Technological University Dublin
ARROW@TU Dublin

Conference Papers | School of Science and Computing

2019

An international study piloting the measuring teacher enacted computing curriculum (METRECC) instrument

Katrina Falkner
University of Adelaide

Sue Sentance
Raspberry Pi Foundation

Rebecca Vivian
University of Adelaide

See next page for additional authors

Follow this and additional works at: https://arrow.tudublin.ie/ittscicon

 Part of the Computer Sciences Commons

Recommended Citation

Katrina Falkner, Sue Sentance, Rebecca Vivian, Sarah Barksdale, Leonard Busuttil, Elizabeth Cole, Christine Liebe, Francesco Maiorana, Monica M. McGill, and Keith Quille. 2019. An International Study Piloting the MEasuring TeacheR Enacted Computing Curriculum (METRECC) Instrument. In Proceedings of the Working Group Reports on Innovation and Technology in Computer Science Education (ITiCSE-WGR '19). Association for Computing Machinery, New York, NY, USA, 111–142. DOI: 10.1145/3344429.3372505

This Conference Paper is brought to you for free and open access by the School of Science and Computing at ARROW@TU Dublin. It has been accepted for inclusion in Conference Papers by an authorized administrator of ARROW@TU Dublin. For more information, please contact arrow.admin@tudublin.ie, aisling.coyne@tudublin.ie, gerard.connolly@tudublin.ie, vera.kilshaw@tudublin.ie.

Authors

Katrina Falkner, Sue Sentance, Rebecca Vivian, Sarah Barksdale, Leonard Busuttil, Elizabeth Cole, Christine Liebe, Francesco Maiorana, Monica M. McGill, and Keith Quille

This conference paper is available at ARROW@TU Dublin: https://arrow.tudublin.ie/ittscicon/25

An International Study Piloting the MEasuring TeacheR Enacted Computing Curriculum (METRECC) Instrument

Katrina Falkner*
School of Computer Science
The University of Adelaide
Adelaide, South Australia, Australia
katrina.falkner@adelaide.edu.au

Sue Sentance*
Raspberry Pi Foundation
Cambridge, England, UK
sue@raspberrypi.org

Rebecca Vivian*
School of Computer Science
The University of Adelaide
Adelaide, South Australia, Australia
rebecca.vivian@adelaide.edu.au

Sarah Barksdale
Department of Curriculum and Instruction
University of Minnesota
Minneapolis, Minnesota, US
barks016@umn.edu

Leonard Busuttil
Department of Technology and Entrepreneurship Education
University of Malta
Msida, Malta
leonard.busuttil@um.edu.mt

Elizabeth Cole
School of Computer Science
University of Glasgow
Glasgow, UK
e.cole.2@research.gla.ac.uk

Christine Liebe
Colorado School of Mines
Golden, Colorado, US
cliebe@mines.edu

Francesco Maiorana†
Department of Computer Science
Kansas State University
Kansas, Manhattan, US
fmaiorana@ksu.edu

Monica M. McGill
Department of Computer Science
Knox College
Galesburg, Illinois, US
mmmcgill@knox.edu

Keith Quille
Department of Computing
TU Dublin
Tallaght, Dublin, Ireland
keith.quille@tudublin.ie

ABSTRACT

As the discipline of K-12 computer science (CS) education evolves, international comparisons of curriculum and teaching provide valuable information for policymakers and educators. Previous academic analyses of K-12 CS intended and enacted curriculum has been conducted via curriculum analyses, country reports, experience reports, and case studies, with K-12 CS comparisons distinctly lacking teacher input.

This report presents the process of an international Working Group to develop, pilot, review and test validity and reliability of the MEasuring TeacheR Enacted Computing Curriculum (METRECC) instrument to survey teachers in K-12 schools about their implementation of CS curriculum to understand pedagogy, practice, resources and experiences in classrooms around the world. The Working Group reviewed and analysed pilot data from 244 teachers across seven countries (Australia, England, Ireland, Italy, Malta, Scotland and the United States). We analysed the pilot results (n=244) and applied four validity tests: face validity, concurrent validity, population validity, sampling validity and construct validity, in addition to a focus group to further revised the instrument.

This report presents the pilot results and outcomes of validity testing, as well as revisions made to the instrument. The resulting METRECC tool combines a country report template and a teacher survey that will provide K-12 teachers with a means to communicate their experience enacting CS curriculum. National and regional policymakers can use METRECC data to inform iterative curriculum revision and implementation. We provide open access to the METRECC instrument and data set.

CCS CONCEPTS

• **Social and professional topics → Computing education**; *K-12 education*;

KEYWORDS

ITiCSE Working Group, K-12, Computer Science Education, Informatics Education, Schools, Survey

ACM Reference Format:
Katrina Falkner, Sue Sentance*, Rebecca Vivian*, Sarah Barksdale, Leonard Busuttil, Elizabeth Cole, Christine Liebe, Francesco Maiorana, Monica M.

*Co-Leader
†Also with Universitá di Urbino Carlo Bo, Italy.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
ITiCSE-WGR '19, July 15–17, 2019, Aberdeen, Scotland Uk
© 2019 Association for Computing Machinery.
ACM ISBN 978-1-4503-7567-2/19/07…$15.00
https://doi.org/10.1145/3344429.3372505

McGill, and Keith Quille. 2019. An International Study Piloting the MEasuring TeacheR Enacted Computing Curriculum (METRECC) Instrument. In *2019 ITiCSE Working Group Reports (ITiCSE-WGR '19), July 15–17, 2019, Aberdeen, Scotland Uk*. ACM, New York, NY, USA, 32 pages. https://doi.org/10.1145/3344429.3372505

1 INTRODUCTION

New primary and secondary school Computer Science (CS) curricula have recently been introduced to a number of countries (e.g. [5, 11, 39, 41, 45, 50, 78, 99]), leading to a number of national and international efforts to develop, support and evaluate curriculum development. While K-12 CS curricula is well entrenched in some countries [3, 34, 52], it is a relatively new phenomenon in many. This poses challenges not only in the development and implementation of new curricula, but also in preparing and supporting teachers as they transition from initial teacher qualifications and experience in other learning areas to the teaching of computing [19]. Significant effort is being undertaken to understand the learning and teaching of computing in the K-12 space, as well as how we are able to build teacher capacity in teaching this subject that is new to many, with a need to support teachers with varying degrees of background skill, confidence, perceptions and knowledge.

To help support evaluation, comparison and reflection on these initiatives, a number of country and regional reports have been produced with the aim of identifying and describing in a comparable, standardised way, the intended curriculum, defined by Porter and Smithson [85] as "such policy tools as curriculum standards, frameworks, or guidelines that outline the curriculum teachers are expected to deliver". Gander [54] and Balanskat and Engelhardt [6] have explored K-12 CS curriculum initiatives across Europe, while several reports have been undertaken for initiatives in the UK [105, 106], the US [57], Wales [76] and Poland [104]. The ACM Transactions on Computing Education (TOCE) dedicated two volumes for a special issue that covered a range of countries as case studies in K-12 CS education (see Vol. 14, no. 2, 2014) [59].

Beyond the intended curriculum, there is a need to understand what curricula content is actually taught in classrooms, and to explore more deeply the pedagogical practices, approaches and tools that teachers utilise to deliver curricula. The enacted curriculum, defined by Porter and Smithson [85] as "actual curricular content that students engage in the classroom", captures what content is being delivered within the classroom, as well as the pedagogical approaches adopted, and - with particular relevance to CS curriculum - the use of technology, physical computing devices and tools used. These descriptions of classroom practice enable us to better understand the alignment of what is actually happening within our classrooms to what we intend our curricula to be, and to identify where resources and support can be most effectively targeted and are most urgently needed.

Related to this Working Group topic, in 2011 another Working Group [61] undertook the process of collecting and evaluating research findings about secondary CS curriculum from different countries, and in the process developed a category system (Darmstadt Model) to support future research activities and comparison of results across regional and national boundaries. Expanding this work, a 2015 Working Group applied the Darmstadt Model to analyse, compare and extract insights from the articles published within two K-12 CS education special issues for TOCE [60]. This work sought to understand CS curriculum topics taught in schools, goals and competencies, programming languages and tools adopted, assessment practices and teacher training; however, the authors acknowledged that a limitation of the study was that it was restricted to the analysis of selected journal publications.

In 2013, a Working Group formed to investigate trends of CS as a subject in schools by inviting CS education and teaching professionals worldwide to complete an online questionnaire about the current state of K-12 CS curriculum in their country [98]. Experts from 22 countries responded, addressing CS curriculum topics and goals covered across K-12 as well as teaching methods; however, a limitation is that results were based on a small group of experts, from across a range of settings (university, school and industry). Participants were invited to reflect on their curriculum and contexts and rank items, such as CS topics and pedagogy, in order of importance. This work interestingly sheds light on CS curriculum in various contexts but relies on personal judgements rather than descriptive information about education systems and intended CS curriculum. With the rate of K-12 CS education evolving rapidly, it is important to find ways to easily capture high-level descriptive information that can be used to track changes over time, with experiences and approaches to implementing enacted curriculum coming directly from the teachers who teach content in classrooms.

Prior work has set a strong foundation for understanding the state of K-12 CS curriculum and implementation efforts from the perspective of experts and academics. However, there is an opportunity to further expand this work to focus on what K-12 teachers are doing in classrooms, capturing their input about enacted curriculum and their experiences via self-report measures. Additionally, there is a need to align work at the broader intended curriculum-level with teachers' self-reports of enacted curriculum in classrooms. With the substantial investment across many countries in teacher professional development, we are also interested in the connection between teacher professional development and enacted curriculum. By understanding how professional development resources are being used by teachers to support their curriculum implementation and classroom practice, we are better able to refine and direct future resources to specific needs.

This Working Group is an initiation of a collective effort for a deeper but scalable investigation into what is happening in schools, based on the input of educators in classrooms. The project aims to understand K-12 teachers' experiences and approaches to CS curriculum implementation in the classroom, including demographics, self-esteem, teaching methods, assessments, use of resources, professional development, and curriculum topics implemented. The project involved developing instruments to capture and measure K-12 CS enacted and intended curriculum for countries across the world. This Working Group presents the development of a survey instrument for K-12 teachers to measure the enacted curriculum and a template to capture the intended curriculum of countries. The Working Group piloted the survey across seven countries (Australia, England, Ireland, Italy, Malta, Scotland and United States) and describe the process undertaken to evaluate the instruments and measure evidence of reliability and validity of the teacher survey.

Duncan and Bell [43] compared English language curriculum for students from ages five to fourteen in England, Australia, and in the

United States. Building on previous work and analysis in English, we chose to create the survey in English. Once future iterations of the survey are deployed and analysed, it may be prudent to develop the survey in other languages. In responding to the observation by Ihantola et al. [62] that there is a critical need for data to be shared for re-validation, all data for this pilot study along with study instruments are available online (see Appendix A). This also enables those in other countries to take the instruments and adapt them to their native language.

1.1 Structure of Report

The focus of our Working Group was to construct a survey instrument with evidence of reliability and validity to capture and contrast national implementation of CS curriculum. As our goal is to design a generalisable survey instrument that can be adopted easily across countries and regions, we begin by defining a common language around CS curriculum, which is outlined in Section 2.

In Section 3 we describe the related work, specifically looking at literature related to curriculum implementation and measurement of intended and enacted curriculum. In this process, we review related survey work within K-12 CS curriculum implementation, and identify specific survey concerns and opportunities in capturing enacted curriculum.

Section 4 defines our broader Working Group objectives and research questions, followed by Section 5 that describes the methods, including the process of developing the study instruments (both the country report capturing the intended curriculum and the teacher survey capturing enacted curriculum). Additionally, we describe the pilot survey study data collection and analyses.

In Section 6 we present the descriptive findings of our pilot study results, followed by the process and results associated with checking the instrument for evidence of reliability and validity using the pilot data (Section 7). In Section 8, we report insights from our evaluation processes that have led to recommended changes in the design of the instruments, followed by a discussion of pilot study obserbations, lessons learned and recommendations for future use of the instruments in Section 9. The report concludes with a discussion around opportunities for future research (Section 10).

2 DEFINITIONS

Our purpose in this study and the development of our instruments was not to strictly define Computational Thinking, programming, or to take a stance on the role of the computer, but rather to provide an instrument for many contexts to use. We specifically want to have openness in interpretation of the survey to allow it to be applicable and understandable in many contexts. This does pose limitations to comparisons between countries, but also allows for local contextualisation to be layered. We aimed to have openness in interpretation so that survey administrators could interpret findings in their particular context.

To provide consistency and operationalise terminology in this report, we present key definitions that describe the essence of broader concepts, such as intended and enacted curriculum or computational thinking, as well as provide guidance for more specific terminology used throughout.

Computer Science (CS) The study of computers and algorithmic processes, including their principles, hardware and software designs, applications, and their impact on society. Many countries use the term computer science [61], however, it is also referred to as "Computing Science", "Computing", "Informatics" and "Digital Technologies".

Computing Any goal-oriented activity requiring, benefiting from, or creating algorithmic processes[38].

Computational Thinking While there is much debate on the exact definition of computational thinking, the following concepts are often agreed upon in the literature and comprise the definition for the purpose of this paper [33]: A way of thinking when computing that uses decomposition, pattern recognition, abstraction, pattern generalisation, and algorithm design to solve problems. It is commonly used in computer science, but it is applicable to many everyday problems, too. It allows us to take a complex problem, understand the problem better by using a computational framework, and develop possible solutions. We can then present these solutions in a way that a computer, a human, or both, can understand [100, 121].

Curriculum (Standards) In many countries the term curriculum is a high-level concept relating to specific learning objectives and measurable outcomes or benchmarks for learning levels. In the US, standards are synonymous with curriculum as defined previously. Educators in the US refer to curriculum as well-articulated bodies of courses, modules, and lesson plans. When educators from outside the US use the term curriculum, people from the US can understand their meaning as a CS framework or standards.

Digital Literacy (Computer and Information Literacy, Technological Literacy) The ability to use digital technology, communication tools or networks to locate, evaluate, use, and create information [77]. Digital literacy is the ability to understand and use information in multiple formats from a wide range of sources when it is presented via computers [56]. A person's ability to perform tasks effectively in a digital environment. Literacy includes the ability to read and interpret media, reproduce data and images through digital manipulation, and evaluate and apply new knowledge gained from digital environments [64]. Other references to the term include digital literacy, technological literacy, computer literacy and information literacy. In all instances, the focus of the terms is on the use of technology and devices, not the creation of solutions applying computing.

Enacted Curriculum Actual curricular content taught by teachers that students engage with in the classroom [85].

Experience Any aspect of teachers' lived perceptions.

Intended Curriculum Policy tools as curriculum standards, frameworks, or guidelines that outline the curriculum teachers are expected to deliver [85].

