2023 Annual Conference & Exposition
Baltimore Convention Center, MD | June 25 - 28, 2023
The Harbor of Engineering Education for 130 Years
Paper ID #39681

Common Metrics: Lessons from Building a Collaborative Process for the Examination of State-level K–12 Computer Science Education Data

Rebecca Zarch, SageFox Consulting Group

Rebecca Zarch is an evaluator and a director of SageFox Consulting Group. She has spent nearly 20 years evaluating and researching projects in STEM education from K-12 through graduate programs.

Sarah T. Dunton
Jayce R. Warner, University of Texas, Austin
Mr. Jeffrey Xavier
Joshua Childs, University of Texas, Austin
Dr. Alan Peterfreund, SAGE

Common Metrics: Lessons from Building a Collaborative Process for the Examination of State-level K–12 Computer Science

Abstract

The broadening participation in computing (BPC) movement has evolved from a focus on increasing the number and diversity of students who participate in computing education to addressing the systemic barriers that prevent equity in computing education. Data has historically played a significant role in the BPC movement, however it is not neutral and can be used to both mask and perpetuate the systemic inequities in computing education. One strategy for ensuring equitable computing experiences is to develop a data infrastructure that democratizes data.
This research paper presents the early results of how two cohorts of multi-state teams focused on BPC in K–14 through equity-explicit changes in policy, pathways and practices have worked to develop state-specific data infrastructure that can help safeguard against these potential harms.
Research Questions: What are the components of a data infrastructure that supports equity driven computer science education efforts? And, how feasible is it to create cross-state data systems in support of the national BPC movement?

Eleven states participated in a four-phase, facilitated process over 9 months to build a multi-state data infrastructure, uncovering state-specific needs as they worked the project. Teams learned from their neighbors, held each other accountable and uncovered the opportunities and barriers within their unique data context. The process involved framing of the BPC goals, assessing current data systems for points of vulnerability and opportunity, a data request from the state systems, data visualization, utilization and reflection.

Data was drawn from reflective team journals; notes, collaborative materials and observations made during collaborative meetings; and the technical assistance requests made during the project.

When developing data infrastructure in support of BPC, diverse teams matter. Teams needed representation from people who can access institutional or state data, understand the practical context of the data to support interpretation, and help tie data to broader advocacy efforts. Ongoing team engagement, both within and across states, allowed the space to consider the complexity of issues and maintain sustained engagement and commitment to the topic in an environment with competing educational priorities.

Teams uncovered assumptions many states had about the access to and quality of their data. Issues with course alignment across and within states about how courses are identified and entered into the data systems; the software systems data storage; and access in a politically charged climate all emerged. Data suppression rules and the implications for invisibilizing populations were apparent. Finally, common-enough definitions for computer science seem possible and desirable but also challenging based on state policy contexts.

This project leads to clear implications for creating data infrastructure; most importantly data needs to be clearly tied to equity driven questions and purpose; data cannot be examined in a void and data systems should be revised according to equity needs. States are often using data systems outside of the intended design and face limitations when trying to surface inequity for populations based on gender, disability, ethnicity and race, which are exacerbated when looking intersectionally. Across states, teams are now asking deeper and more complex questions about pathways, policy, and purpose.

Introduction

Building off prior efforts to understand the evolution of measurement approaches in the K–12 broadening participation in computing (BPC) movement [1], this work provides a detailed look into collaborative processes for examining state-based data within and across states in support of equity in computer science education (CSEd). The BPC movement was initiated with the goal of significantly increasing the number of students pursuing post-secondary degrees in computing disciplines, focusing on demographic diversity [2]. Over time the movement has expanded to address systemic barriers that limit schools' ability to equitably reach all students, promoting diversity in the classroom and among degree holders in higher education [3], [4]. As with many emerging efforts, as the movement matured, BPC approaches moved from ad-hoc implementation and measurement efforts [5], [6] to more coordinated approaches in CSEd from K–12 through the workforce [7]–[9].

With increased BPC-focused education policy, pedagogical practices and programs are scaling rapidly, and knowing what works and for whom [10] is an important component for developing strategies and monitoring progress against those approaches [10]–[12]. Equity-focused measurement can also help shift the deficit perspectives of minoritized students and communities that have historically been misrepresented within educational data.

Creating effective and equity-focused measurement systems, however, has proven to be a challenging and ongoing process, both at a national level and for individual initiatives [13]–[15]. The highly localized nature of educational decision-making in the United States, combined with varying policies, definitions, and data collection systems, makes it challenging to understand the landscape of CSEd. Furthermore, the lack of coherence that exists between different levels of the U.S. educational system complicates aligning interventions that reach down to students inside of classrooms. Despite these challenges, these efforts have significantly advanced our knowledge of BPC initiatives and the difficulties in measuring their progress. These efforts have been aided by a culture of collaboration in the CSEd measurement space [16]–[18].

Theoretical Framing: The CAPE framework

As CSEd opportunities have grown, the BPC movement has worked to ensure that every student not only has access to, but is recruited and retained in computing classes [1], [19], [20]. The CAPE framework [4] is a holistic guide that has the potential to bolster BPC efforts by addressing four key components of CSEd: the Capacity for, Access to, Participation in, and Experience (CAPE) of CSEd (See Fig. 1). These dimensions are progressive in that success at one level relies on success at the level preceding it. For students to have good experiences in CSEd, they first need to participate in computer science (CS) courses. To enroll in CS courses, students must first have access to those courses (i.e., their schools must offer the courses). To offer CS courses, schools and districts must first have the capacity (i.e., funding, staffing, resources, etc.) to do so. The central idea of the CAPE framework is that there are equity issues to be addressed at each of these levels. The capacity to offer CSEd and students' access to CSEd may be unevenly distributed across different types of schools and districts. When students do have access to courses, there may be disparities in enrollment rates between different student subgroups. When students do enroll in CS courses, there still may be inequities in terms of which students feel included and which students ultimately benefit from participating in those courses. The relationships between the four components of CAPE and examples of equity issues to address within each component are represented in Figure 1. In our work, we utilized CAPE as our framework for understanding how to measure and address equity in CSEd.

