foundation under Grant Nos. 1625005, 1625335, and 1757402.

REFERENCES

[1] V. L. Almstrum, O. Hazzan, M. Guzdial, and M. Petre, "Challenges to computer science education research," *SIGCSE Bull.*, vol. 37, no. 1, pp. 191–192, Feb. 2005. [Online]. Available: http://doi.acm.org/10.1145/1047124.1047415

[2] Codecademy. (2017) Codecademylearn to code interactively, for free. [Online]. Available: https://www.codecademy.com/

[3] Code.org. (2017) Code.org. [Online]. Available: https://code.org

[4] Computer Science Education Week. (2017) Cs ed week. [Online]. Available: https://csedweek.org

[5] csteachers.org. (2017) Computer science teachers association. [Online]. Available: https://csteachers.org

[6] Girls Who Code. (2017) Girls who code: Join 40,000 girls who code. [Online]. Available: https://girlswhocode.com/

[7] National Science Foundation. (2017) Broadening participation in computing alliance program (bpc-a). [Online]. Available: https://www.nsf.gov/

[8] A. Decker and M. M. McGill, "Towards a common framework for evaluating computing outreach activities," ser. SIGCSE '16. New York, NY, USA: ACM, 2016.

[9] A. Decker and M. McGill, "Pre-college computing outreach research: Towards improving the practice," ser. SIGCSE '17. New York, NY, USA: ACM, 2017.

[10] S. Fincher and M. Petre, *Computer Science Education Research*. London, UK: Routledge Falmer, 2004.

[11] P. Liamputtong, *Focus Group Methodology: Principles and Practice*. London, UK: SAGE Publications Ltd., 2004.

[12] S. E. J. Randolph, G. Julnes and S. Lehman, "A methodological review of computer science education research," *Journal of Information Technology Education*, vol. 7, p. 135162, 2008.

[13] J. Tenenberg, "Asking research questions: Theoretical presuppositions," *Trans. Comput. Educ.*, vol. 14, 2014.

[14] D. W. Valentine, "Cs educational research: A meta-analysis of sigcse technical symposium proceedings," in *Proceedings of the 35th SIGCSE Technical Symposium on Computer Science Education*, ser. SIGCSE '04. New York, NY, USA: ACM, 2004, pp. 255–259. [Online]. Available: http://doi.acm.org/10.1145/971300.971391

[15] M. M. McGill, A. Decker, and Z. Abbott, "Improving research and experience reports of pre-college computing activities: A gap analysis," ser. SIGCSE '18. New York, NY, USA: ACM, 2018.

[16] A. Craig, "Theorising about gender and computing interventions through an evaluation framework," *Info Systems J*, vol. 26, p. 585611, 2016.

[17]

[18] K. Sanders, B. Richards, J. E. Moström, V. Almstrum, S. Edwards, S. Fincher, K. Gunion, M. Hall, B. Hanks, S. Lonergan, R. McCartney, B. Morrison, J. Spacco, and L. Thomas, "Dcer: Sharing empirical computer science education data," in *Proceedings of the Fourth International Workshop on Computing Education Research*, ser. ICER '08. New York, NY, USA: ACM, 2008, pp. 137–148. [Online]. Available: http://doi.acm.org/10.1145/1404520.1404534

[19] National Science Foundation. (September 2005) Long-lived digital data collections enabling research and education in the 21st century. technical report nsb-05-40. [Online]. Available: http://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf

[20] M. Plakhotnik and T. Rocco, "Four components of a manuscript reporting qualitative empirical studies: A brief overview," in *Proceedings of the Eighth Annual College of Education and GSN Research Conference*, Miami: Florida International University, 2009, pp. 105–111.

[21] What Works Clearinghouse. (2017) What works clearinghouse. [Online]. Available: http://ies.ed.gov/ncee/wwc/

[22] A. Onwuegbuzie, W. Dickinson, N. Leech, and A. Zoran, “A qualitative framework for collecting and analyzing data in focus group research,” *International Journal of Qualitative Methods*, vol. 8, 2009.

[23] M. Patton, *Qualitative research and evaluation methods (3rd ed.)*. Thousand Oaks, CA: Sage, 2002.

[24] J. Crewsell, *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Upper Saddle River, NJ, USA: Merrill Prentice Hall., 2008.

[25] A. Strauss and J. Corbin, *Basics of Qualitative Research Techniques and Procedures for Developing Grounded Theory*. London, UK: Sage Publications, 1998.

[26] M. Joy, J. Sinclair, S. Sun, J. Sitthiworachart, and J. Lopez-Gonzales, “Categorising computer science education research,” *Education and Information Technologies*, vol. 14, no. 2, pp. 105–126, 2009.