Pedagogy Pedagogy commonly is defined as the intentional activity of one person influencing learning in another. Instructional strategies are methods teachers use to execute pedagogy. Pedagogy is the experience and communication exchange between teachers and students [118].

Practice The activities that support successful pedagogy and teaching improvement such as planning, record-keeping, questioning, developing relationships, creating assessments, and reflecting on teaching. Practice includes pedagogy as a broader term [123].

Programming The practice of reading, writing, and analysing computer code using text-based and block-based (i.e., drag and drop coding) general computer languages to implement a plan/algorithm including debugging [43].

Professional Development The training offered to pre-service or in-service teachers. Professional development options often include university or college courses, workshops, hybrid training, and online courses.

Resources A term that refers broadly to purchased or free (open source) materials online providing lesson plans, modules, assessments, videos and other teaching aides. In some locations, such as the US, resources are referred to broadly as curriculum and/or instructional materials.

3 RELATED WORK

In this section, we start with a theoretical discussion by what we mean in terms of curriculum and particularly enacted and intended curriculum. We present related work on efforts monitoring and scoping K-12 CS education and implementation efforts that this study has built upon.

3.1 Curriculum Implementation

Curriculum is defined broadly as the academic content to be delivered in a specific course or program, with a more expansive definition including the learning objectives, the content topics to be delivered, measurable outcomes, the units that are to be taught, assessments used to measure and guide learning and materials providing support and resourcing.

There is a difference between the intended curriculum, defined by relevant standards, and the enacted or implemented curriculum, which is taught by teachers in the classroom [85]. Nolet and McLaughlin [81] define the enacted curriculum as the operationalisation of intended curriculum, embodying the decisions teachers make in terms of what actually is taught, and how. In Porter and Smithson's model [85], the intended and enacted curriculum are further elaborated through descriptions of the assessed and the learned curriculum. Remillard [92] separates this broader view of enacted curriculum into the teacher intended curriculum, which encapsulates teacher beliefs and knowledge of pedagogy and curriculum access to resources, understanding of student needs and local context, and the enacted curriculum, which is student and teacher beliefs, and access to resources. In this model, the intended curriculum represents only standards and policy definitions. According to Van den Akker et al. [113], language learning is influenced in part by an intended ideal curriculum (standards) and the enacted, or implemented operational curriculum, through teachers' aims and objectives, content, and learning activities.

Large scale survey results, such as in Trends in International Mathematics and Science Study (TIMSS), recognise that textbooks are the mediators between "the intent of curricular policy and the instruction that occurs in the classroom" [83, 112]. This notion can be extended to all types of resources developed by researchers, teachers, organisations and educators. These resources can be the mediators to facilitate the implementation of the intended curriculum or could arise in the field from the process of teaching enacted curricula. This approach has benefited the introduction of CS in countries where there is not a clear indication of intended curricula, however also poses challenges in terms of measuring what teachers view as enacted curricula. Regarding Europe, there is a movement to introduce Informatics from primary school [22] and to achieve this goal there is a strong effort to integrate Informatics inside the teaching of all school subjects by means of supporting teachers' creative approach as well as support from non-profit organisations such as "Coder Dojo" and initiatives led by the national research agency and Ministry in Italy and Europe who create CS education resources [10, 12, 51, 66, 70, 71, 80].

Curriculum change, either through curriculum reform or the introduction of a new curriculum, poses many challenges and may take several years for full implementation to occur, particularly with alignment between enacted and intended curriculum. Broadly, teachers have identified lack of resources and time as key obstacles in implementation of a new curriculum [23], lack of theoretical and/or technical knowledge [68], addition to the complexity of developing a clear understanding of curriculum standards [96].

We have observed these challenges in the recent rise of CS curriculum implementation efforts. There is a common confusion amongst teachers [116], but also present in government and school leadership [11, 18], regarding the distinction between Information and Communications Technology (ICT) literacy and CS, clearly indicating complexity in understanding curriculum standards. The required degree of technological awareness in CS curriculum is a further challenge, with teachers' lack of confidence and familiarity with CS tools and physical devices leading to deviation from lesson plans [75]. Black et al. [16] describe a related experience, with early adopters focusing more on fun activities, engaging with impressive technology or physical computing devices, rather than providing opportunities for deep learning of computational thinking. Additionally, the plethora of free scripted lesson plans allow CS teachers to disengage from offering intentional pedagogy. Teachers unfamiliar with technology, national curriculum or standards may sacrifice their agency as a teacher assuming that by duly following prescribed lesson plans or modules they are teaching CS. Thus, an instrument to elucidate teacher confusion about intended and enacted curriculum is warranted.

The choice of programming language to be used within any CS curriculum is a further complication, with many curriculum standards silent on programming language, and paradigm choice. This flexibility aids teachers in that they are able to identify what best suits their immediate context, however this poses challenges in terms of development of suitable resources, assessment and professional development, as well as increased expectations on teacher capability. Bell et al. [11] provide one example of the complexity of this single point of curriculum enactment within the New Zealand curriculum implementation.

It is useful to understand and study enacted curriculum in that it assists in identifying key areas for future resourcing, as well as potential problematic aspects of intended curriculum. Targeting resources and support can increase alignment between intended

and enacted curriculum. Gamoran et al. [53] and Kurz [67] identify correlation between the alignment of intended and enacted curriculum with student performance. Gamoran et al. study this alignment in the context of transition Mathematics courses, analysing the performance of n=882 students across 42 courses, identifying that curriculum alignment accounted for the majority of achievement difference between the courses. Careful analysis of the relationship of the intended and enacted curriculum can inform the efforts of professional development providers and preservice teacher programs. Ultimately, the conscious monitoring of efforts from intended curriculum stakeholders and teachers enacting curriculum can facilitate robust K-12 CS education.

Furthermore, there is a unique opportunity afforded through the study of enacted curriculum within the K-12 CS curriculum space. Larke's [68] qualitative study looks at how educators in England are (not) meeting the "actual policy" (or the intended curriculum: all policy documents distributed from officials/government, like curriculum) goals of the computing curriculum by studying how educators interpret and enact the curriculum, or "policy-in-use." Her conclusion is that teachers are the gatekeeper to computing education as they choose to interpret and/or reject the intended curriculum for a variety of reasons, including lack of training, experience, resources, and time needed to teach. Due to its infancy, there is much that we do not understand as yet about K-12 CS pedagogical practice and suitability of tools, programming languages, and physical computing. Thus, an instrument that can assist in the evaluation of the intended and enacted curriculum longitudinally and internationally will speed the development of K-12 CS pedagogy.

In this report, we view enacted curriculum as encompassing both Remillard's [92] view of enacted curriculum and the teacher intended curriculum [85] together. We offer the METRECC survey as an instrument that can help obtain descriptive comparisons internationally of K-12 CS intended and enacted curriculum. The proposed METRECC instrument is a tool that can assist CS researchers and CS teacher practitioners in achieving a common goal - improving K-12 CS education. Teachers can share their experience and knowledge of the classroom by reflecting on their enacted curriculum, which can then be used to further inform and refine the intended curriculum, pedagogy, research and policy directions. In this way, teachers who participate in the study gain agency and identity as K-12 CS pedagogical experts.

3.2 Measuring K-12 CS Intended Curriculum

In this section, we will briefly review papers that have used existing survey instruments to collect CS landscape information. To identify relevant research, a systematic process was adopted to identify relevant work published on Scopus, Web of Science, Google Scholar and EBESCO, restricting the search to the last six years. Search phrases included terms such as "Computing", "Computer Science", "Informatics", "Intended", "Enacted", "Curriculum", "K-12", "Recommendations", "Survey", "Review", "Comparison", "National" and "International". The search results were independently analysed by the three working group members who reviewed the abstract of relevant publications, and, if deemed applicable to the scope of the literature review, analysed the full paper. The search resulted in 20 papers which were reviewed by means of a collective discussion among the Working Group members who shaped the selection of final content presented in Table 1 depicting the most relevant work for the scope of the study. The table provides an overview of 11 relevant studies that have utilised surveys to measure K-12 CS education, or aspects of it. We discuss these survey studies in this section along with broader research that informs the discussion.

A number of country and regional reports have been produced, detailing K-12 CS education curriculum and initiatives across Europe [6, 54] and the UK [105, 106]. The International Conference on Informatics in Schools (ISSEP) have also welcomed country reports, resulting in publications about the state of K-12 CSED in the United States (US) [57], Wales [76] and Poland [104], among others.

Efforts have been made to measure CS and Computational Thinking (CT) implementation across countries. In 2012, an international study was conducted to explore the state of the art and current activities regarding the teaching of CS in K-12 schools [35, 97]. The questionnaire focused on the national state of K-12 CS education in schools, and the associated situation and education of K-12 CS teachers. The survey was completed by CS education experts familiar with the school system of 22 different countries. Since this study focused on the responses of experts, the data surveyed related to the prescribed rather than the enacted curricula of the countries under study. Similarly, another international survey focused on measuring CT in particular [72]. This was achieved through a 2014 ITiCSE working group who distributed a survey to K-9 teachers in Finland, Italy, Lithuania, Netherlands, and Sweden to better understand how CT was being integrated into the K-9 classroom. Based on their survey results, they were able to see how teachers' perceived CT and how they enacted curriculum to teach CT. They concluded with a recommendation to work toward quality professional development, resources, and materials.

In the US, a report on the state of K-12 CS education [1] was conducted by comparing how each state aligned with Code.org's nine policies to make CS fundamental, examining elements of: state plans, CS standards, state funding for CS, certification pathways for CS teachers, pre-service teacher preperation in CS, state supervision, high school offerings of CS, and admission requirements of CS for higher education systems. Using these nine policies as benchmarks, they were able to identify how states compared in regards to the policies and track changes over time, while also highlighting each state's complexities and differences. In addition, CSTA conducted a national survey of high school CS teachers [32] and a survey of high school administrators [31] to see the ways CS education is conceptualised within US high schools, how it is being taught, and to better understand the accessibility of CS education. Many landscape surveys of K-12 CS education have been conducted by individual states to expand on these findings, particularly those states who are a part of the NSF Alliance Organization: Expanding Computing Education Pathways Alliance (ECEP). Constructing the landscape reports often involved states administering independently designed surveys to teachers, administration, and sometimes industry leaders within the state. Outcomes included a mixture of better understanding the K-12 CS curriculum within schools, teachers' understanding of CS standards, CS teacher certification and professional development opportunities, demographic information of K-12 CS students and teachers, and capturing enablers/barriers to including CS in K-12 education [28, 40, 49, 63].

The ACM Transactions on Computing Education (TOCE) dedicated two volumes for a special issue that covered a range of countries as case studies in K-12 CSED (see Vol. 14, no. 2, 2014) [59] and increasingly researchers have published works about their country curricula and/or initiatives, including in France [24], the Netherlands [9], Australia [47] and England [18], to name a few. Many of the studies utilise the Darmstadt Model in order to provide a unifying framework that allows a well-founded critical discussion about the future of K-12 CSE. The Darmstadt model looks at:

- Educational system: Organisational aspects of subject, Enrollment, School type,
- Socio-Cultural related Factors: History of ICT and Informatics in School, Age, Gender, Social and Immigration Background, Family Socialisation, Public opinion, Techno-economic development,
- Policies: Research and Funding Policies, Education Policies, Quality Management,
- Teacher Qualification: Teacher Education, Professional Experience, Motivation: Student, Teacher Intentions: Learning Objectives, Competencies, Standards
- Knowledge: Computer Science, ICT Curriculum Issues Examination/Certification Teaching Methods, CSE, General Education
- Extracurricular Activities: Contest
- Media: Technical infrastructure, Textbooks, Tools, Didactical software, Visualisation software, Unplugged Media, Haptic media Research

Researchers have also performed comprehensive curriculum analyses across countries and resources to identify the emergence of common K-12 CS education topics and concepts to inform future curriculum developments and research [10, 43, 119]. This international effort put forward in the last decade has shown the importance of a shift of focus from learning computing and Information Technologies applications to a shared rigorous computing academic foundation. Zendler et al. [125] surveyed CS professors in 2010 to determine CS learning objectives that should be included in the intended K-12 CS curriculum related to computing content and process concepts. They expanded on their work to see if this intended curriculum can be generalised across different contexts by surveying experts in both Germany and the USA in 2015 [124]. Findings revealed not only key content and process concepts to be included in higher education curriculum to prepare K-12 CS teachers, but also that curricula can be generalised and differentiated across international contexts. As core CS objectives are identified and integrated into intended curriculum, research should continue to evaluate and understand the curriculum that is enacted in the K-12 CS classroom.

Additional surveys have been used to understand perceptions of key stakeholders in order to understand factors that impede implementation of the intended curriculum. Wang et al. [117] identified student, parent, teacher, and school administration perceptions of, and access/barriers to K-12 CS education through a mixture of telephone and online surveys. Key outcomes were the realisation of a need for shared understanding of what CS is and is not, that there is a high demand for CS but a lack of availability and access, as well as the necessity of support for CS teachers and a need for flexible CS curriculum. Wong et al. [122] similarly identified challenges as including a lack of teacher training and unified curriculum in their survey of Hong Kong schools perceptions of coding education. As K-12 CS education continues to grow and expand, it is necessary to ensure an agreed-upon CS curriculum and access to quality professional development opportunities.

Student performance in K-12 CS improves with opportunities to assert agency [69, 111]. Marzano and Kendall's [73] new taxonomy of critical thinking lists the self-system as an important aspect of developing student agency. Thus, CS teachers who are empowered with the ability to interpret the intended curriculum offering students options to learn CT and develop agency will positively influence educational attainment. For instance, in 2016 Wang et al. [117] explored perceptions, access, and barriers to K-12 CS education with a survey of students, parents, teachers, and school administration. Wang et al. conclude with a recommendation for more research on equitable pedagogy in K-12 CS to "help inform teaching and enlighten us on methods to lessen biases and discrepancies in exposure in order to create opportunities for all to advance with computer science" (p. 650).

3.3 Measuring Enacted Curriculum

Porter and Smithson [85] define a systematic description of four curriculum components: intended, enacted, assessed and learned, to support evaluation and comparison of curriculum at different points. Blank et al. [17] extend on this foundation to define a standardised Survey of Enacted Curriculum (SEC) for Mathematics education organised into three dimensions: topic coverage (for associated Mathematics topics), cognitive demand and mode of presentation. Cognitive demand identified nine descriptors, including memorisation, understanding concepts, and solving routine problems; mode of presentation included seven descriptors, including exposition, pictorial models and fieldwork. Each lesson articulated by a teacher could then be described along these three dimensions. This provided a rich, but relatively simple mechanism for capturing a subset of enacted curriculum. Their instrument was employed in a three year study of mathematics and science education [102], making recommendations for targeting teacher professional development resources.