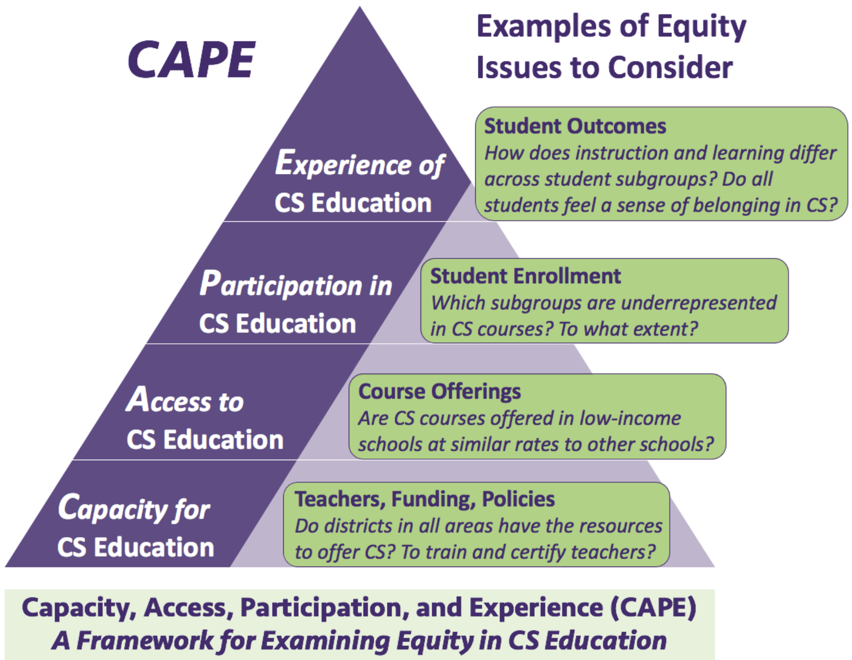

Figure 1: CAPE Framework

The Expanding Computing Education Pathways Alliance as a Laboratory for Data

The Expanding Computing Education Pathways (ECEP) Alliance, a National Science Foundation-funded Broadening Participation in Computing (BPC) Alliance, seeks to increase the number and diversity of students in the pipeline to computing and computing-intensive degrees by promoting state-level CSEd reform [21]. ECEP supports state leadership teams in order to develop effective and replicable interventions to broaden participation in computing (BPC) and to create state-level infrastructure to foster equitable computing education policies.

Data has historically played a significant role in the BPC movement, however it is not neutral and can be used to both mask and perpetuate the systemic inequities in computing education. ECEP states have moved from ad-hoc measurement efforts such as surveys of known CS teachers to including CS in their state data systems [22], [23]. ECEP has begun to test a data infrastructure project that democratizes data, the Common Metrics Project (CMP). The CMP grew out of 10+ years of work in systems change efforts designed to build equitable pathways in K–16 CSEd. The objective of the CMP is to facilitate the collaborative development of a common framework for monitoring progress in K–16 CSEd and advocacy with an emphasis on broadening participation. This project engages state teams in data gathering, data analysis, data reporting, and data utilization activities that have led to an increase in evidence-based BPC advocacy, policy development, and implementation. The team engagement work, as well as ECEP's 5-Stage Model for State Change, serve as guides for both state change and ECEP's collective impact work [24]. The CMP takes a deep dive into the 'Build/utilize data infrastructure to provide evidence to inform strategic BPC efforts' stage.

This research paper presents the early results of how two cohorts of multi-state teams focused on BPC in K–12 through equity-explicit changes in policy, pathways and practices have worked to develop state-specific data infrastructure that can help safeguard against potential harms. Due to the early and exploratory nature of this research, states will remain anonymous in this paper with the exception of Texas, whose data ecosystem map (Fig. 2) is shared with permission.

Research Questions:

With two cohorts of states completing the common metrics process (CMP) and a third underway, this paper explores the way the CMP has evolved, its utility to teams and the potential for this process to work across states. Specifically, this paper explores:

1. What are the components of a data infrastructure that supports equity-driven computer science education efforts? And,
2. How feasible is it to create cross-state data systems in support of the national BPC movement?

Although the ECEP and ECEP CMP teams do not seek to privilege quantitative data from state data systems over other kinds of data and knowledge, this paper does focus on state data systems as one sustainable and relatively low-cost approach to understanding a longitudinal perspective associated with access and participation in CSEd at the K–12 level. Anyone doing comprehensive BPC work will need to consider other sources of information to appropriately understand the capacity to offer high quality CSEd, the experiences of students in CSEd, and even what may be influencing access and participation. This includes working with local communities to be led by their perspectives and insight on possible BPC/CSEd interventions. It also includes gathering qualitative information, through empathetic interviews, that can inform the analysis of quantitative datasets.

Methods

Engagement process

Eleven states participated in a four-phase, facilitated process over 12 months to build a model for a multi-state data infrastructure, uncovering state-specific needs as they worked the project (as of this writing, a third and fourth cohort, each consisting of 7 states, are engaging in the CMP).

States vary in data capacity and in policy structure (see Tab. 1). For example, most of the cohort 1 states have a decentralized model where curricular and course decisions are made at the district or school level with little influence from the state. In cohort 2, most of the states operate within a top-down approach where curriculum and graduation requirements are set at the state level.

Table 1: State data capacity at baseline

State features (at baseline)	**Cohort 1 (n=6)**	**Cohort 2 (n=5)**
State identified which courses count as CS at the high school (9–12) level	3	5
State has a common, statewide course numbering system	4	5
State agency uses SCED codes to define courses	4	1
Team has analyzed state level data in support of your BPC efforts	4	5
Disaggregated data available: gender (binary), race and ethnicity	5	5
Disaggregated data available: disability	4	4

For the purposes of the project, we limited our data focus to student-only data in stand-alone courses in grades 9–12. Teams learned from their colleagues, held each other accountable, and uncovered the opportunities and barriers within their unique data context. The process involved framing of the BPC goals of state-level change efforts, assessing current data systems for points of vulnerability and opportunity, data gathering, data visualization, data utilization, and analysis. The four phases are as follows:

Phase I: Onboarding

ECEP state teams begin learning about the CMP opportunity. State teams begin to form a team of 3–5 leaders who focus on common metrics work. Members who are adept in the areas of data gathering, data analysis, data reporting, and data utilization are ideal. This phase accounts for 3 months of the year-long project. This allows ample time for team development and completion of an initiation Data Approach Survey (DAS). The DAS helps teams to self-assess their purpose and capacity and to participate in a subsequent team interview to clarify and highlight notable findings in their survey.