[27] L. Malmi, J. Sheard, Simon, R. Bednarik, J. Helminen, P. Kinnunen, A. Korhonen, N. Myller, J. Sorva, and A. Taherkhani, “Theoretical underpinnings of computing education research: What is the evidence?” in *Proceedings of the Tenth Annual Conference on International Computing Education Research*, ser. ICER ’14. New York, NY, USA: ACM, 2014, pp. 27–34. [Online]. Available: http://doi.acm.org/10.1145/2632320.2632358

[28] A. Robins, “The ongoing challenges of computer science education research,” vol. 25, p. 115119, 2015.

[29] S. Fitzgerald, B. Hanks, and R. McCauley, “Collaborative research in computer science education: A case study,” in *Proceedings of the 41st ACM Technical Symposium on Computer Science Education*, ser. SIGCSE ’10. New York, NY, USA: ACM, 2010, pp. 305–309. [Online]. Available: http://doi.acm.org/10.1145/1734263.1734367

[30] O. Hazzan, Y. Dubinsky, L. Eidelman, V. Sakhnini, and M. Teif, “Qualitative research in computer science education,” *SIGCSE Bull.*, vol. 38, no. 1, pp. 408–412, Mar. 2006. [Online]. Available: http://doi.acm.org/10.1145/1124706.1121469

[31] A. Berglund, M. Daniels, and A. Pears, “Qualitative research projects in computing education research: An overview,” in *Proceedings of the 8th Australasian Conference on Computing Education - Volume 52*, ser. ACE ’06. Darlinghurst, Australia, Australia: Australian Computer Society, Inc., 2006, pp. 25–33. [Online]. Available: http://dl.acm.org/citation.cfm?id=1151869.1151875

[32] P. Kinnunen and B. Simon, “Building theory about computing education phenomena: A discussion of grounded theory,” in *Proceedings of the 10th Koli Calling International Conference on Computing Education Research*, ser. Koli Calling ’10. New York, NY, USA: ACM, 2010, pp. 37–42. [Online]. Available: http://doi.acm.org/10.1145/1930464.1930469

[33] T. Clear, “Valuing computer science education research?” in *Proceedings of the 6th Baltic Sea Conference on Computing Education Research: Koli Calling 2006*, ser. Baltic Sea ’06. New York, NY, USA: ACM, 2006, pp. 8–18. [Online]. Available: http://doi.acm.org/10.1145/1315803.1315806

[34] J. Cuny, “Cs education: Catching the wave,” in *Proceedings of the 47th ACM Technical Symposium on Computing Science Education*, ser. SIGCSE ’16. New York, NY, USA: ACM, 2016, pp. 3–3. [Online]. Available: http://doi.acm.org/10.1145/2839509.2844551

[35] M. Daniels and A. Pears, “Models and methods for computing education research,” in *Proceedings of the Fourteenth Australasian Computing Education Conference - Volume 123*, ser. ACE ’12. Darlinghurst, Australia, Australia: Australian Computer Society, Inc., 2012, pp. 95–102. [Online]. Available: http://dl.acm.org/citation.cfm?id=2483716.2483728

[36] A. Lishinski, J. Good, P. Sands, and A. Yadav, “Methodological rigor and theoretical foundations of cs education research,” in *Proceedings of the 2016 ACM Conference on International Computing Education Research*, ser. ICER ’16. New York, NY, USA: ACM, 2016, pp. 161–169. [Online]. Available: http://doi.acm.org/10.1145/2960310.2960328

[37] A. Berglund, P. Kinnunen, and L. Malmi, “A doctoral course in research methods in computing education research how should we teach it?” in *Proceedings of the Seventh Baltic Sea Conference on Computing Education Research - Volume 88*, ser. Koli Calling ’07. Darlinghurst, Australia, Australia: Australian Computer Society, Inc., 2007, pp. 175–178. [Online]. Available: http://dl.acm.org/citation.cfm?id=2449323.2449345

[38] ICPSR. (2017) Icpsr from university of michigan. [Online]. Available: https://www.icpsr.umich.edu/icpsrweb/ICPSR/

[39] J. Rubin and D. Chisnell, *Handbook of Usability Testing, Second Edition: How to Plan, Design, and Conduct Effective Tests*. Indianapolis, IN: Wiley Publishing, Inc., 2008.

[40] J. Tidwell, *Designing Interfaces*. Sebastopol, CA: OReilly Media, 2006.

[41] D. Franklin, C. Hill, H. Dwyer, A. Iveland, A. Killian, and D. Harlow, “Getting started in teaching and researching computer science in the elementary classroom,” in *Proceedings of the 46th ACM Technical Symposium on Computer Science Education*, ser. SIGCSE ’15. New York, NY, USA: ACM, 2015, pp. 552–557. [Online]. Available: http://doi.acm.org/10.1145/2676723.2677288

[42] D. T. Campbell, J. C. Stanley, and N. Gage, *Experimental and Quasi-Experimental Designs for Research*. Boston, MA, USA: Houghton Mifflin, 2006.