The SEC survey [17] serves as a guide for future development of surveys for enacted curriculum, defining the importance of capturing teacher beliefs, perceptions and activities. The SEC survey instrument is extensive, capturing hundreds of data points around opinion, practice, instructional content (associated with Mathematics and Science education), professional development, and teacher and environment demographics. This serves as a basis for the identification of survey points (e.g. curriculum, self-efficacy, resources, pedagogy, practice, assessment, and professional development) in this survey instrument, although it is recognised that due to the specific curriculum context and date of development, SEC does not address K-12 CS curriculum content, recent pedagogical practice, modern tools and resources, or professional development opportunities.

A more recent large-scale international survey is the Organisation for Economic Co-operation and Development (OECD) Teaching and Learning International Survey (TALIS), which first launched in

Table 1: Overview of surveys presented within the literature review

Scope/Domain	Country	Year of survey	Respondents	Sample Size
CS [49]	NY, US	2018	Teachers & administrators	344
CS [28]	CT, US	2018	Teachers, administrators, & school counselors	207
CS [48]	Australia	2018	Teachers	119
CS [63]	IN, US	2017	Teachers & administrators	355
CS [117]	US	2016	Teachers, administrators, parents, & students	15,929
CS [125]	US	2015	Teachers	48
CS [32]	US	2015	High school CS teachers	1,354
CS [122]	Hong Kong	2015	Schools	42
CT [72]	Finland, Italy, Lithuania, Netherlands; Sweden	2014	K-8 teachers	961
CS [31]	US	2014	High school administrators	503
CS [40]	Maryland, US	2013	Teachers	85

2008 and covers around 260,000 teachers in 15,000 schools across 48 countries [2]. TALIS is an annual survey of teachers, school leaders and learning environments. TALIS uses a survey instrument with evidence of reliability and validity and provides analysis that helps countries identify others facing similar challenges, informing policy across the world. The latest conceptual framework [2] describes how TALIS addresses topics relating to professional characteristics and pedagogical practices at the institutional and individual levels in schools, including: teachers' educational background and training, professional development, implemented instructional and professional practices, teacher self-efficacy and teachers' workplace and job satisfaction. At the institutional level, TALIS addresses issues of school leadership, feedback systems, and school culture. As being a survey instrument with evidence of reliability and validity adopted across numerous countries with detailed conceptual framework [2], the TALIS can provide a founding model and exemplar for informing the design of teacher surveys as well as comparable data source when comparing CS education with general teaching practices and teacher experiences around the world.

van Veen et al. [114] identified early in the development of CS curriculum the need for supporting both intended and enacted curriculum in curriculum descriptions. However, there has been little work within the CS domain on capturing and describing enacted modern CS curriculum, less so in a generalisable and standardised manner.Some efforts have made progress in CS to explore aspects of curriculum. For example, Bienkowski and Snow [15] have undertaken initial work with a mixed-methods instrument to study secondary curriculum enactment and teaching quality, focusing on inquiry and CT practices. Additionally, Rutstein et al. [95] describe their evaluation instrument for curriculum implementation within the content of the ECS (Exploring Computer Science) program, aiming to characterise the relationship between measures of K-12 curriculum implementation and student learning outcome. In this work, enactment is defined purely as "lessons modified, skipped, and added", however factors that impact enactment are identified and captured within the analysis, including the learning context, as aspects of human capital, social capital and technical and resource capital. They describe the development of a series of six surveys, encompassing background and teacher professional development, with the remaining surveys addressing specific units within the ECS curriculum.

In New Zealand, Bell et al. [11] describe a rich case study of secondary computing curriculum implementation. They discuss both the intended curriculum and standards environment, as well as a case study of enacted curriculum, analysing responses across two survey periods for an unpublished survey, with n=91 [108] and n=109 [107] respectively over the two periods. In their survey, they gather information on teacher motivation, background demographics, implementation of standards, programming language selection, and confidence levels (including related explanation of Mathematics concepts). Researchers are also diving deeper into the classroom context to gather valuable data about teachers' enacted curriculum. For example, Prescott et al. [86] explore the experience of two middle school science teachers integrating CT concepts into their science class. Although this is a less scalable method, it provides rich data and insight into the experience of teachers implementing CT in the classroom.

Researchers have started to collect data about, not only what is being taught in classrooms, but factors that impact on teachers' implementation of CS curriculum. For example, Vivian and Falkner [115] conducted a survey of Australian Digital Technologies (CS) teachers (n=113) to gather information on K-12 enacted curriculum, with a focus on assessment practices, reporting confidence and self-efficacy [7] against teaching and assessing a range of CS topics. Teachers were asked to describe "any formative and summative assessment activities, processes, dialogue, instruments or resources" that they used with the context of assessing a programming activity, providing a rich description of enacted curriculum for this specific aspect of K-12 CS.

Many of the studies discussed have adopted teacher surveys as a means to capture classroom activities and practice, however, self-report measures are widely critiqued for its validity and scientific rigour [55]. Despite self-report measures being one of the most widely used measurement strategies in fields such as education and

psychology, a limitation is that it relies solely on participants being truthful and their ability to accurately recall information. However, researchers [25] have found that teachers' self-reports are a reasonably accurate reflection of classroom practice when comparing both teacher surveys and observational data about teachers and their self-reporting of student behaviour and classroom management strategies. Haeffel and Howard [55] state that prior work has found that although individuals may not be able to always accurately reflect on their cognitive processes, they are reasonably capable of being able to validly respond to questions about constructs such as attributions, plans, attitudes, and beliefs. Porter and Smithson [85], amongst others [20, 21, 25], identify the need to gather further evidence of reliability and validity in survey instruments through the collection and comparison with alternative data sources, such as teacher observation studies, teacher logs and examples of student work. This represents an area of future work, both at the level of individual country or region reports, and at an international level.

4 OBJECTIVES AND RESEARCH QUESTIONS

In coming together, the broad objectives of the Working Group were to:

- To build an international research collaboration and strategy for measuring K-12 CSED implementation in schools.
- To initiate a scalable, collective effort for a deeper investigation into what is happening in schools, based on the experiences of educators in classrooms.
- To develop an open source teacher survey instrument that can be implemented across countries.

This Working Group builds on prior work in K-12 CS education measurement and reporting to develop an instrument with evidence of reliability and validity for surveying K-12 teachers about enacted curriculum and their experiences and approaches to CS curriculum implementation in the classroom, including demographics, self-esteem, teaching methods, use of resources and curriculum topics implemented. The goal of the Group is to develop an instrument with evidence of reliability and validity that can transcend boundaries and be used for consistent, ongoing data collection.

The working group sought to investigate the following research questions:

- What are the similarities and differences in K-12 teachers' perspectives regarding their pedagogy, practice, assessment, resources, and experiences?
- What would a universal evaluation instrument with evidence of reliability and validity to capture these perspectives from across the world contain?

This report presents the development of a country report template and K-12 teacher survey instrument and presents the pilot survey findings with teacher cohorts from across the seven working group member countries. While it is acknowledged that the research questions focus on the third objective, this report by it's nature (the international Working Group members and the pilot study) broadly aims to satisfy the first two objectives.

5 METHODS

5.1 Research Design

This study adopts a mixed-methods design process centred around the development and evaluation of a teacher survey instrument, that included a review of related K-12 CS survey instruments and development of instruments for this study, along with a pilot of the instruments and a focus group to revise the teacher survey instrument. This process is captured in Figure 1. This process is supported by approaches in educational and psychological testing [58] that use a combination of theory and expert opinion as the basis for the development and selection of testing items, paired with an iterative and multi-stage process in evaluating test items (in this case being teacher survey items). In the following section we describe the processes involved in developing the two instruments used in this study: the *country report* and the *teacher survey*.

5.2 Instrument Development

As discussed in the Related Work, there are a number of papers and reports presenting country data about CS education with a particular focus on "intended curriculum", with few capturing the "enacted curriculum". Although primarily setting out to investigate the enacted curriculum, the Working Group also identified a need to capture information about countries that may support analysis and comparisons. Comparisons between countries are only useful when a survey administrator is able to conclude that any emerging differences between countries are unlikely to have arisen by chance, thus being able to explain reasons for the observed differences [2]. The enacted curriculum should reflect the curriculum policies of the state (the intended curriculum) [85].

It was therefore determined that two instruments would be developed: *a country report template* and a *teacher survey instrument* (see Appendix A). The country report template is to be completed by the survey administrator and the survey instrument by classroom teachers. Having survey administers complete a country report for their survey cohort as a pre-survey phase had a number of perceived benefits. It allowed the capture of data of K-12 CS curricula and implementation across countries for comparing and contrasting, and longitudinal analysis of changes over time, as well as providing background information about the context of the cohort being surveyed. Additionally, it reduced duplication and the need for teachers to answer the same questions about their intended curriculum, thus reducing survey completion time and possible variations between responses that might emerge. This additional time also allowed the teacher survey instrument to focus on teachers' own demographics, school contexts and what is happening in the classroom - the enacted curriculum.

5.2.1 Country Reports. To develop the country report template, a number of reports and papers capturing international and regional data were used as a basis to identify potential key categories relevant to comparing and contrasting school demographics and intended CS curriculum across countries (e.g. [60], [106], [35], [61], [1]). The Working Group searched and curated relevant papers from the ACM Digital Library and Google Scholar. Search terms such as "informatics", "computing", "digital technologies", were included to

capture reports for countries referring to CS curriculum in alternative ways. The reference list of the identified papers were used as a basis to identify other key papers. These references were curated into a spreadsheet with details entered for each of the headings (e.g. date of publication, year levels, country, methods, etc.). Papers were included if they captured or reported on country or multinational K-12 CS education from an intended curriculum perspective (e.g. details around topics, age bands). From here a new spreadsheet was devised to curate key categories and survey item questions from these prior studies to form a country report template.

There was a challenge in capturing implementation of CS topics across countries, due to the differences of CS curriculum between countries as well because it was dependent on whether a specific CS curriculum was available. Therefore, it was decided that a comprehensive measure of CS topics being implemented was a key consideration of the enacted curriculum and would be captured via the survey instrument. However, as a broad comparison across countries, we reviewed various curriculum analysis reports [72][9], previously mentioned country reports, country curriculum documents (of those represented by Working Group Members) and the CSTA standards [99]. We used curriculum documents and CSTA standards as a starting point of broad CS topics which were expanded on by the Working Group for a high-level comparison across countries. The recurrence of CT within literature merited its inclusion as a high-level topic. The goal is that the METRECC instrument would seek to identify more specific and fine-grained topics which would be used to inform the revised country report instrument. All the curated categories were organised in a spreadsheet and presented as a template to be completed by survey administrators.

The Working Group reviewed the draft country report template to determine which categories would be eliminated, adapted or kept, taking into account considerations toward language, nuances and transferability across countries. Items which the group deemed difficult to clarify were removed. For example, the provision of national funding (taken from [1])) was eliminated due to idenfication of the vast differences across countries funding schemes.

Additionally, in recognising the challenge of mapping CS curriculum availability and implementation requirements across countries due to differing age groups for grades, it was decided that student ages would be included alongside grades for ease of completion. To test the template, each Working Group member took the template and completed it for their respective country. As members completed the template, they noted any confusion around language, categories or problematic categories. No significant changes were made but it was decided that to support administrators to complete the country report template, instructions and a glossary would be provided.

The following information is captured in the country report template:

- Country demographics and information relating to schools (e.g. total population, number of schools, number of teachers).
- CS curriculum state or country plan standards and requirements.
- Year Level (with age for comparisons) mapped to prescribed curriculum and programming requirements.
- General CS topics covered.

A threat to the validity with the country report is that there is the risk of the local survey administrator or representative misunderstanding their own curriculum, or K-12 school system, and therefore entering data incorrectly. To reduce this, we would encourage where possible, administrators to cross-validate their country reports with colleagues or other local experts.

We present the results of the curated data for our pilot study in Section 6.

5.2.2 Teacher Survey. The Working Group undertook a collaborative, iterative process to develop a teacher survey instrument that could be transferable across countries. This section describes the process of designing and preparing the survey instrument as seen in parts 1-3 in Figure 1.

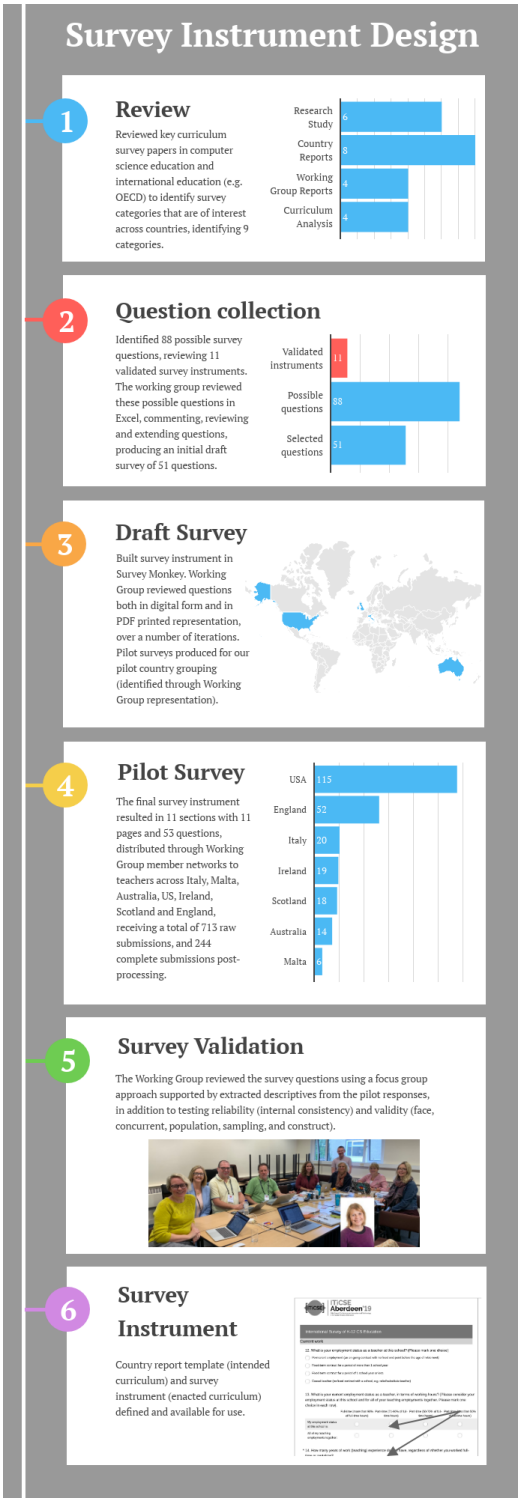
The Working Group broadly undertook the following steps to define the survey categories and questions:

(1) Curation and review of CS and education survey papers and reports, identifying those that included survey instruments with evidence of reliability and validity.
(2) Identification of survey categories.
(3) Curation of survey questions from surveys with reliability and validity evidence that aligned with survey categories.
(4) Addition of new survey questions for categories that were not found in surveys with evidence of reliability and validity.
(5) Refinement of survey categories and questions and selection of questions for inclusion in the survey.
(6) Building of the online survey and final survey reviewed by all members.