Phase II: Initial Framing

The DAS results form the basis for the overall work of each cohort. Teams meet in person to develop common definitions of CS, discuss centering equity in the project, learn about the data systems in other states in the cohort, and review current data efforts from within states and across the U.S., as well as prior CMP cohorts. Teams begin to develop a comprehensive list of computing courses while considering how to build a common measurement system. The in-person meeting is highly interactive and includes sessions that address both the technical infrastructure and data ecosystem in the state, as well as sessions on subjectivity and bias in data reporting.

Phase III: Data Collection

Teams enter currently-available state data into a shared spreadsheet, so all teams in the cohort can see where data similarities exist. Data is only utilized within the project to develop a system for common metrics, allowing states to take risks and be more creative as they consider how to build a comprehensive, multi-state data project. An early look at visualizations for the states helps the teams consider where they may be able to improve the quality of their data.

Phase IV: Data Review & Dissemination

Being able to make the case for BPC efforts through data visualization and utilization is vital to BPC strategies in ECEP and across individual states. The CMP project helps states learn best practices in data dissemination in order to provide more stakeholders with more access to the data that will drive BPC-focused systems change efforts.

Contributing data:

To understand the process more fully and answer the research questions, we draw data from reflective team journals; meeting notes; and collaborative materials such as data templates, observations made during collaborative meetings, and the technical assistance requests made during the project.

Team Journals

Team journals are kept as live, shared digital documents to encourage collaboration among state teams. The journals consist of a set of reflective questions for each phase of the project. This serves to guide the state team in thinking through their own data approach and to help the CMP leadership guide the work of the cohort while improving the process for future cohorts. Example questions for Phase 1 include:

- How did the Data Approach Survey and interview support organizing and orienting your team to current capacities and potential outcomes?
- Please describe your experience of forming a team.
- What changes would you suggest to this phase that will improve scaling in other states?

Examples from the final CMP reflection section, which is answered at the conclusion of the project, include:

- How has collaboration across the cohort states supported and informed your equity-focused CSEd work?

- How will this project inform your future broadening participation in computing data collection and dissemination efforts?
- How has this project adjusted the strategies and actions you are taking to address the lack of diversity in K–12 CSEd pathways?
- What aspects of this project would you adjust or adapt to advance ECEP's ability to impact systems change within and across states?

Phase 2 and 3 questions are more specific to the data collection effort itself, for example "Please describe your experience of gathering the state based data. (e.g., ability to capture data from one system, having local data available at state level, school vs district granularity, what does your data network look like?).

Meeting materials

During our online and face-to-face meetings teams collaborate to create materials and contribute to shared notes documents. An example of materials generated is the "Data Ecosystem Map" that each team produces. This map is made in Jamboard, an online tool for collaborative visualization, or on posterboard, and traces data from point of entry at the local level through point of exit at the state level for a variety of reporting purposes. See Figure 2 for an example of a state data ecosystem map created during a meeting.

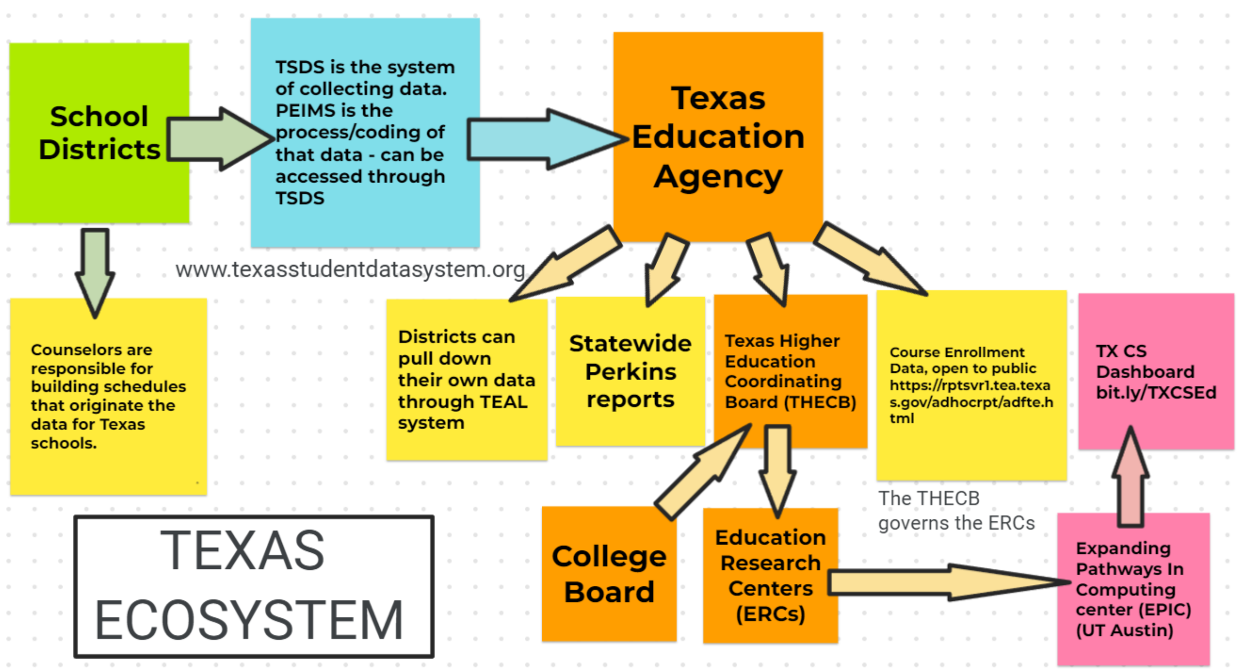

Figure 2: Texas Team's Data Ecosystem Map developed under this project

Technical Assistance Requests

In the third phase of the project, all teams are asked to complete the same data request. It is during this phase that assumptions are tested and many teams need additional clarification or support. For example, teams that assumed they had unbridled access to disaggregated data because the state collects and reports on it, have found that when making queries about CS specifically, they often trigger masking rules in which data is suppressed to protect student privacy (see discussion below on FERPA). Similarly, teams are sometimes surprised that data easily available in other subject areas is more ambiguous in the CSEd space due to inconsistent course names, descriptions, and how courses were coded at the point-of-entry into the data system. Understanding the needs of teams during this phase helps us understand the particulars of state data systems across the nation.

Lessons Learned: Components of a data infrastructure supporting equity-driven computer science education efforts.

Teams

Formation

When developing data infrastructure in support of BPC, teams matter. Teams need representation from people who can access data, understand the practical context of the data to support interpretation, and help tie data to broader advocacy efforts. In the first cohort we had some sense of what the team makeup should entail. There was already a commitment from state education agencies (SEAs), however we knew that a well-rounded team that included K–12 educators and administrators, as well as data leaders from local education agencies (LEAs), and state-level data leaders would enhance the team's ability to understand where data existed, how to access the data, and how to adapt the data for utilization in the final stages of the project. As presented in [23, Fig. 1], the number of team members and the diverse roles and perspectives among team members dramatically increased over the course of the project. Each new cohort has seen a slightly quicker team development process.

Ongoing engagement

Ongoing team engagement, both within and across states, allows participants to consider the complexity of issues and maintain high levels of engagement and commitment to the project. Maintaining commitment in an environment with competing educational priorities, and with all cohorts being facilitated during the COVID-19 pandemic, shows how relevant this project is to the needs of each state as well as the understanding that our work could provide a framework for national-level data efforts.

Establishing values a priori

Establishing the values guiding data work a priori ensures that data practices are ethical, responsible, and aligned with the goals and values of the group [25], [26]. ECEP centers equity in all BPC work across the Alliance. This is also an explicit goal of the CMP project, intended to ensure that data is not collected for the sake of gathering numbers but for the purpose of discovering which students do not have access to, or are engaging in, high quality computing. State teams are asked to establish their state's BPC goals as a condition of membership in the ECEP alliance; however, the CMP encourages teams to bring new people to the team, particularly representatives from the State Departments of Education and data leaders. The Orientation and Initial Framing phases of the work provide an opportunity to delve into how equity is centered. These phases provide context for the 'why' of the work they will engage in, and highlight who will benefit from the project and how the data will be used in support of BPC goals.

Questions that are central to the CMP project and focus the work around equity include: How do we center equity in data practices without being exclusionary?

1. What are the policies that are necessary for expanding CSEd access, participation, engagement, and inclusion? How do we leverage data to meet our desired outcomes/ends?
2. What should collaboration look like as we work towards equitable systems and outcomes in our BPC efforts?

The Data Approach Survey (DAS) is the first opportunity for teams to self-assess their purpose and capacity. During this time state teams are able to collectively set priorities and expectations. The DAS has also uncovered potential strengths and weaknesses in the data ecosystem. Having this information before collaborating with other states allows team members to take advantage of the learning community by positioning their own opportunities and limitations in relation to their peers. It also helps manage the expectations among the various state team members who often come to the CMP with different perspectives and motivations. For the CMP leadership team it helps with understanding the pressing issues including legislative opportunities (new CSEd coordinators, call for state CSEd plans, or funding allocated for CSEd).

The sessions during the Initial framing phase on subjectivity and bias and, relatedly, on engaging others with the data story are important in enabling the teams to understand in advance a) how they want to own their state BPC story and b) how others in the BPC arena make choices about how they present data in ways that may be counterproductive to the ECEP team's BPC advocacy work. These sessions are both cautionary and empowering for the teams.

Understanding the data ecosystem

Within the first 6 months of the first CMP cohort the research team designed a process that allows for a clear understanding of where data sits within a state, and how high school level data moves within a state. The models developed by CMP teams benefit the project overall and help states to both share their work and engage in models from other states. Each state is asked to draw a map of their data 'ecosystem,' essentially a lifecycle of K–12 CSEd data in their state. The figure below (Fig. 3) shows the framing for the exercise.

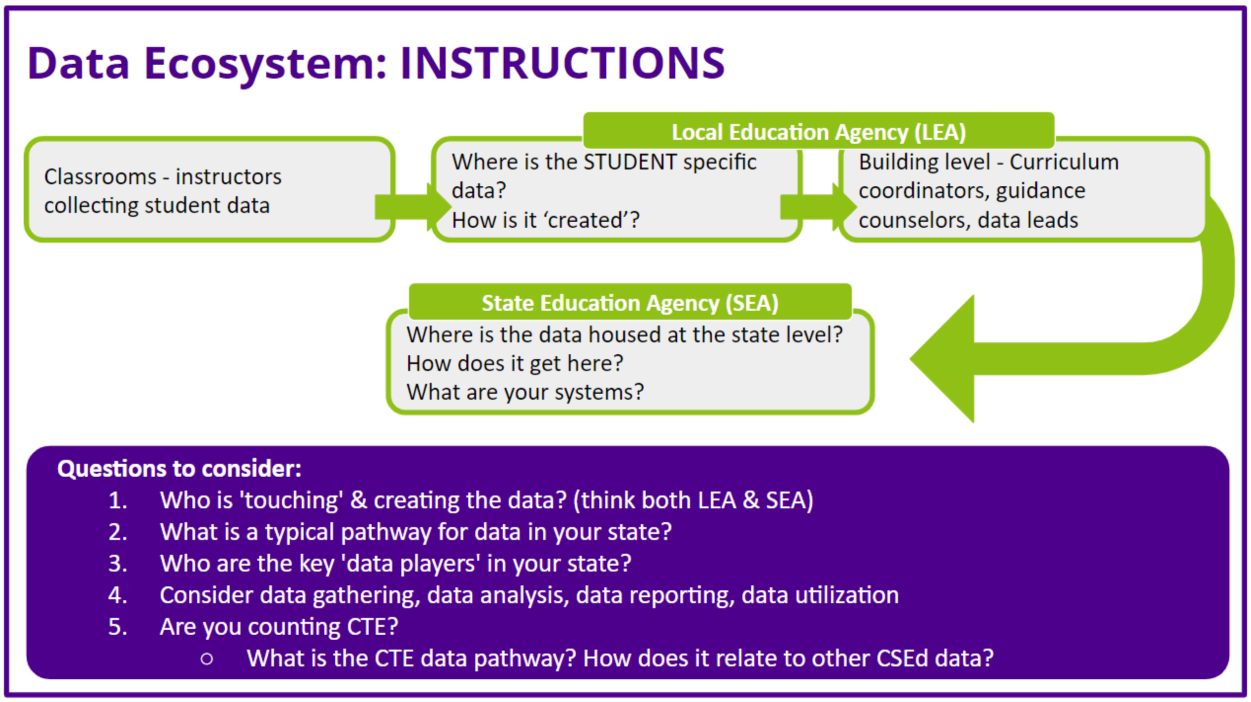

Figure 3: Data ecosystem map framing

While states describe pathways for data, various data collection tools, and named offices or positions responsible for data efforts, the process of visualizing these paths yields more details and more clarity around system strengths and weaknesses. The visualizations allow state teams as well as ECEP CMP facilitators to observe and discuss points of vulnerability, such as K–12 data entry positions at the building level that are often underfunded and experience a great deal of turnover. In some states it is surprising to see the number of data systems that are in place, but not necessarily connected in a manner that will allow teacher and student data to be paired or Career and Technical Education (CTE) data to be aligned with other CSEd data. The adage "junk in creates junk out" can be traced in these visualizations. The values of a state also surface through this exercise, for example whether states choose to collect demographic variables. In many cases, understanding the way data moves from classroom to state systems sheds light on the reporting schedule and when data will be available for research purposes. On the plus side, it gives states the opportunity to highlight where changes are needed and how supports can be developed that will create better, cleaner data.