The Working Group leaders curated and reviewed key CS education survey papers and reports with survey instruments that could guide possible survey categories and questions. The ACM Digital Library and Google Scholar were used to locate relevant articles, using search terms such as "informatics", "digital technologies" and "computing" in addition to CS to capture international work. These were paired with search terms such as "teacher survey" and "survey instrument". Papers were included if they had included survey items (or linked to instruments) in their paper. The reference list of identified papers was used to check for any additional relevant papers not captured in the database search. Relevant papers were curated and organised using a Mendeley Group and entered into a Google spreadsheet with metadata and annotations (such as author name, country covered, year published, topic and sub-topic, method, etc.). The Working Group leaders developed a set of key categories that might be of interest internationally as a starting point. The categories were shared with Working Group members for review, alterations and the addition of new categories. Although initially seeking to identify CS education surveys and articles reporting on teacher surveys, the search was broadened to also review known international education survey instruments such as the TALIS Survey [2] that could provide valuable survey items with evidence of validity and reliability for demographics and teaching practices.

Once a set of draft categories were agreed upon by the Working Group, these formed separate sheet labels in a Google sheet. Collaboratively, Working Group members curated and added questions from surveys with evidence of reliability and validity, including

Figure 1: Survey design and validation methods



identifying metadata such as the "sub-category" (e.g. classroom equipment), "response options" (e.g. laptop, computer, tablet, other), the "source" (e.g. TALIS Survey [2]), the measure (e.g. Likert, checkbox, multiple choice), possible threats to validity and whether or not the questions were from a survey instrument with previous evidence of reliability and validity. This resulted in 88 initial example questions from 11 sources [2, 8, 36, 44, 65, 79, 82, 84, 85, 89, 103].

Upon developing the questions, consideration was taken with regards to the best way to measure responses. Here, we discuss some examples and how prior survey instruments that have evidence of validity have been utilised. To capture teacher demographic data and teachers' classroom composition in Sections two to four of Table 2 (e.g. gender, low-socio-economic status, disability, gifted students), we adopted a majority of TALIS [2] questions as these have been found to translate across 48 countries. For classroom composition, we use teachers' estimations of how many students have various characteristics against a percentage. We also utilised TALIS questions and items about professional development activities and barriers for section 10 relating to professional development. This allows us to compare benchmark results against TALIS survey reports and also allow administrators to compare estimations against their country report breakdowns.

To better understand teachers' instructional practices, we reviewed questions in works by [2] and [85], as the authors provide guidance around capturing classroom practice. [2] recommend using frequency of instructional practices rather than measuring teachers' agreement towards the adoption of practices. Similarly, [85] invite teachers to estimate and nominate time spent against various instructional practices in terms of percentage of implementation (e.g. 25-49% on "whole class instruction"). The authors' reason that this measurement facilitates comparisons across classrooms, types of courses, and types of student populations and that they have the advantage of being easy to respond to (i.e. in cases when teachers teach multiple classes or for helping teachers reflect on time spent against practices as they can estimate using various time measures, such as a week or a year of instruction). However, a major disadvantage is that such measures provide a crude estimate. To reduce complexity, we did not include the full matrix columns by [85] that invited teachers to reflect on practices across Bloom's Taxonomy items.

Some items from the Research-Practice Partnerships CS For ALL (RPPforCS) Survey Instruments [27] were adopted in section 3 around teachers' current work and section 10 inquiring about their professional development and use of professional development materials in the classroom. The RPPforCS project collects participation data about teachers participating in the CS for All: Research Practitioner Partnership Project. RPPforCS is focusing on the projects preparing teachers to offer a stand-alone high school course in CS, however, they have made their instruments available to support others in capturing CS implementation.

The survey component measuring CS self-esteem utilised the Bergin Self-Esteem Instrument [14] that was developed as part of a longitudinal study, also utilised by [89] with CS student cohorts. Bergin had developed the instrument as a modification of the Rosenberg self-esteem scale, which has generally been shown to have evidence of high inter-item and test-retest reliability evidence [93] to apply to programming. The 10 items used in the Bergin [14] study were added to the instrument, however, the domain-specific subject was adapted from "programming" to "Computer Science"

to reflect the broader K-12 CS curriculum that the survey was investigating. Teachers responded to statements on a 7-point Likert scale, from "Strongly Disagree" to "Strongly Agree". The items were generally about CS capabilities and we wished to measure teachers' self-esteem to determine how impact on classroom practice.

Some additional question development were devised using survey instruments used in other studies, such as by [115], and those developed as a collaborative exercise by Working Group members. Our Working Group investigation to evaluate this teacher survey instrument will involve checking these questions for evidence of validity.

The Working Group held an online meeting in which the group worked through the curated questions to determine whether to "keep" or remove them as well as considering and discussing the language of questions, duplicates, response options and the transferability of questions across the various countries and alignment with the study objectives (e.g. to investigate the enacted curriculum). This process was undertaken twice (once offline) and resulted in the final set of categories (now referred to as sections) and 51 key research questions that were ready for import into the SurveyMonkey tool. A number of sub-sections and questions, particularly within teacher confidence and motivation, were excluded from the final survey. The final draft survey was downloaded from SurveyMonkey as a PDF and emailed to Working Group members for review, with required amendments made. Two researchers tested a copy of the digital survey on SurveyMonkey.

The final survey instrument resulted in 11 sections with 11 pages and 53 questions (two questions being administrative). The survey overview is presented in Table 2. 33 (58.5%) of the total survey questions were set as "required" with the remaining as optional. Required questions were determined as those key to answering our Working Group research question that focused on the enacted curriculum, with optional being extensions and as useful to providing additional supporting data. In the following section, where relevant, we broadly describe some of the survey sections and where questions and measures were sourced from.

A final question asked teachers if they would be willing to consent to their anonymous data being shared with the computer science education research community for future use.

A large portion of survey questions (39.6%, n=21) related to teacher demographics, their current role and qualifications/experience (see Table 2). The second highest portion of questions related to what teachers are doing in the classroom and the resources and practices they are adopting (39.6%, n=21), aligning with our survey goal of investigating the enacted curriculum. Additional question topics related to student cohort composition, professional development and teacher' perceived CS self-esteem.

Examining an overview of the types of questions utilised in the survey instrument, there were a reasonable split between multiple choice questions (35.8%, n=19) and matrix questions (34.0%, n= 18) that used Likert style.

5.3 Data Collection

5.3.1 Country Reports. The seven countries for the pilot study were those represented by the ten Working Group members (*note: at the time of the study, one member was located in Cantina, Italy, but has since relocated to the US*). During Working Group member selection process, consideration was made to include members from a sample of countries, including English as Second Language (e.g. Italy and Malta).

Each Working Group member was responsible for completing their respective country report information using the designed template in an Excel spreadsheet. One Working Group member led the collation of the data from each of the separate country reports into consolidated tables to visualise findings for comparison across countries. The consolidated tables were then presented and cross-checked for accuracy with each of the working group members. Each member verified the data inputs in their country report data revised tables. As mentioned previously, it is acknowledged that a risk to validity is that administrators may provide incorrect or out-of-date information, however, we hoped that the process of cross-checking information would reduce discrepancies.

5.3.2 Teacher Survey. Prior to data collection, each Working Group member was responsible for seeking appropriate institutional and country-specific Human Research Ethics Committee (HREC) approval. In the survey information and advertisements it was stated that teachers participating in the survey from the respective Working Group countries listed, who go on to complete the survey, were able to go into the draw to with a $100 Gift Card (one available per country). If participants were interested in entering the draw they were invited to go to a separate URL (to a Google Form) that collected entries. This was to maintain participant anonymity as contact information was kept separately from their survey response. The survey instrument was built in the English language using the Survey Monkey software, "Simple design". To address participant anonymity, participants were not invited to share personally identifiable information in the survey and the survey was set to "anonymous" with no collection of IP addresses. Three separate surveys were created for targeted regions (Australia, United States, and a general international survey) to accommodate regional HREC information and to include state/territory specific drop-down lists for ease of analysis.

Recall in Section 2: Definitions that our survey instrument does not aim to explicitly define CT, programming or CS, but rather allow for interpretation and use in many contexts. To ensure that teachers understood that the survey was investigating CS as a discipline/learning area and that the term "CS" is used throughout but can encompass whatever their context uses to describe this discipline, the cover page of the survey instrument, states that "*throughout this survey, we refer to "computer science" or "CS" as being synonymous with computing, informatics, computational thinking, computing science and other variations of the term. Being an international survey, it is possible that some questions do not fit very well within your national context. In these cases, please answer as best as you can*".

The survey was shared with the Working Group for distribution on the 29th of June at 5:00am GMT+1. Surveys were advertised among working group member networks (e.g. social media, email listservs) related. The survey was open between 30th of June (GMT+1) and 12th July, with surveys being closed at 11:00am GMT+1 and being downloaded at 12:30pm GMT+1 with a sample size of n=713 (see Table 4). As the data were collected from three

Table 2: Overview of survey instrument sections, topics and question numbers

Section number & heading	Section topics covered	Question numbers	Questions (n and % of total)	Required (n and % of total)
1. Introduction	Study information; Consent to participate	1	1 (1.9%)	1 (3.2%)
2. Demographics	Teacher demographics (e.g. age, location); School demographics (e.g. socioeconomic, remoteness)	2-11	10 (18.9%)	8 (25.8%)
3. Current work	Employment; Teaching role; Subject expertise; Experience teaching CS	12-18	7 (13.2%)	2 (6.5%)
4. Qualifications	Qualifications in teaching, computing and other subjects; Participation in classroom research	19-22	4 (7.5%)	1 (3.2%)
5. Student composition	Student cohorts; Classes taught and class size; Demographics of students (reported)	23-25	3 (5.7%)	2 (6.5%)
6. Support and resourcing	Access to infrastructure, facilities and equipment; Available school support (people, PD) and perceived needs; Place of CS classes; Local CS outreach engagement and awareness; CS topics taught and unplugged/plugged; Curriculum document/s used (if any); Access to CS and general teaching materials and technology	26-38	13 (24.5%)	7 (22.6%)
7. Assessment of student learning	Implemented assessment approaches in CS; Reporting required or not	39-40	2 (3.8%)	2 (6.5%)
8. Classroom practice	Learning and teaching strategies (CS specific and general); Programming environments and motivation for use	41-46	6 (11.3%)	4 (12.9%)
9. CS self-esteem	Teachers' self-esteem in CS	47	1 (1.9%)	1 (3.2%)
10. Professional development	Participation in types of PD activities; Structure/benefits of PD activities; Perceived PD needs; Extent PD resources used in classroom	48-52	5 (9.4%)	2 (6.5%)
11. Open access data	Consent for anonymous data to be included in open access	53	1 (1.9%)	1 (3.2%)
Total		1-53	53 (100%)	31 (100%)

Table 3: Overview of survey question types

Question type	Number (n)	Percentage (%)
Multiple choice	19	35.8%
Matrix	18	34.0%
Checkbox	10	18.9%
Open Text	5	9.4%
Dropdown	1	1.9%
Total	**53**	**100%**
(Extensions) Textbox "other" option	21	39.6%
(Extensions) Explain textbox	4	7.5%

separate regional surveys, the separate files were combined into a single Excel spreadsheet for pre-processing. Very minor alterations were made to the structure and a buddy system was used to cross-check any structural changes. The total spreadsheet contained 379 columns.

Table 4: Raw individual survey completion statistics (pre-processing)

Survey Link	Respondents	Average time	Completion rate
International	407	16m;3s	43%
United States	271	19m;38s	46%
Australia	35	15m;35s	46%
Total	713		

5.4 Data Pre-Processing

This section describes the data pre-processing process for the teacher survey data that resulted in the 244 responses included in our final dataset.

As a starting point, the following individual participant responses were removed:

- Respondents from countries outside of those represented by members in the working group (n=63).
- Those that selected "do not consent" (n=6) and "under 18" (n=2).

- Responses where the participant consented to the survey but exited after the consent page without entering/saving any additional data (as SurveyMonkey saves the data after each page, n=104).

A day into releasing the survey it was noticed that a high number of participants were exiting the survey and it was decided that to ensure the Working Group had some data to work with, within the short survey release period, and that we had sufficient data across questions for instrument validation that key questions throughout the survey would be set to "required" (see breakdown of required questions by section in Table 2).

In our dataset, we kept those respondents who had answered all required questions (31 of the 53 questions) as a minimum, resulting in a final dataset of 252 responses for gathering evidence of reliability and validity and pilot analysis. This ensured a more robust dataset with minimal missing values and that participants had not exited early due to wanting to withdraw from the study. One of our final questions asked participants if they would be willing to have their anonymous data shared with the CS education research community. Some of the remaining respondents (n=8) opted "no" to sharing their data and it was decided that to enable us to share a complete dataset that we have validated with the CS community, we would remove this small number of responses. This resulted in a total and final dataset of 244 respondents for analysis. Of this dataset, 85 respondents answered all 53 questions and the average questions answered were 42.

5.5 Pilot study sample

Our pilot sample is comprised of seven countries and a total of 244 teacher survey participants. Firstly, we present the an overview of the country contexts involved in the study, followed by the demographics of our teacher survey participants.

5.5.1 Country Contexts. Our sample is comprised of seven countries, including Australia, England, Ireland, Italy Malta, Scotland and the United States. Tables 5-7 presents the country demographics. Table 8 presents the structure of the school system across countries, with Tables 9-11 presenting CS curriculum implementation across those year levels.

Tables 5-7 show a snapshot of the Working Group members' understanding of the intended CS curriculum within their country or state. It became clear that provision across snapshot categories for K-6 and Year 7-12 are sufficiently different and should be captured separately. Working Group members also provided key contextual information to expand their country/state tabulated snapshot data, which we present here.

5.5.2 Country Information. In capturing country report data, it was identified there is a need to provide additional information to expand on the data in the tables to explain some of the intricacies and to provide supporting contextual information. We include the descriptions for countries in this study below.

In *Australia* CS commences from the first year of school until year 10. No national curriculum is mandated at the final stages of secondary school (Grade 11 and Grade 12) because courses are optional for students and align to final certification. CS curriculum is at the early stages of implementation, with each state or territory determining reporting requirements. As a result, reporting expectations vary for both government and privately funded schools. Formal pre-service training and in-service professional CS learning varies in terms of requirements and availability.