Lessons Learned: Feasibility of creating cross-state data systems in support of the national BPC movement

Working across states

Working across states to develop a shared data system is challenging as all the state-specific issues identified above are compounded when developing a multi-state system. In 2018 ECEP attempted an early iteration of this work with the then 16 states and the territory of Puerto Rico. While the 2-day engagement at an ECEP Summit did yield an understanding of the vast differences in state data collection efforts, it didn't yield any common enough examples that would potentially scale across states. The shared data request was an Excel template in which states were asked to provide enrollment data at the state and course level for every CS course offered [23]. When making the official ask, teams uncovered assumptions they had about access to and quality of the data that may not have emerged during the data ecosystem exercise, including alignment of definitions and data suppression challenges.

Creating "common enough" definitions

Working across states requires shared definitions of CS. A facilitated exercise has the states explore the definitions of CS that they have utilized formally (in state policy), or informally (shared between districts, or in state-led professional development). Three commonly-used definitions[1] are shared and teams are asked to agree to one, or build a new common-enough definition. Each of the cohorts has had different responses to this exercise, with the first (and now also the third) cohort arriving at the definition in the K–12 Framework as common enough, and the second cohort continually revisiting the definitions. It became clear in the CMP that there are really two types of definitions. The first is a conceptual definition. This definition needs to be flexible enough that it can encompass any state definition and, more importantly, guide the operational definition which is created by identifying which courses in each state would be counted as CS. This clarity is important for bringing states to a "common enough" agreement on definitions and avoids the nuances between states, but also within states. Including both operational and conceptual definitions in the fourth cohort will provide us with more data on the

---

[1] Drawn from the K–12 Computer Science Framework [27], the US Department of Education Civil Rights Survey [28], and the Association for Computing Machinery K–12 Task Force Curriculum Committee [29].

effectiveness. The designation of common-enough is intentional and allows states freedom to explore within their own formal or informal definitions. It also acknowledges that in several states there is no codified definition of CS, or that the definition that the Department of Education uses may not align with that of CSEd advocates. For these states team members have expressed the need to not become locked into a definition while they are pursuing policy and CSEd strategies.

A conceptual definition of CS

The conceptual definition to help us understand collectively what CS is that has been generally agreed upon was taken from the ACM Task Force Curriculum Committee and reads "Computer science (CS) is the study of computers and algorithmic processes, including their principles, their hardware and software designs, their applications, and their impact on society" [29]. State teams have individually found that this definition generally works well enough, though some states would have liked to include additional information on what CS is NOT or examples of what CS is. It also lacks the specificity to be used as the sole determining factor in selecting which courses count (or do not count) as CS.

An operational definition of CS

The operational definition of CS, in which states have to identify which courses count as CS, ranges from easy for some states (particularly those that have CS standards in place) to nearly impossible for others. In many cases it isn't until states attempt to complete the data request itself that the challenges with identifying courses surface. To work across states courses are grouped into two types: CS and CS-Related. Those in the CS group are courses that state teams can confidently identify as CS within their states. CS-Related courses are those that *might* be CS; however, they are either not considered core CS courses within the state, not consistently taught across the state, or only exist in one state but not others within the cohort. Those that are most consistently grouped as "CS" include the two AP courses in CS. The courses for which there are the most open questions include web design, game development, cybersecurity, robotics, and database courses as these tend to focus either on specialized subfields within CS or be more focused on engineering applications. One method used to help clarify which courses should be grouped is to have teams categorize their courses within a shared spreadsheet. This allows teams to discuss with peers how and why they are categorizing their courses and also observe how other cohorts have worked with the categories.

The distinction between the *conceptual* and *operational* definitions became necessary after the second cohort, where how states defined CS created an insurmountable barrier to measuring CS across states. One state considered a CS course to need to include all five of the following concepts:

- Computing systems
- Neworks and the internet
- Data and analysis
- Algorithms and programming
- Impacts of computing.

With this approach, a course like Advanced Placement Computer Science - A, which is a programming-based course, would not “count” as CS, and yet it is one of only two CS courses that can be objectively and consistently measured across the nation. Ultimately, the cohort agreed to a set of courses that would be acceptable for the purposes of this project, while each state would need to retain their own unique definitions for internal policy efforts.

Depending on the state, adoption of SCED codes (School Course Exchange of Data), which are a nationally defined set of course codes, makes identifying the CS-related courses easier, though within a state there could be variability within districts for how they are applied. Some state teams have already participated in efforts to create crosswalks of their state-level course codes with SCED codes, creating the possibility of applying lessons from this approach across state teams.

Operationally, most states agree that courses that are most consistent across states include the following (though sometimes exact names might be different):

- Introduction to computing
- Computer literacy
- Advanced Placement Computer Science - A
- Advanced Placement Computer Science Principles
- Exploring Computer Science
- Computer Science Essentials
- Programming (I and II).

Those with more variation across states include:

- Cybersecurity related courses
- Robotics
- Web design
- Game development
- Networking
- Artificial Intelligence.

In some states they are clearly CS driven but in others they may not be considered as CS.

Perhaps ironically, although cohort 1 was more likely to use SCED codes than cohort 2, they were applied so inconsistently that many of the state teams questioned the quality of the data. The inconsistency was in part due to the versions of the SCED codes being used by states, which again varied greatly. Those in cohort 2 had more disparate course naming/numbering systems from each other but greater confidence that the course names/numbers were being applied uniformly across the state.