England has a mandatory computing curriculum in state-funded schools from age 5-16 (Year 1-11) which covers computer science, information technology and digital literacy. This can be seen in the representation of teaching year levels of teachers from England who took the survey (see Table 18). At age 14, students can additionally elect to take a GCSE in CS, and at age 16, an A Level in CS. Postgraduate initial teacher training courses have been available, with financial incentives, for secondary computing teachers since 2013. The government supported the Network of Excellence [101] [101] for in-service professional development of computing from 2013-2018 with a small amount of funding, and then massively increased the amount of support by forming the National Centre for Computing Education in 2018 to support in-service teachers.

In *Ireland* secondary school education is in two phases including the Junior Cycle at age 12-15 followed by the Leaving Certificate (which includes fifth and sixth year). These phases/years are mandatory across all schools. There is an optional year, TY (also known as transition year or fourth year). In the Junior cycle students undertake short courses across a range of subject areas which includes an optional in coding. In 2018 Ireland finalised the pilot upper secondary CS curriculum and by September 2020 all schools will be eligible to implement the CS curriculum at their own choosing. In primary, the CS curriculum is under development and is expected to be launched in 2022. The pilot phase involved a school choosing their own concepts and content which will be used to develop the curriculum. Although the secondary curriculum is optional teachers have control to decide on resources and pedagogy.

In *Italy* the secondary schools vary in specialisation, including academic, technical and vocational. CS is not mandatory in all types of high school but it is delivered in secondary schools specialising in technology or science. Object orientated programming is mandatory in the higher stages of technical schools. CS is promoted in primary and lower secondary, with CS guidance that includes "computational thinking" concepts. Formal reporting takes place in some secondary schools.

Since 2018-2019 in *Malta*, all pupils from year 7-11 follow an ICT C3 certificate which includes CS education. In the primary years Computational Thinking learning objectives are embedded in the Digital Literacy cross-curricula theme and the teacher decides how and when to implement them. These are not formally assessed. CS is a standalone subject at year 9 comprised of two branches, one being VET IT (based on networking and vocational/hands-on) and the other Computing (including programming, databases, computer architecture). Secondary schools formally report on CS in years 7-11. Pre-service CS training is compulsory for teachers delivering CS from years 7-11.

In *Scotland* all pupils have an entitlement from pre-school up to 3rd year in secondary school to a Broad General Education (BGE). Across the BGE computing science guidelines are organised into a discrete subject. However, teachers and schools have ownership on its delivery. Fourth year to 6th year computing science is optional for qualifications. In 2016 the computing science curriculum K-10

Table 5: Overall school-related demographic information for countries.

COUNTRY/USA STATE	AUSTRALIA (AUS)	COLORADO (US-CO)	ENGLAND (ENG)	IRELAND (IRL)	ITALY (ITA)	ILLINOIS (US-IL)	MALTA (MLT)	SCOTLAND (SCO)
Population (million)	25.09	5.69	55.62	4.70	60.50	12.7	0.47	5.44
No. of schools	9,477	1,900	29,972	3,961	8,636	4,266	170	2,400
No. of students	3,893,834	911,536	8,378,809	920,867	8,422,419	2,072,880	46,247	693,251
No. of teachers (FTE)	288,583	59,989	498,100	66,327	872,268	135,701	2,976	51,959

Table 6: Curriculum concepts across pilot study states and countries: Explicit (✓) Implicit (❖) Not covered (×)

Concepts	AUS	US-CO	ENG	IRL	ITA	US-IL	MLT	SCO
Computational Thinking	✓	✓	✓	✓	✓	×	✓	✓
Computer Systems	✓	❖	✓	✓	✓	×	❖	✓
Networks and Internet	✓	❖	✓	✓	✓	×	✓	✓
Data & Analysis	✓	✓	✓	✓	✓	×	✓	✓
Algorithms and Programming	❖	❖	✓	✓	✓	×	❖	✓
Impact of Computing	✓	✓	✓	✓	✓	×	✓	✓

Table 7: Demographics of pilot study countries/states education systems.
(i) Yes (✓) No (×) Additional information (❖)
(ii) Pre-service training - Varies(V) Compulsory (✓), Elective (E) *Date previous CS curriculum refreshed.

COUNTRY/USA STATE	AUS	US-CO	ENG	IRL	ITA	US-IL	MLT	SCO
CS State or country plan	✓	×	✓	×	❖	×	✓	✓
CS Curriculum K-6 standards defined	✓	×	✓	❖	❖	×	✓	✓
CS Curriculum: Y7+ standards defined	✓	✓	✓	❖	❖	✓	✓	✓
CS Guidelines - standalone subject	✓	✓	✓	❖	❖	×	❖	✓
CS Guidelines - across disciplines	×	×	×	×	❖	×	❖	×
Teacher autonomy to implement state/country guidelines as standalone or cross discipline	✓	✓	✓	×	✓	✓	❖	✓
CS Formal reporting	V	×	×*	×	❖	×	❖	❖
CS in pre-service training primary	E	E	✓	E	V	E	×	E
CS in pre-service training secondary	E	E	✓	E	V	E	✓	✓
CS training for in-service primary?	V	✓	✓	V	V	×	✓	
CS training for in-service secondary?	V	✓	✓	V	V	✓	✓	
Year endorsed	2015	2018	2013/14	×	×	×	2018*	2016*

Broad General Education for curriculum content for computing science was refreshed.

In the *United States (US)* there is no national CS curriculum, however, individual states can mandate their own CS curriculum to be implemented. If there is no state or district wide curriculum formally adopted then primary and secondary schools have autonomy to implement CS curriculum and/or classes, often using the Computer Science Teachers Association (CSTA) standards as a framework. For the purposes of this paper, and as we are presenting a pilot comparing countries, the US have used the CSTA standards to reflect on implementation due to the variances between states. As US state CS curricula evolves, it would be prudent for future work or use of this survey instrument to reflect on the context of their state curriculum. State funding is sometimes available for CS in-service professional development through various initiatives.

In Illinois, currently there are no state standards but districts have the ability to implement their own. Chicago public schools,

Table 8: Approximate age and school placements across authors' countries/states education systems. *Youngest age at the start of the school session. For example, in Scotland for session 2019/2020 almost all children between the ages of 4 and a half and 5 years old will start primary school at the start of term in August. Children who attain the age of 5 years between 1 March 2019 and 28 February 2020 should be registered for education in January 2019 to start school in August 2019.

COUNTRY	AUSTRALIA (AUS)	ENGLAND (ENG)	IRELAND (IRL)	ITALY (ITA)	MALTA (MLT)	SCOTLAND (SCO)	US
AGE* (Years)							
2+				Pre-school	Kindergarten	Pre-school	Pre-school
3	Pre-school	Pre-school	Pre-school	Kindergarten	Kindergarten	Pre-school	Pre-school
4	Kindergarten	Pre-school	Junior Infants	Kindergarten	Kindergarten	Pre-school	Pre-school
4-5	Reception/ Foundation	Reception	Senior Infants	Kindergarten	Year 1	Primary 1	Pre-school
5-6	Year 1	Year 1	First Class	First class primary	Year 2	Primary 2	Kindergarten
6-7	Year 2	Year 2	Second Class	Second class primary	Year 3	Primary 3	Grade 1
7-8	Year 3	Year 3	Third Class	Third class primary	Year 4	Primary 4	Grade 2
8-9	Year 4	Year 4	Fourth Class	Fourth class primary	Year 5	Primary 5	Grade 3
9-10	Year 5	Year 5	Fifth Class	Fifth class primary	Year 6	Primary 6	Grade 4
10-11	Year 6	Year 6	Sixth Class	First class lower high school	Year 7	Primary 6	Grade 5
11-12	Year 7	Year 7	First Year	Second class lower high school	Year 8	Primary 7	Grade 6
12-13	Year 8	Year 8	Second Year	Third class lower high school	Year 9	S1	Grade 7
13-14	Year 9	Year 9	Third Year	First class higher school	Year 10	S2	Grade 8
14-15	Year 10	Year 10	Transition Yr.	Second class higher school	Year 11	S3	Grade 9
15-16	Year 11	Year 11	Fifth Year	Third class higher school	Sixth form lower	S4	Grade 10
16-17	Year 12	Year 12	Sixth Year	Fourth class higher school	Sixth form higher	S5	Grade 11
17-18		Year 13		Fifth class higher school		S6	Grade 12

for example, implemented a graduation requirement that all high school students have one year of CS education.

In Minnesota (US) each school district decides if CS is a standalone subject. The state government is trying to include computational thinking within the performing arts and science standards revisions. Although students do not experience a state mandated CS curriculum, teachers can choose to incorporate CS into their classrooms.

We present an analysis of the results of the intended and enacted curriculum across our countries in Section 6.

5.5.3 Teacher Survey Participants. Recommendations on sample size for a pilot study varies in the literature [110]. On one hand, some researchers suggest using a respondent-to-item ratio when determining adequate sample size, then sets a guideline ratio such as 5:1, 10:1, or even 30:1 [110]. Other researchers recommend stating a range of sample sizes as acceptable for pilot studies, and go

Table 9: Programming language curriculum specification across pilot study states and countries.
Age at the start of the school session
Not specified (×) Visual Programming (VP) General Purpose Programming (GPP) Object Oriented Programming (OOP)

Age*/ US Grade	AUS	US-CO	ENG	IRL	ITA	US-IL	MLT	SCO
3-4 Pre-school		×				×		VP
4-5 Pre-school	×	×				×		VP
5-6 Kinder.	×	×	VP	×	VP	×	×	VP
6-7 Grade 1	×	×	VP	×	VP	×	×	VP
7-8 Grade 2	VP	×	VP	×	VP	×	×	VP
8-9 Grade 3	VP	×	VP	×	VP	×	×	VP
9-10 Grade 4	VP	×	VP	×	VP	×	×	VP
10-11 Grade 5	VP	×	VP	×	VP/GPP	×	×	VP
11-12 Grade 6	GPP	×	GPP	VP/GPP	VP/GPP	×	VP	VP/GPP
12-13 Grade 7	GPP	×	GPP	VP/GPP	VP/GPP	×	VP	VP/GPP
13-14 Grade 8	OOP	×	GPP	VP/GPP	VP/GPP/OOP	×	GPP	VP/GPP
14-15 Grade 9	OOP	×	GPP	VP/GPP	VP/GPP/OOP	×	GPP*	GPP
15-16 Grade 10	×	×	GPP	VP/GPP	VP/GPP/OOP	×	GPP*	GPP
16-17 Grade 11	×	×	GPP	VP/GPP	VP/GPP/OOP	×		OOP
17-18 Grade 12		×	GPP/OOP	VP/GPP	VP/GPP/OOP		×	

Table 10: Teacher autonomy in curriculum specification across pilot study states and countries.
Age at the start of the school session
TEACHER CONTROL (TC): The extent to which teachers can decide on individual learning activities or the curriculum predefines at a granular level through pre-defined resources, lesson plans and content. YES(✓) Achievement standards but no description how to get there, teachers can decide within a broad framework. NO (N) Prescribed daily materials explicitly defining what is to be taught and when.
Can decide within a broad framework (S) No CS curriculum defined (×)

Age*/ US Grade	AUS	US-CO	ENG	IRL	ITA	US-IL	MLT	SCO
3-4 Pre-school								
4-5 Pre-school	✓	✓						S
5-6 Kinder.	✓	✓	S	×	S	×	✓	S
6-7 Grade 1	✓	✓	S	×	S	×	✓	S
7-8 Grade 2	✓	✓	S	×	S	×	✓	S
8-9 Grade 3	✓	✓	S	×	S	×	✓	S
9-10 Grade 4	✓	✓	S	×	S	×	✓	S
10-11 Grade 5	✓	✓	S	×	S	×	✓	S
11-12 Grade 6	✓	✓	S	✓	S	×	✓	S
12-13 Grade 7	✓	✓	S	✓	S	×	S	S
13-14 Grade 8	✓	✓	S	✓	S	×	S	S
14-15 Grade 9	✓	✓	N	✓	S	×	S	N
15-16 Grade 10	N	✓	N	✓	S	×	S	N
16-17 Grade 11	N	✓	N	✓	S	×		N
17-18 Grade 12		✓	N	✓	S	×		S

further to describe a recommended scale, with under 50 responses considered very poor, 51-100 as poor, 101-200 as fair, and 201-300 as good. 301-500 is then very good and 501-1000 or more is excellent. This demonstrates the wide range of acceptable sample sizes in the literature.

For our study, we concluded the survey with 713 responses. Once we cleaned the data, 244 responses remained. Given that there were 51 questions on the survey, we achieved a ratio of 4.8:1, closely meeting the basic respondent-to-item ratio size (5:1). With a sample size of 244, we are in the fair to good range for sample size [110]. We address representation of these respondents to the population validity (how these respondents are divided across various groups of populations) in Section 7.

Table 12 presents a breakdown of participants per country. The majority of participants (68%) were from the USA (n=115) and England (n=52). Italy, Ireland, and Scotland make up 24% of the sample, and Australia and Malta represent 8% of the participants (Table 12). Table 13 shows a larger female representation from most

Table 11: Computing curriculum specification across pilot study states and countries.
Age at the start of the school session
COMPUTING CURRICULUM (CC): Optional (O) Elective (E) Compulsory (✓) Not defined (×)

Age*/ US Grade	AUS	US-CO	ENG	IRL	ITA	US-IL	MLT	SCO
3-4 Pre-school						×		
4-5 Pre-school	✓	O				×		✓
5-6 Kinder.	✓	O	✓	NA	O	×	O	✓
6-7 Grade 1	✓	O	✓	NA	O	×	O	✓
7-8 Grade 2	✓	O	✓	NA	O	×	O	✓
8-9 Grade 3	✓	O	✓	NA	O	×	O	✓
9-10 Grade 4	✓	O	✓	NA	O	×	O	✓
10-11 Grade 5	✓	O	✓	NA	O	×	O	✓
11-12 Grade 6	✓	O	✓	O	O	×	C	✓
12-13 Grade 7	✓	O	✓	O	O	×	C	✓
13-14 Grade 8	E	O	✓	O	C/O/N/✓	×	C	✓
14-15 Grade 9	E	O	E/✓	O	C/O/N/✓	×	C	O
15-16 Grade 10	NA	O	E/✓	O	C/O/N/✓	×	C	O
16-17 Grade 11	NA	O	E	O	C/O/N/✓	×	✓	
17-18 Grade 12		O	E	O	C/O/N/✓	×	C	

countries (61%), with Scotland being the only country who had more male participants (n=10) represented than female respondents (n=8). Over half of teachers were between the ages of 40-59 (63%) and a little less than a quarter were between the ages of 30-39 (24%) (see Table 14).

Table 12: Participants per country

Country	**N**	**%**
US	115	47
England	52	21
Italy	20	8
Ireland	19	8
Scotland	18	7
Australia	14	6
Malta	6	2
Total	**244**	**100**

Table 13: Gender breakdown per country

	Male		Female		Prefer not to say		Other	
Country	N	%	N	%	N	%	N	%
US	38	**33%**	74	**64%**	2	**2%**	1	**1%**
England	23	**44%**	28	**54%**	1	**2%**	0	**0%**
Italy	8	**40%**	12	**60%**	0	**0%**	0	**0%**
Ireland	8	**42%**	11	**58%**	0	**0%**	0	**0%**
Scotland	10	**56%**	8	**44%**	0	**0%**	0	**0%**
Australia	2	**14%**	11	**79%**	1	**7%**	0	**0%**
Malta	1	**17%**	5	**83%**	0	**0%**	0	**0%**
Total	90	**37%**	149	**61%**	4	**1.6%**	1	**.4%**

5.6 Data Analysis

5.6.1 Country Reports. Data collected and curated were analysed by using manual comparative analysis of data across countries. Data from the country reports were also used as a benchmark to review and compare results emerging from the survey instrument (e.g. comparing topics within intended curriculum with topics reported by teachers as being taught in the classroom as enacted curriculum).

5.6.2 Teacher Survey. Curiosity about the definition of robust K-12 CS is the broader context in which we framed the survey development. The instrument is intentionally a mixed method tool to not only ascertain affirmation and frequencies or percentages (e.g. time offering robotics or number of years teaching CS) but also to illuminate reasons for the affirmative or negative answers. The pilot survey was specifically designed to gather quantitative and open-text information to help us decide which information was most useful and which was superfluous.

Processed survey data were downloaded into into Excel. From here data were analysed using a combination of statistical analysis scripts in Python and from importing the data into the statistical software package, SPSS Statistics. For this pilot study, data analysis of the survey is primarily descriptive. We provide further information about statistical analysis with the Results in Section 6.

The evaluation of the survey instrument was primarily conducted using statistical analysis of the pilot survey data (n=244) for evidence of validity and reliability with Python scripts in Excel. Two working group members led the analysis and cross-checked with one another. We present the precise statistical tests with the results in Section 7.

The revision of the survey instrument was based on the results that emerged from a self-directed focus group, chaired by two working group leaders. During the ITiCSE conference, an afternoon block was allocated for the Working Group to present survey instrument analysis and key findings. The evaluation of the survey instrument

Table 14: Teacher age by country

	18-29		30-39		40-49		50-59		60 and over	
Country	N	%	N	%	N	%	N	%	N	%
US	4	3%	26	23%	29	25%	41	36%	15	13%
England	2	4%	11	21%	21	40%	15	29%	3	6%
Italy		0%	4	20%	7	35%	8	40%	1	5%
Ireland		0%	7	37%	6	32%	4	21%	2	11%
Scotland	2	11%	5	28%	7	39%	4	22%		0%
Australia	1	7%	3	21%	4	29%	5	36%	1	7%
Malta	1	17%	2	33%	1	17%	2	33%		0%
Total	10	4%	58	24%	75	31%	79	32%	22	9%

within the focus group session was conducted drawing on a combination of data about participant exit points, pilot survey results, evidence of validity and reliability results, and keyword analysis of the survey language. Following this, the group went question-by-question together, viewing the results on a shared projector screen, and discussed whether to leave, remove or adapt questions based on data from teacher responses and analysis insights. The group collectively discussed and decided on whether to: 1) keep the item the same, 2) amend the item, or 3) to remove the survey item. Two working group leaders noted any required modifications and these were synthesised and updated in the survey (resulting in version 2). We present the results of our focus group evaluation in Section 8. The working group reviewed the final amended survey to check that it addressed all required amendments raised in the focus group.

6 PILOT STUDY RESULTS

In this section we present the pilot results of our country reports (intended curriculum) and descriptive results of our pilot survey (n=244) from participants across seven countries (enacted curriculum).

6.1 Intended Curriculum Observations

Within the pilot sample, England was the first to endorse a CS curriculum in 2014. To date, formal curriculum (or standards/ frameworks) have been endorsed across all countries, except Ireland, Italy and in the US where it is state-dependent. Although some countries have national CS curricula, there are observed variations regarding formal reporting of student learning outcomes in CS. In a study of Australian teachers [115], it was found that teacher self-efficacy increased with formal reporting requirements as teachers had developed more experience in assessing student learning. This suggest that this is something worth investigating and monitoring across countries.

Although some locations, such as Colorado, England and Malta, indicated that they have compulsory CS training for primary and secondary teachers, it is clear that this is not something that has been standardised across other regions, irrespective of a formal curriculum being introduced. Additionally, pre-service teacher training is only provided in England, with a majority of other locations having this as an optional study elective at this stage.

Findings from the Working Group responses categorise the CS curricula into three broad types: those with a state plan for CS in place, those with no state plan for CS in place and those whose CS state plan is in development. CS guidance for those with a state plan was either through standalone delivery or embedded across disciplines. All teachers have flexibility of implementation within their state plan curricula. They all have the opportunity to plan delivery of lessons and choose resources. Lesson structure, delivery and content is not prescribed.

All of the countries, except the US, cover some aspect of the CS concepts presented in Table 6. Within the US, we can see that Illinois does not cover any of the concepts explicitly but Colorado does. In four out of the nine countries/states with a K-6 national/ plan, all cover "Computational Thinking", "Algorithms and Programming" and "Impact of Computing". In the seven countries/states with a state plan for students Year 7 onwards, curriculum concepts include "Computational Thinking", "Computer Systems", "Networks and Internets", "Data and Analysis", "Algorithms and Programming" and "Impact of Computing".

We observe that in Table 9 some countries have defined programming languages that are to be taught at specific year levels and others have not. Australia, England, Italy and Scotland have all defined programming languages from primary years of schooling. Programming Languages for those countries/states with a national/state plan use Visual Programming through K-6. From Year 7 onwards, General Purpose Programming is used, moving to OOP in later grades.

6.2 Enacted Curriculum Observations

Due to the small sample size of this pilot study and the newness of K-12 CS education, we aggregate the reporting of teacher demographics and descriptives as a model for exploring this data. However, we acknowledge that teacher preparation, expectations, and experiences in CS varies across across grade bands (i.e., primary, middle years, and secondary) may impact on enacted CS curriculum. Survey administrators can choose to break up the reporting of grade bands. Future research using the survey will investigate results across these grade bands for more useful and detailed reporting.

Tables 16, 17, 18, 19 provide insight from the survey on what teachers' enacted curriculum looks like across countries, including what they are using to teach CS, what resources they perceive is needed to teach CS, who they are teaching, and what they are teaching. For instance, all seven countries were similar in that the

most common workplace equipment teachers used to teach CS were desktop (n=183) and laptop (n=128) computers, with smart phones being the least utilised, if at all (see Table 16). Although Table 17 shows the two most commonly selected needed resources across countries were classroom lesson resources (n=136) and professional development (n=124), looking more closely at individual countries responses illustrates a wide array of contexts. For example, teachers in England and Scotland selected wanting more CS-specific technology (n=30 and n=14) more often than other resources, compared to Malta who selected non-CS specific technology equipment as their most common needed resource.

Across the countries, the most common areas of expertise teachers selected were computer science (n=180) and ICT (n=111), followed by Math and numeracy (n=66) (see Table 15). Some teachers stating that they did not teach CS may actually teach aspects of CS. Possibly, the courses they teach are called STEM, robotics, or something else, but they include aspects of programming and computational thinking into their course.

The levels teachers are instructing are also important to consider when looking at teacher pedagogy and what occurs within the classroom, as well as when making comparisons to intended curriculum, since it often varies depending on the level and/or age range of students. Across countries, lower secondary (13-15 years old) (n=156) and secondary (16-17 years old) (n=150) teaching year levels were the two most common selections (see Table 18) in our pilot survey. In contrast, more respondents from Australia work with upper primary (11-12 years old) while Italy had more teachers working with secondary and senior secondary (18-19 years old) levels.

The variety in student levels also helps to explain the breadth visible in the content taught (see table 19). Across countries, programming skills and concepts (n=219) and algorithms (n=204) were the most frequent CS content being taught by teachers whereas machine learning (n=46) and artificial intelligence (n=67) were the least commonly taught (see Table 19), reflecting some of the similarities in the intended curriculum across these seven countries. There are some differences in enacted curriculum across countries, for example, robotics, which is taught by a higher percentage of teachers in US and Australia than other countries. Moving forward, we anticipate that breaking down the content taught by each country and comparing it to the intended curriculum there can highlight how teachers are enacting and perceiving the intended curriculum, and provide some focus for future resource development. Further analysis and filtering of the data by year level may provide additional insights into the enacted curriculum, as well as offering a way to make connections to each country's intended curriculum.

7 EVIDENCE OF RELIABILITY AND VALIDITY

Many of the survey items are designed to gather descriptive data (questions 2-45, 48-53), while two are designed to measure specific constructs (46, 47). As we considered how to gather evidence of reliability and validity of the instrument in its entirety, we considered the different types of data collected and how to best measure it.

7.1 Measuring the Instrument for Reliability

Reliable instruments yield the same (or statistically significant "same") results each time they are taken. To be as comprehensive as possible with measuring the reliability of the instrument, we selected the following three tests that would be appropriate for this instrument:

- Internal consistency reliability,
- Inter-rater reliability, and
- Test-retest reliability.

7.1.1 Internal consistency. For all questions that are designed to measure a single construct (#46-Motivation, #47-Self-Esteem and Confidence), Cronbach's α (alpha) can be used to measure their internal consistency [29, 30]. In order to state that these constructs (or factors) are reliable, a Cronbach's α of 0.80 or higher indicates good reliability and a Cronbach's α of 0.70 or higher indicates acceptable reliability [30]. Items that contribute to a lower reliability can be examined and dropped as appropriate.

Motivation. The Motivation construct (Question #46 in the Pilot Survey) is designed to capture what motivates teachers to select particular programming environments and languages to use with their students. It uses a 5-point Likert scale (Strongly Agree to Strongly Disagree) to capture this across the 11 items that are part of this construct. We coded Strongly Agree as 1, Agree as 2, Neutral as 3, Disagree as 4, and Strongly Disagree as 5. We also removed the "Other" item since it was not evaluated on a Likert scale. When we conducted Cronbach's α on this construct, we measured Cronbach's α to be 0.78.

Though this is somewhat lower than what we would like, it still is within the acceptable range. Due to our limited time and scope, we did not analyse this any further. However, we recommend that if someone utilizes this survey, they consider performing a factor analysis to determine if dropping one or more of the items in the construct would improve its reliability.

Self-esteem. The Self-Esteem construct is based on the items in the Bergin Self-Esteem survey as discussed in Section 5.2.2 [14, 87]. Bergin's self-esteem survey was developed and validated over 14 years, in multiple studies [88, 89, 91]. The Bergin self-esteem survey was based on the Rosenberg self-esteem questionnaire, but modified to reflect a student's perception of their programming ability [94]. Cronbach's α values were compared to investigate if the Bergin self-esteem questionnaire had an equivalent or greater internal consistency than the Rosenberg self-esteem questionnaire. The Cronbach's α values for the unmodified Rosenberg self-esteem questionnaire were in the range of 0.82 to 0.88 and for the modified Bergin questionnaire, the alpha value was 0.91. When we conducted Cronbach's α on this construct with teachers, we measured Cronbach's α to be 0.89, which indicates good reliability.

7.1.2 Inter-rater reliability. Some questions in the pilot survey require coding the portion of each question that contain an "Other" response with a textbox (Questions 4, 5, 9, 15, 18-21, 26-36, 38-46, 48, 49, and 51 on the pilot survey). For the pilot survey, we presented a number of these textboxes in order to determine if we were presenting adequate options in the selections for each. Instead of conducting a formal coding process for these textboxes, we chose instead to do a keyword count. This decision was made

Table 15: Frequency of teachers with other core subject area teaching expertise by country

Country	English, Literacy	Math, Numeracy	Computer Science	ICT	Design and Technology	The Arts	Languages	Sciences	Physical Education & Health	All areas
US	12	47	85	30	31	5	4	15	4	15
England		1	46	38		1	3	2		3
Italy	1	5	13	7	2	1	2	3		2
Ireland		9	13	11				2	1	5
Scotland		1	18	15	1					2
Australia	2	3	3	6	2	1	1	3		10
Malta			2	4						2
Total	15	66	180	111	36	8	10	25	5	39

Table 16: Frequency of reported classroom equipment usage by country

Country	Laptop	Desktop	Tablet	Smart Phone
US	64	78	38	16
England	17	44	15	2
Italy	10	15	6	6
Ireland	12	17	8	4
Scotland	11	18	5	3
Australia	11	6	10	
Malta	3	5		
Total	128	183	82	31

based on the fact that there were no more than 10-15% of open responses for each of these questions and with the small number of responses for each, coding did not seem an appropriate way to present their results. For future versions of the survey, if more than one researcher codes these answers, inter-rater reliability should be checked with the goal of achieving a consistency at or above 90% [26, 74, 109].

7.1.3 Test-retest reliability. The test-retest reliability measure provides confidence that if given the same survey instrument at different times, participants will answer the items the same [109]. For example, if we were to conduct our survey March 1 through March 14 and then send the survey to the same participants on April 1 (with the second survey perhaps ending on April 14), the aggregated results would be statistically the same for each question.

Given the fact that our initial survey was a pilot study, we did not measure test-retest reliability. However, we would advise those who choose to use the evaluation instrument in the future to consider testing for reliability using the test-retest reliability measure. We recommend using the Pearson Correlation Coefficient, with the goal of achieving 0.70 or above.

7.2 Measuring the Instrument for Validity

Measuring the validity of instruments helps determine to what extent it measures what it was developed to measure. For measuring validity for this instrument, we selected four tests that would be appropriate for this instrument:

- Construct Validity (measured by Face Validity and Concurrent Validity),
- Population Validity, and
- Sampling Validity.

Each of these measures are described below.

7.2.1 Construct validity. Construct validity can be measured for actual constructs with an instrument, with we have two (motivation (Q46 in Pilot Survey) and self-esteem (Q47 in Pilot Survey)). According to Westen and Rosenthal, construct validity is "...an estimate of the extent to which variance in the measure reflects variance in the underlying construct." [120, p. 609]. We use Drost's definition of Construct Validity and choose one Translational measure (Face Validity) and one Criterion-Related Validity measure (Concurrent Validity) to provide evidence of construct validity [42]. Each of these are described below.