Defining CS and making the data request itself can have consequences. For states without a strong guidance for CS coming from within, those requesting data externally may inadvertently define CS for the state. For example, the annual State of Computer Science Education report [30] is an annual report on "the State of CS" in which data is requested from each state's Department of Education on participation in a set of computing education courses. This data request is defined by the report writing team, and does not necessarily reflect the specificities of how a particular state defines or conceptualizes computing. Although the report provides a clear methodology, the data is often used by state legislators to make decisions about resource allocations and educational priorities. If the operational definition used in the resulting analyses and visualizations does not align with the state team's BPC definition, it can lead to undesired or unanticipated policy implications. The ECEP CMP recognizes this tension (ECEP has contributed to the State of Computer Science Education report since 2018) and has intentionally designed the CMP to be a community-led initiative. ECEP doesn't assume expertise, and doesn't bring a top down approach to this work. Instead the project works with state teams in this year-long discovery project with the aim being a common data framework derived from state input.

Navigating the Family Educational Rights and Privacy Act (FERPA)

One of the most challenging parts of the data request is navigating each state's data suppression rules. States must comply with the Family Educational Rights and Privacy Act (FERPA), which regulates the disclosure of personally identifiable information that may include direct identifiers (name, addresses, ID numbers) indirect identifiers (date of birth, place of birth) or "other information that, alone or in combination, is linked or linkable to a specific student that would allow a reasonable person in the school community, who does not have personal knowledge of the relevant circumstances, to identify the student with reasonable certainty" [31]. No part of the data request includes direct or indirect identifiers, and so it is with this third clause regarding "other information" that states have to be concerned in terms of safeguarding student privacy.

FERPA leaves it to individual states to determine how to ensure data privacy under these guidelines. The act does not include any specific rules or requirements regarding how or when data must be suppressed. A more recent law, the Every Student Succeeds Act (ESSA) stipulates that each state must establish a minimum student subgroup size for which it will not publicly report assessment data. In conducting the data requests, however, we have found that many states adhere to blanket rules of suppressing any small group count, even when the counts are not related to assessment data. Although such rules make it easier to ensure data privacy, they have the ill effect in this project of leading to more data being suppressed than is necessary. For example, some states suppress small cell counts of basic enrollment information (e.g., number of students enrolled at a school) when the unsuppressed data is already publicly available through federal websites. To minimize the amount of data that has to be suppressed, some states combine the race categories with fewer students (i.e., American Indian or Alaska Native, Native Hawaiian or Pacific Islander, and two or more races) into one "other races" category. On one hand, this workaround allows for the data for those students to be included in analyses where it would otherwise be excluded due to data suppression. On the other hand, it removes the possibility of considering these racial subpopulations in the metrics being reported, effectively invisibilizing those populations in the analysis and report.

Specific examples from our Cohort 2 participants include:

- State_4 and State_5 suppress values less than 10, including "0."
- State_3 imputes values for small (<5) and complementary cell suppression, while zero values are not suppressed.
- State_1 and State_2 do not suppress any data at all, however once the dashboard was shared with the cohort, one of these states reported that their data needed to be suppressed and thought the research team would handle that on their behalf.

Data visualization

We seek to bring together the data from each state into a common reporting method by constructing data dashboards using Tableau software. We have chosen to use interactive data dashboards rather than traditional static reports for two primary reasons. Firstly, because the dashboards allow users to interact with the data, there is less of a need for a detailed narrative of what the data shows. Instead, users can interrogate the data on their own and generate their own conclusions, which, in turn, can help to generate meaningful discussion about both the message being communicated through the data and the way the data is reported. Secondly, data dashboards are live in the sense that they can be updated in real time as new data is made available or errors are corrected. This is important given the fact that we intend this project to be an iterative process of testing and improving what data is collected, how the data is collected, and how the data is visualized and reported in the dashboard.

Numerous issues have arisen as we attempt to combine data across states. Some states have had data available for multiple years while others have only had data for the latest school year or for just one school year that is not consistent with the other states. Similarly, some states have had data for certain fields in the dataset but not others. For example, among four states in cohort 2, State_2 was unable to provide counts of students with disabilities at the school level and State_1 was unable to provide counts of students with disabilities at the school or state level. These gaps in the data that occur with some states create challenges when it comes to using the data to make comparisons across states. However, the dashboard format of reporting the data has proved useful in working through these challenges because it allows us to create visualizations for particular metrics that can accommodate this missingness by including in the visualization whatever data is available based on whichever filters are being applied at a given time. For example, one graph shows the percent of female students enrolled in CS courses by state and school (Fig. 4). When the school year filter is set to "2020–21," each state is included in the graph except for State_1, which was only able to provide data for the 2018–19 school year. Changing the school year to "2018–19" then causes State_1's data to show up in the graph while also causing State_5's data to disappear since they did not provide data for that year.