Face validity. Face validity determines whether or not the instrument appears to measure what it is intended to measure [109]. For the pilot survey (see Appendix A), gathering evidence for face validity consisted of a construction of the items in the instrument through its careful development by the ten researchers (as discussed in Section 5.2.2), all of whom have experience constructing instruments within the context of CS education. Each researcher reviewed the entire instrument more than once, offering constructive critiques of questions and items within each question. When needed, additional discussion was held about the particular items and questions. The pilot survey went through a continual set of revisions based on this input before it was finalized and launched.

After the pilot survey was launched and the data was analysed, we reviewed the entire survey question by question to critically analyse each and determine if they should remain in the survey, be modified, or be removed. This was performed by the entire Working Group (all ten researchers) and is discussed in detail in Section 8.

We also recommend that face validity should be conducted on any future versions of the instrument. We recommend that the instrument be evaluated by others who are not authors on this paper (part of the original team) but by six to eight other experts in educational research (with a minimum of one from each country in which the study will be conducted). Face validity should be conducted prior to the release of the survey, and feedback provided

Table 17: Frequency of teachers reporting needed resources by country.

Country	Non-CS specific technology equipment	CS-specific technology	Improved technology infrastructure	Professional development	Classroom lesson resources	CS Professional Mentor	School collaboration	Support for classroom research
US	26	52	29	52	62	35	43	29
England	13	30	16	27	24	18	20	24
Italy	4	6	6	10	7	6	7	6
Ireland	5	8	6	13	14	10	9	5
Scotland	8	14	12	10	13	3	7	3
Australia	1	4	5	10	13	7	5	6
Malta	4	2	3	2	3	2	1	3
Total	61	116	77	124	136	81	92	76

Table 18: Frequency of teachers teaching at year level bands by country.

Country	Pre-primary (3-5 years old)	Junior Primary (6-7 years old)	Primary (8-10 years old)	Upper Primary (11-12 years old)	Lower Secondary (13-15 years old)	Secondary (16-17 years old)	Senior Secondary (18-19 years old)
US	5	15	17	36	76	70	56
England	5	10	12	33	38	36	27
Italy	1	2	4	2	8	14	13
Ireland		2	5	7	13	13	8
Scotland	1	1	2	9	16	15	6
Australia	3	6	9	12	3	1	
Malta		2		2	2	1	1
Total	15	38	49	101	156	150	111

Table 19: Frequency of teachers reporting content taught by country.

Country	Programming skills and concepts	Algorithms	Cybersecurity	Robotics	Artificial Intelligence	Machine Learning	Networks and Digital Systems	Information Systems	Web Systems	Hardware	Ethics	Data representation	Privacy	Databases	Data analysis and visualisation	Computational Thinking	Design process (or Design Thinking)
US	100	94	66	54	34	24	52	40	44	70	86	78	74	31	50	83	83
England	52	52	43	17	23	12	47	30	32	47	46	46	40	37	23	50	28
Italy	16	14	7	8	2	1	8	6	10	11	7	9	8	9	5	9	4
Ireland	19	13	3	8	6	5	3	4	7	13	11	10	8	8	5	13	11
Scotland	18	18	13	2	1	2	7	13	17	17	10	18	11	16	2	16	10
Australia	11	11	10	11	1	1	9	7	5	4	4	8	9	2	4	8	12
Malta	3	2	1	3		1	1	2	1	3		2	1	1		1	1
Total	219	204	143	103	67	46	127	102	116	165	164	171	151	104	89	180	149

by these experts should be carefully considered and incorporated into the survey as appropriate.

Concurrent validity. We set out to determine concurrent validity–that is, to compare the results of these responses (like Cronbach's α or other psychometric data) against results of existing instruments previously studied that also measure similar items. We selectively chose two of the constructs to compare against other data to determine if the results from these questions align with results from other similar data.

For the self-esteem construct (Question #47), this set of ten questions came directly from the 2006 Bergin Survey instrument [14]. This survey was adapted by Bergin from the [93] self-esteem scale to be contextualised in the programming context. The adapted instrument has found to be used in several previous studies examining CS students' self-esteem in programming with a Cronbach's α of 0.89.

For the Motivation construct (Question #46), which was designed to measure "strong driving motivations for why you select particular programming environments and languages to use with your students", we were unable to find any instrument that measures a similar construct for comparison. We believe these questions were uniquely structured as described in Section 5.2.2 and further analysis is planned on this portion of the instrument.

7.2.2 Population validity. Population validity is conducted to determine the representation of the targeted sample population to ensure its validity across various groups in the context of the analysis being performed. Given that this study is intended to benchmark enacted curriculum from the teacher perspective as well as various demographic data, each of the following groups were examined:

- Countries being targeted (Question #4),
- School location (#6),
- Socio-economic levels (#7),
- Primary and secondary levels (#10), and
- Experience in teaching computing (#17).

Though we conducted population validity for the pilot survey, this validity measure will in large extent be dependent upon the context in which the instrument is being used. For example, if the instrument will only be distributed to secondary school teachers, the K-12 spectrum analysis would not be conducted–or it may be conducted differently, taking into account the actual grade levels being taught. Further analysis will be conducted here for the final report (with reference to Table 20).

Number of teachers per country. For this validity measure, we recommend that each country is represented by the number of teachers in each country. Table 21 shows the total number of teachers from each country as of 2018. We calculated the percentage of teachers in each country and compared that to the study participants and ran a goodness-of-fit test using Cramer's V, which resulted in $\chi^2(6,N=196)=385.45$, $p < .0001$, $V=0.51$. Given some of the lower values, the Chi-squared approximation may be incorrect. Therefore, though the result indicates a very significant result, and there appears to be a significant relationship between the expected respondents and the actual, this should be considered cautiously. We also know that the expected proportions are not equal across the categories, so Cramer's V is difficult to interpret without further analysis.

Since goodness-of-fit is significant, we conducted a post-hoc analysis to determine the extent to which counts differ from the expected proportions. Standardized residuals indicated that only Australia was below the 1.96 threshold (residual value of 0.06), indicating that it was the only country with adequate representation (see Figure 2). Future administrations of the survey across countries should look at these values carefully and adjust their sample sizes accordingly.

Number of teachers. For the remaining population validity measures, for thoroughness we recommend analyzing teachers in each country. Again, we use England as an exemplar, and run this analysis on only this country. See Table 22. Testing for goodness-of-fit, we find $\chi^2(1,N=52)=17.626$, $p < .0001$, $V=0.58$, indicating significance with a large effect size.

Socio-economic Status. The SES for England is defined as the percentage of pupils that are eligible for free meals, while the number of teachers reported whether or not they taught at a "disadvantaged school". Given that these are two very distinct set of values, we decided a comparison would not be appropriate. However, better definitions of "disadvantaged" in the survey may lead to better reporting and interpretation of results.

School Location. As we worked through the process of analyzing representation across pupils in urban/rural locations in England, we found that there were six classifications for 2018. Reported numbers at the end of Key Stage 4 in England include: Urban - major conurbation (36.3%), Urban - minor conurbation (3.5%), Urban - city and town (43.6%), Rural ares - town and fringe (8.9%), Rural areas - village (5.0%), and Rural areas - hamlet & isolated dwelling (2.8%). For this analysis, we considered representation by labelling the Urban - major and minor conurbation categories as "Metro" and Urban - city and town as "Urban". Table 23 provides a summary.

Testing for goodness-of-fit, we find $\chi^2(2,N=50)=18.291$, $p < .0001$, $V=0.43$, indicating a large effect size. Further analysis was conducted to identify the standardized residuals and determine which survey counts differed from the desired counts. Two (Rural and Remote and Metro) had absolute values greater than 1.96, indicating that they were underrepresented.

CS Teaching Experience. Although we planned on comparing CS Teaching Experience results to national statistics in England, there is currently no data in England on CS teaching experience. Computing only came into the curriculum in England in 2014 and only teachers of A level computing (16-18 year old students) will have more than 5 years' experience. Therefore, though we think this is an important comparison to be made, we find that our survey may be the benchmark data that others may use to compare in future studies.

Looking at the distribution of the respondents in this category shows that the majority of those responding fit in the 4-5 year category, which may reflect the curriculum changes made in 2014. We propose in future usage of the instrument that this data be interpreted through the lens of the national curriculum. In the case of England, we would suggest that additional efforts should be made to include participants who have only taught computing for 1-3 years, in order to get a more representative view of enacted curriculum from those just integrating CS into their curriculum.

Table 20: Proposed Population Validity Measures.

Population	Scope	Measures
Primary/Secondary Spectrum	Country	Ratio based on number of primary and secondary teachers within each country
Countries	International	Ratio based on number of teachers across each country
Disadvantaged/Socio-economic Status	Country	Ratio based on the number of disadvantaged schools within each country
School Location	Country	Ratio based on location of schools within each country
CS Teaching Experience	Country	Ratio based on cs teaching experience within each country

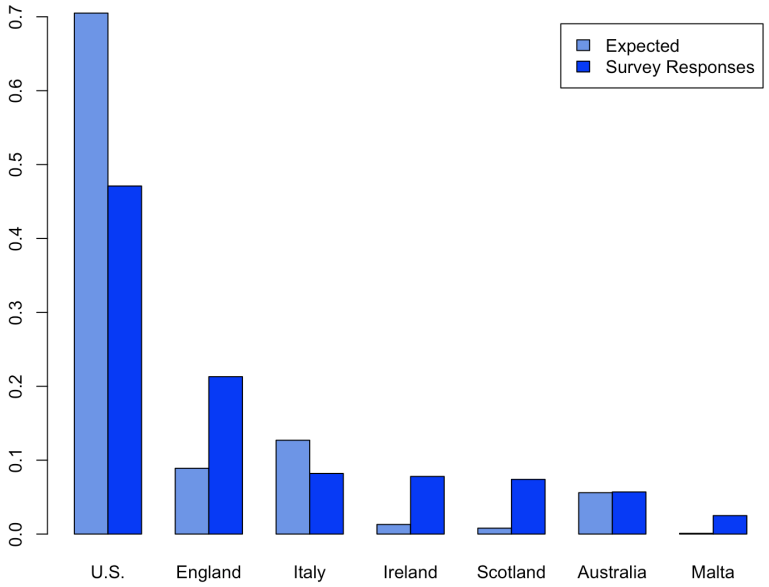


Figure 2: Expected versus Actual Responses across Countries

Table 21: Country Representation

Country	# of Teachers	% of Teachers	# in Study	% in Study
US	3,600,000	70.52%	115	47.1%
England	453,411	8.88%	52	21.3%
Italy	649,495	12.72%	20	8.2%
Ireland	66,327	1.30%	19	7.8%
Scotland	51,138	0.81%	18	1.00%
Australia	288,583	5.65%	14	5.7%
Malta	5,923	0.12%	6	2.5%

Table 22: Primary and Secondary Teacher Representation - England

Level	# of Teachers	% of Teachers	# in Study	% in Study
Primary	222,100	52.16	12	23.08%
Secondary	203,700	47.84	40	76.92%

Table 23: School Location Representation - England

School Location	# of Teachers	% of Teachers	Target (Actual)
Rural and Remote	17	34.00%	16.6%
Metro	7	14.00%	39.8%
Urban	26	52.00%	43.6%

Table 24: CS Teaching Experience Representation - England

Teacher Experience	# of Teachers	% of Teachers
1 Year	3	5.77%
2-3 Years	3	5.77%
4-5 Years	20	38.46%
6-10 Years	14	26.92%
11-15 Years	8	15.38%
16+ years	4	7.69%

7.2.3 Sampling validity. Prior to the pilot study, we reviewed each question and each item to ensure that coverage of the array of variables important to identifying the state of K-12 education internationally was intentional and contained a variety of data. A natural grouping formed, as represented in Table 2. The Working Group leaders started with a small number of categories (or areas to sample), including demographics, current work, qualifications, student composition, and support and resources. As these were opened to the entire group for questions and items to be added to these groups, additional categories were added (assessment of student learning, classroom practice, self-esteem and confidence, and professional development). Based on our collective experiences in CS education, we believe this is a strong sampling across the board.

During the phase of determining the categories and survey items to include in the teacher survey, it was decided by the group to remove certain questions in order to shorten the survey while maintaining a solid sampling across these categories. The following questions were removed in the survey development process:

- Category 2 (Demographics): What is your ethnicity/race?
- Category 3 (Current work): How many years have you taught a core computer science subject (AP Principles, etc)?
- Category 3 (Current work): In the last 12 months, approximately how many students did you teach CS or CT to?
- Category 3 (Current work): Do you lead student teams for competitions?
- Category 5 (Student composition): Is CS a required course at your school?
- Category 6 (Support and resources): Do you know of any out-of-school computing or computer science activities/opportunities in your area?
- Category 8 (Classroom practice): Do you combine CS instruction with other subjects?
- Category 8 (Classroom practice): Do you focus on teaching thinking skills (problem-solving, creativity, decomposition, generalization, hypothesizing, etc.)?
- Category 8 (Classroom practice): Identify the strengths and weaknesses of your curriculum.
- Category 10 (PD): What can the research community do to help sustain your motivation?
- Category 10 (PD): Previous experience in PD, including whether it had a negative experience
- Category 10 (PD): What PD resources were used in the PD that they attended?
- Category 10 (PD): What do you think is needed for you to be effective in introducing CS into your discipline?

In addition, we removed the following constructs from the survey:

- General confidence
- Anxiety
- Motivation (intrinsic and extrinsic)
- Mindset
- Stress in the teaching profession

Though these constructs are valuable in assessing the state of K-12 CS education in schools, they were deemed to not be as critical in this initial international assessment and would add additional time to take an already lengthy survey.

During the discussion of the pilot survey based on the analysis of the data, we further refined the survey by reviewing the results and determining if other questions were or were not needed. Further discussion is contained in Section 8. However, what we discuss here is results of our review within each category of questions to determine completeness.

After reviewing Section 2 (demographics), we determined that the teacher disability question could be optional based on the evaluators' needs. We also determined that we did not capture whether or not the teacher taught in their native language and determined that a question concerning native language would be appropriate in capturing this independent variable that may affect professional development or other factors.

After reviewing Section 3 (current work), we discussed its completeness, and decided to remove questions 12 and 13 after serious discussion on their relationship to CS academic achievement of students.

For Section 4 (qualifications), we decided that a confidence question, "do they feel they are confident in teaching?" belonged in the self-esteem and Confidence category (Section 9).

For the remaining sections (Section 5 through 11), we concluded that these categories were complete.

Upon the conclusion of the review of all of the questions, we again addressed the question "Are there any major categories that we excluded that should or could be included in the categories?" As a group, we came to the conclusion that the survey was complete and comprehensive in establishing benchmark data on enacted curriculum (and support for it) from a teacher perspective.

8 TEACHER SURVEY REVISION

Following our pilot survey and our preliminary analysis and validation of the survey instrument, Working Group members collectively undertook a comprehensive review of the survey instrument taking into the pilot and validation and reliability findings, as well as by conducting further critical analysis of the survey structure. The following sections below discuss our survey revision review.