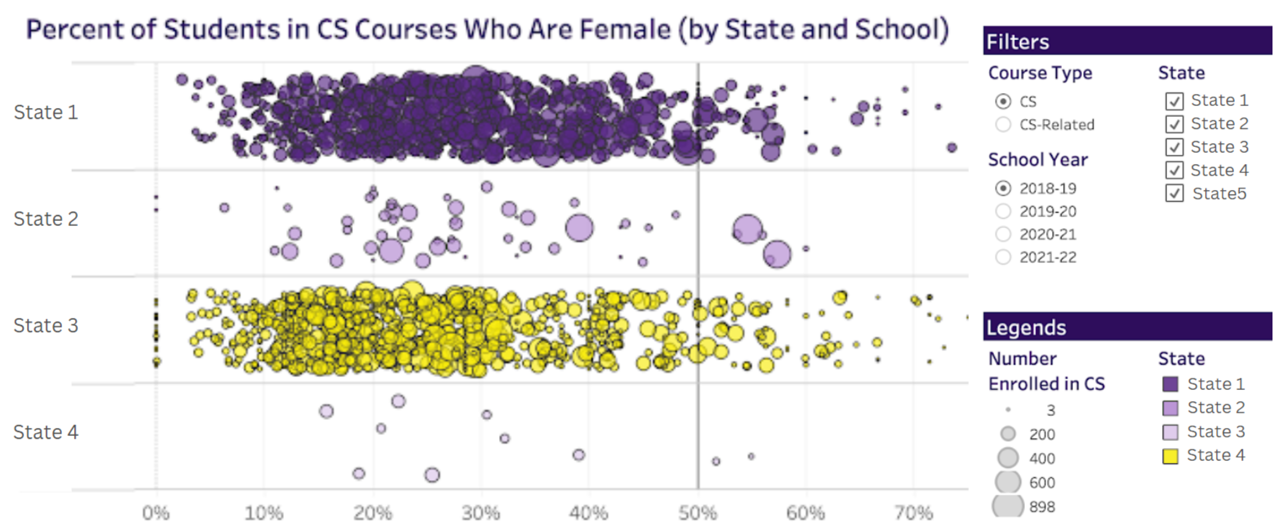

Figure 4: Example visual from the data dashboard

Other differences in the data between states are more problematic and not as easily resolved. For example, one state originally provided us with data on students' race and ethnicity using categories that were not mutually exclusive, which we had to exclude entirely from the dashboard until they were able to recompile their data and provide us with mutually exclusive counts. Other differences have no possibility of resolution. One state did not have the ability to count unique students in their enrollment numbers. They could count the number of enrollments for a given course, but they were unable to distinguish, for example, whether any of the students who enrolled in Computer Programming also enrolled in Web Development. Thus, their measure of the number of students who took one or more CS courses was based on total enrollments (i.e., it counted twice students who enrolled in more than one CS course) rather than total students like the other states. The only solution we had for this anomaly was to include a disclaimer in the dashboard that noted the different method of counting CS students for that particular state.

The data dashboards are beneficial in that they provide meaningful information in terms of the progress of broadening participation in computing both within each state and collectively across states. State teams appreciate the opportunity to use the dashboards to compare their state to others and to dig deeper into their own state's data by drilling down within the various visualizations to focus on specific districts, schools, or student subpopulations. State teams have also provided feedback on the dashboards themselves, making suggestions to improve their clarity and functionality. Some teams began making plans for how they could utilize the dashboards to advocate for policy within their states.

Discussion and Conclusions

States often use data systems outside of the intended design and face limitations when trying to surface inequity for populations based on gender, disability, ethnicity and race, which are exacerbated when looking intersectionally [26]. The politics within states around data access and use may also create barriers for state teams to navigate [32]. Using the CAPE framework allows CMP to overcome some of these challenges. Recognizing that no one individual is in a position to comprehensively define or understand data associated with the *capacity, access, participation* and *experience* related to offering high quality CSEd means a team approach is critical. Similarly, the framework helps teams think through the various elements of the unique data ecosystems within states and come to common-enough definitions by having a shared purpose across states.

Working across states is feasible but requires significant time and commitment to defining the purpose of the shared data collection and definitions being used. In the CMP project, states have already developed trust with the ECEP CMP leadership team and each other and have a demonstrated commitment to equity efforts. Without the commitment to collective impact [33], [34] across the states, garnered through the ECEP participation, teams may not have been as open to cross-state sharing.

As a result of this project, deeper questions are emerging as states look across their data. The ECEP community has a shared appreciation for the potential and limitations of their own state data systems and will work collectively to tackle some of these questions. Topics that have emerged include:

- How do we measure BPC in K–8 settings, or in courses where computing is integrated into another subject area?
- What can we learn about pathways? Do students who are exposed to a foundational CS course continue on to more advanced CS courses? This question is particularly important for states that have adopted a CS for All approach which requires CSEd in middle or early high school or as a graduation requirement.
- Does the classification of a CS course as a CTE or comprehensive course influence who can teach it, changing access for students? Does it change who participates?

States have also recognized that data is useful at all levels of advocacy, particularly when they're in control of defining the purpose, analysis and use.

Going beyond state data

A call for better evidence and improved state data reporting systems is important if BPC efforts are going to be realized in U.S. CSEd classrooms, and systems change strategies. Although state data has been helpful for obtaining BPC progress in recent years, the underrepresentation of minoritized students in CSEd continues to be an ongoing wicked problem [35]. Since the beginning of the 21st century we have seen an abundance of data emerge outside of traditional governmental agencies and institutions that historically had a monopoly on data. This publicly available data has called for actors outside of the formal government hierarchy to analyze and implement policies that will advance equitable practices in various social sectors, including CSEd. Originally designed for reporting, state data systems are not intended for social science research and investigating intersectionality of theories and concepts [36]. Over the past decade, state education departments and legislatures have implemented tighter policies to protect student-level data. Therefore, practitioners, researchers, and policymakers now must either link or access data outside of state data systems. Large education-related datasets have grown over the years, which has offered new opportunities to leverage data to address questions of policy and practice. Finally, going beyond state data provides stakeholders the ability to critically examine questions that center around race, class, and gender, and how issues of CSEd intersect with other social sector phenomena (i.e. transportation, housing, health & wellness, etc.).

Acknowledgements

This material is based upon work supported by the U.S. National Science Foundation under Grant No. 2137834. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. We also thank the states that participated in the common metrics process with openness and transparency.

References

[1] R. Zarch and S. Dunton, “Looking back to move forward: Measuring K–12 computer science education requires an equity explicit perspective,” presented at the Conference on Research on Equity and Sustained Participation in Engineering, Computing, and Technology (RESPECT), Philadelphia, PA, USA, May 23–24, 2022.

[2] J. Kurose, “Dear Colleague letter: Pursuing meaningful actions in support of Broadening Participation in Computing (BPC) (NSF 17-110).” nsf.gov. https://www.nsf.gov/pubs/2017/nsf17110/nsf17110.jsp (accessed Feb. 13, 2023).

[3] J. Margolis, R. Estrella, J. Goode, J. Jellison Holme, and K. Nao, *Stuck in the Shallow End: Education, Race, and Computing*. Cambridge, MA: MIT Press, 2008.

[4] C.L. Fletcher and J.R. Warner, “CAPE: A framework for assessing equity throughout the computer science education ecosystem,” *Communications of the ACM*, vol. 64, no. 2, pp. 23–25, Feb. 2021, doi: 10.1145/3442373.

[5] L. Goodyear, G. Silverstein, and L. P. Thurston, “The influence and promise of alliances,” *Communications of the ACM*, vol. 60, no. 6, pp. 29–30, 2017.

[6] D. Chubin, B. Bizot, T. McKlin, and A Peterfreund, “NSF’s BPC alliances program – A report on common core indicators of breadth and depth of participation,” NSF, USA, Rep. Aug. 17, 2012.

[7] National Science Foundation, “Broadening Participation in Computing (BPC) (NSF 21-571).” nsf.gov. https://www.nsf.gov/pubs/2021/nsf21571/nsf21571.htm (accessed Feb. 9, 2023).

[8] M. Smith, “Computer Science for All,” *The White House, President Barack Obama*, Jan. 30, 2016.

[9] D. E. Chubin and R. Y. Johnson, “A program greater than the sum of its parts: The BPC alliances,” *Communications of the ACM*, vol. 54, no. 3, pp. 35–37, 2011.

[10] R. Zarch, S. Dunton, J. Childs, and A. Leftwich, “Through the looking glass: Computer science education and the unintended consequences of broadening participation policy efforts,” in *2020 Research on Equity and Sustained Participation in Engineering,*

Computing, and Technology (RESPECT), Portland, OR, USA, Mar. 11, 2020, pp. 169–172.

[11] E. A. Locke and G. P. Latham, *A Theory of Goal Setting & Task Performance*. Englewood Cliffs, NJ: Prentice-Hall, 1990.

[12] M. Giovannini, “What gets measured gets done,” *Journal for Quality & Participation*, vol. 24, no. 4, pp. 21–27, 2004.

[13] Code.org, “Does your school teach Computer Science?” code.org. https://code.org/yourschool#form (accessed Dec. 22, 2020).

[14] Google LLC and Gallup Inc. “Searching for computer science: Access and barriers in U.S. K–12 education,” Rep. 2015. Available: https://services.google.com/fh/files/misc/searching-for-computer-science_report.pdf

[15] E. Banilower and L. Craven, “Factors associated with high-quality computer science instruction: Data from a nationally representative sample of high school teachers,” in *SIGCSE ‘20: Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, Portland, OR, USA, Mar. 2020, pp. 360–365.

[16] A. Decker, M. M. McGill, J. Ravitz, E. Snow, and R. Zarch, “Connecting evaluation and computing education research: Why is it so important?,” in *SIGCSE ‘18: Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, Baltimore, MD, USA, Feb. 2018, p. 818, doi: 10.1145/3159450.3159642.

[17] A. Decker *et al*, “Repositories you shouldn’t be living without,” in *SIGCSE ‘18: Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, Baltimore, MD, USA, Feb. 2018, pp. 920–921.

[18] J. Xavier, M. M. McGill, and A. Decker, “Designing and developing a resource center for primary and secondary computing education researchers,” in *2020 IEEE Frontiers in Education Conference (FIE)*, Uppsala, Sweden, Oct. 2020, doi: 10.1109/FIE44824.2020.9274252.

[19] J. Margolis and J. Goode, “Ten lessons for computer science for all,” *ACM Inroads*, vol. 7, no. 4, pp. 52–56, 2016.

[20] J. Margolis, J. Goode, and G. Chapman, "An equity lens for scaling: A critical juncture for exploring computer science," *ACM Inroads*, vol. 6, no. 3, pp. 58–66, 2015.

[21] ECEP. "Expanding Computing Education Pathways (ECEP) alliance: Broadening participation in computing state by state." ecepalliance.org. https://ecepalliance.org (accessed Feb. 8, 2023).

[22] ECEP. "State data dashboards." ecepalliance.org. https://ecepalliance.org/cs-data/state-data-dashboards/ (accessed Feb. 8, 2023).

[23] S. Dunton, R. Zarch, J. Xavier, J. Warner, and A. Peterfreund, "Determining metrics for broadening participation in computing: Connecting data to multi-state computer science education policy efforts," *Policy Futures in Education*, Jan. 2022, doi: 10.1177/14782103211064443.

[24] A. T. Ottenbreit-Leftwich et al., "How to change a state: Broadening participation in K–12 computer science education," *Policy Futures in Education*, Sep. 2022, doi: 10.1177/14782103221123363.

[25] National Academies of Sciences, Engineering, and Medicine, *Monitoring Educational Equity*. Washington, DC: The National Academies Press, 2019, doi: 10.17226/25389.

[26] *IEEE Standard Model Process for Addressing Ethical Concerns during System Design*, IEEE Standard 7000-2021, Sept. 15, 2021. doi: 10.1109/IEEESTD.2021.9536679.

[27] "K–12 computer science framework." k12cs.org. https://k12cs.org/ (accessed Feb. 16, 2023).

[28] Office for Civil Rights, "2020–21 Civil Rights Data Collection — School Form," U. S. Department of Education, p. 22. Available: https://www2.ed.gov/about/offices/list/ocr/docs/2020-21-crdc-school-form.pdf.

[29] A. Tucker, "A model curriculum for K–12 computer science: Final report of the ACM K–12 task force curriculum committee," Association for Computing Machinery, New York, NY, USA, Tech. Rep. 2003.

[30] Code.org, CSTA, and ECEP Alliance, “2022 state of computer science education: understanding our national imperative,” Rep. 2022. Available: https://advocacy.code.org/stateofcs.

[31] Privacy Technical Assistance Center, “Frequently asked questions — Disclosure avoidance,” 2013. Available: https://studentprivacy.ed.gov/sites/default/files/resource_document/file/FAQs_disclosure_avoidance.pdf.

[32] J. Fensterwald and T. Peele. “State delays public release of English, math and science test score results to later this year.” EdSource.org. https://edsource.org/2022/state-delays-public-release-of-english-math-and-science-test-score-results-to-later-this-year/678462 (accessed April 10, 2023).

[33] J. Kania, F. Hanleybrown, and J. S. Juster, “Essential mindset shifts for collective impact,” *Stanford Social Innovation Review*, vol. 12, no. 4, pp. A2–A5, 2014, doi: 10.48558/VV1R-C414.

[34] J. E. Cuny, J. Payton, A. Gates, C. L. Fletcher, and A. Peterfreund, “National CS ed equity-focused consortia and their value to the community,” in *SIGCSE ‘20: Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, Feb. 2020, pp. 1116–1117, doi: 10.1145/3328778.3366982.

[35] J. Childs and R. Lofton, “Masking attendance: How education policy distracts from the wicked problem(s) of chronic absenteeism,” *Educational Policy*, vol. 35, no. 2, pp. 213–234, 2021.

[36] C. Conaway, V. Keesler, and N. Schwartz, “What research do state education agencies really need? The promise and limitations of state longitudinal data systems,” *Educational Evaluation and Policy Analysis*, vol. 37, no. 1S, pp. 16S–28S, 2015.