8.1 Survey Terminology and Language

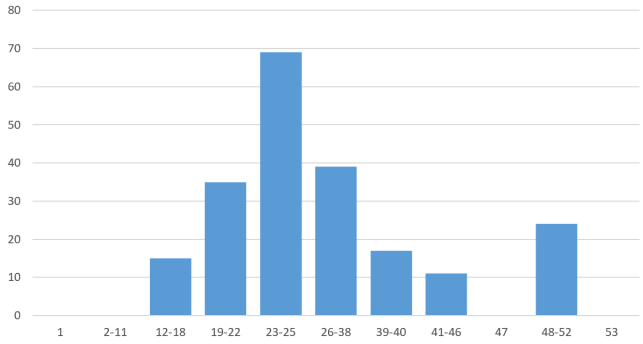
During the analysis of the pilot survey and Working Group discussions, we clarified definitions and synonyms of common terms referenced in the teacher survey. Keyword analysis of each survey item and open-ended responses revealed inconsistencies and words or phrases that needed clarity either referencing synonyms or providing specificity. Based on this, some of the following were adapted: clarification around the term "remote" to include the elaboration "extremely rural" and that "curriculum" refers to "standards/frameworks". Changes were also made to terminology based on focus group discussions, as covered below.

8.2 Survey Sections and Questions

As part of our evaluation, we analysed a breakdown of the survey structure by looking at the number of sections and questions represented across the instrument (as seen in Table 2). We identified

some sections contained larger potions of the survey's total questions. As a number of participants exited the survey early, we were interested in determining the point at which they left to potentially target our attention at areas that may need revision. In Figure 3, we present the frequency of participants (n=210 in total) and their completion of the survey question ranges on the horizontal axis. These 210 participants had made it past the first section of the survey but had dropped out of the survey at some point. Due to Survey Monkey capturing data within page ranges, we can provide the question range at which participants had no further data recorded, indicating they had left. The frequency of participants presented at each range in Figure 3 represents those who had responded to questions within that section but who had left in the following section of questions. The chart shows that a majority of participants exited between questions 26-28 (Section 6: Support and Resourcing, as seen in Table 2). This section contained the highest number of survey questions (n=13, 24.5%), some with lengthy lists of items, suggesting a need to review this section to determine if there are questions that can be revised as well as a possible need to break the section up. The graph also shows a number of participants made it to the end of the survey, however, due to ethics reasons we had excluded their data as they may have wished to exit the survey. Future design of the survey will involve more accurately capturing exit points as well as time taken to complete survey questions and sections for more fine grained analysis.

Figure 3: Question completion bands.



As part of our focus group evaluation of the survey instrument questions, we examined the survey questions that featured an "other" option to determine if there were question items that had not been considered but that should be included within the survey. Based on this process some alterations to our revised survey had been made, such as refining CS technologies used in the classroom as identified by teachers and refinement and clarification of the list of subject areas of expertise in question 15.

8.3 Working Group Evaluation

In this section we present the results of our Working Group evaluation. As mentioned in section 5.6, the evaluation for survey revision was operated as a focus group during the ITiCSE conference during an afternoon block. Following this, the group went question-by-question together to discuss whether to leave, remove or adapt questions based on data from teacher responses and insights. In the following, we discuss the working decisions and amendments made to the survey sections (see Table 2 for the survey outline).

8.3.1 Overall Survey Structure. Because the average completion time for the survey was 1 hour and 14 minutes with the final sample (n= 244), wherever possible we tried to eliminate, combine, or simplify questions. The group agreed to eliminate questions 7,12,13,14,18,21,24,25,36, 38, 40 and 50. The group agreed to move the "qualification" section to follow the demographic section. By removing these questions and making our adjustments, described below, we went from 53 to 46 questions and were able to reduce Survey Monkey's estimated completion time from 41 minutes to 30 minutes.

We discussed making all questions mandatory, or only requiring some compulsory questions as in the pilot survey. The group decided that in order to validate the survey in the future, all questions should be mandatory.

8.3.2 Demographics. In question 6, we decided to combine rural/remote even though in Australia there are extreme rural locations. We also decided to combine the other/unsure category. In question 8, we decided to rename the less than 25 option as 1 to 25 percent. We decided to eliminate the terminology in front of the ages for each grade level band in question 10 because the terminology is different from country to country. The group agreed to leave question 11 about teacher disabilities as an optional question. We reasoned that as CS teacher resources and professional development advances, it will be important to make sure all teachers have accessible resources. The group also decided to add a question to find out if the school where the teacher works is a single-gender or mixed-gender school. The group included a question asking respondents what is their primary native language. Languages listed are from the United Nations Official language list, including Arabic, Chinese, English, French, Russian and Spanish. We also included Maltese and Italian to take into account our cohort countries and have provided an "other" option. The recommendation would be for survey administrators to expand the list, as needed.

8.3.3 Current Work. Regarding current work, we agreed to keep Question 14 asking about total years as a teacher but increase the granularity of answers to less than 1 year, 1-4 years, 5-8 years, 9-12 years, 13-16 years and 17 years or more. The additional responses teachers added in Question 15 helped us to identify the following additional content areas to add: Humanities (Social Studies, Civics); change English/Literacy to English Reading and Composition; change Languages to Foreign Languages/World Languages; and change Design and Technology to Engineering and Design. The group also decided to revisit Question 16 during the revision process.

The group decided to change Question 17 to state: "How long have you taught computer science concepts?" to capture the experience of teachers offering courses that include aspects of CS and full CS courses.

8.3.4 Qualifications. In Question 19, the group decided to add an additional option of "Other Teaching Certificate or Diploma". In question 20, the group decided to remove the word qualification from each answer and add "What other CS teaching qualifications or

professional development have you completed?" with the following options: CS professional development or Bootcamp and CS formal qualification options (e.g. CS Bachelor).

In Question 22, the group decided to rephrase the question as "Have you ever undertaken or participated in CS classroom research?" with the answers: yes, no, unsure. We also decided to add an item in the qualification section, ""I feel qualified to teach CS", with the same seven-point Likert scale options as in question 47 (strongly disagree to strongly agree, in an effort to obtain more information about teacher self-esteem related to preparation.

8.3.5 Student Composition. In the student composition section, the group agreed to change English as a Second Language students to Non-native English speakers and N/A to Unsure. In question 23, the group agreed to revise the percentages for the answers which were based on the TALIS survey [2] to 0%, 1-10%, 11-25%, 26-50%, 51-75%, and 76% or more.

The group discussed asking teachers for school demographics such as the number of CS teachers and the total number of students at the school, but did not come to a consensus and decided to revisit school-based demographic questions.

8.3.6 Support and Resourcing. The group agreed to change the category in question 27 from none to 0%, and add two categories, STEM/ Makerspace and CS Suite/Lab. We decided to add the device options, laptops, desktop computers, tablets, and smartphones to the options in question 35.

The group decided to add (Standards/Framework) after the word curriculum in question 29. In question 30, the group agreed to remove the words "external visit to" from the last two options; add professional organisation to last option; and add Parent, Community, Volunteer as another option.

Regarding question 31, the group chose to rephrase "how much of your time do you spend using unplugged activities?" to be "how much of your time spent teaching Computer Science is done without computers?" with a percentage answer. The group also indicated the need for a new question "How much of your classroom time do you spend teaching computer science?".

In question 32, the group chose to add game design as a category, change Web Systems to Web Dev/Web 2.0, and remove "explicitly" from the computational thinking category.

The group decided to ask a question about computational thinking, although no previously validated questions could be found. A question related to ascertain enacted computational thinking was proposed: "Please check the cognitive and affective skills that you teach: decomposition, generalization, sequencing, logic, abstraction, problem-solving, empathy, persistence, pattern recognition, analyzing solutions, creativity, other."

In question 33, the group decided to add a network/community category.

In question 34, the group decided the answers were too specific. The group decided to change the answers to coding club, community group, University/College Outreach, Competition/Challenges, and Industry mentorship/experience, removing the Other option.

Regarding question 35, the group decided only ask teachers to indicate which resources they use, and remove the requirement to rank their top 3. We decided to remove the words "most commonly" from the instructions. Question 35 would allow teachers to select as many resources as they have used. The group decided to statistically determine which resources teachers used the most instead of asking teachers to take their time and rank the resources.

The group decided in Question 37 to remove the option to choose the top 3 physical computing devices and merge the question into 35. The group decided the answers in question 37 were too specific and they needed to be broader, such as robotics, programmable cards (Raspberry Pi, Microbit), virtual reality devices, STEM kits, and other). The group decided on moving question 35 up closer to the beginning of the survey to hopefully capture these enacted experiences early before exit.

8.3.7 Assessment of Student Learning. In question 39, the group chose to remove descriptions, and change the introductory wording to "commonly use", plus remove the Likert options and replace with two columns, formative and summative. The group decided to explore adding pop-up descriptions of terms with examples to minimize reading time.

8.3.8 Classroom Practice. In regard to question 41, we agreed to elaborate on our description of small group work to include pair programming and related concepts. We agreed to combine and re-structure questions 41, 42 and 43 to revisit categories for granularity and relevance, as well as removing the requirement for participants to address their top 3.

Further revisions to the descriptions in question 45, including the ability to provide further information in hidden tip boxes were agreed. In addition, it was agreed to simplify the scale to "none; a little; a lot", and to remove the optional explanation section.

We are undertaking a further factor analysis of question 46 to determine if we are able to combine or remove categories, in order to simplify this question.

8.3.9 Self-Esteem. Originally, we had referred to this survey section as "self-efficacy" but in our revised instrument, we clarified that the construct is "self-esteem" to align with the use of the instrument in prior work [13]. The group decided that in future work it would be worthwhile to explore the addition of related self-esteem sub-questions, beyond the scope of the Bergin scale[13], to further explore self-esteem and related constructs, such as self-efficacy [7], in relation to the teaching of CS concepts. There are several instruments that have been created to measure explicit self-esteem among students; however, instruments to measure self-esteem in the context of CS are difficult to find, particularly for use with teachers [37].

8.3.10 Professional Development. The Working Group agreed to simplify the operation of these questions, using a combination of simplified check-boxes and drop-down boxes, to record the professional development activities that participants have undertaken in the last 12 months, and to identify the key activity they would like to participate in in the coming 12 months.

In this survey section, we have adopted questions from the TALIS instrument [2]. While some concern was expressed around the complexity of these questions, it was also identified that it was desirable to retain these questions in a format that would aid comparison across other TALIS surveys. The Working Group agreed to explore analysis options and exemplars using TALIS to determine any possible options for simplification.

9 DISCUSSION

The final sample of teachers who completed all survey questions was sufficient for a pilot study but limited. The completed pilot survey was represented the experience of 244 teachers in 7 countries that primarily use English as a native language. The majority, 115 teachers, came from 27 US states. Only 14 teachers completed the survey from Australia and only 6 teachers from Malta. As we offer the survey to additional countries, we will need to reevaluate the item data to ensure the English matches local phrases and terminology for common definitions. If the survey is translated into other languages, it will need to be re-validated[4]. In this section we discuss some key pilot study observations, share our lessons learned during the project and suggest recommendations for future use of the teacher survey instrument.

9.1 Pilot Study Curriculum Observations

Our presentation of pilot results provides a sample demonstration of the insights that can be gathered from the country reports and teacher survey instruments. Although working with a small sample size and aggregated results across year bands for countries, we are able to provide early observations around aspects on intended and enacted curriculum descriptives.

In terms of intended curriculum, our results demonstrate how contextual information gathered through the country report can assist in making comparisons of schooling contexts and CS curriculum requirements. We observe three broad type of CS curriculum implementation that includes those with a state plan for CS in place, those with no state plan for CS in place and those whose CS state plan is in development. Additionally, we observe interesting patterns across countries for curriculum requirements, such as CS topics and programming languages, noting that some countries have defined programming languages in their intended curriculum that are to be taught at specific year levels and others have not.

In terms of enacted curriculum we observe what CS topics teachers are implementing in the classroom, irrespective of their intended curriculum. We identified that programming skills and concepts, and algorithms were the most common CS content being taught, with machine learning and artificial intelligence being less popular, reflecting the intended curriculum. As [68] mentions, teachers are the gatekeeper to CS education as they choose to interpret and/or reject the intended curriculum. As CS becomes more mainstream in schools and teachers move through phases of curriculum implementation, it will be interesting to determine if enacted CS content diversifies and whether content taught more closely aligns with intended curricula.

The survey also captures information about the equipment and resources teachers use in the classroom for CS and their perceived needs. Interestingly, the results highlight differences between countries, demonstrating the value of such comparisons and that a one-size-fits-all approach may not work for making recommendations about how to support teachers with CS curriculum. For example, differences emerged in relation to teachers requesting CS-specific technology in England and Scotland more than any other resource, compared to Malta who identified needing non-CS specific technology.

This report has highlighted some of the early observations via descriptive reporting. However, the power of the instruments will be realised in future analyses where we can align findings from the intended and enacted curriculum, as has been demonstrated in a recent publication of the work comparing teachers' implementation of programming languages and CS topics in comparison to intended curriculum using the METRECC instrument [46]. Additionally, exploring differences in enacted curriculum according to year level bands, in alignment with intended curriculum, will help strengthen our understanding of what is happening in different classroom contexts and required support for primary and secondary teachers. The METRECC instrument not only captures country level reports of intended curriculum, but also enacted curriculum directly from teachers. Up until now enacted curriculum surveys have largely focused on perceptions of what is happening in classrooms [35, 97] or narrow areas of CS such as CT [72].

9.2 Lessons Learned

The analysis of the survey, and the pre-processing of the data highlighted several areas that could be improved or what the group found as successful approaches. This may be of value, when considering future survey tools and the processing of the data set, opposed to the survey instrument itself. This section aims to highlight points that may be of value to the CS education community when considering an international benchmark study/survey, perhaps aimed at K-12. While this may not be applicable in all cases, this Working Group feel that they are of value to highlight, and are in order of appearance and not importance.

9.2.1 Survey Testing/Local Pilot. While the survey was tested, one finding was that some jurisdictions struggled with one or two questions, where the Working Group representative reviewed the survey. Perhaps this may have been a minority, but perhaps a subtest or pilot survey per jurisdiction would have proven useful.

9.2.2 Ethics Approval. The Working Group leaders sought and were granted ethical approval or met requirements to conduct the survey in their jurisdictions. The Working Group recommend that an early investigation be conducted to determine if the process could be coordinated across multiple jurisdictions (if at all possible), and suggest that this would be worth preliminary investigation prior to the individual effort.

9.2.3 Working Group Collaboration. The collaborative approach taken worked very well for this Working Group. During the initial survey design, any question that was validated from another study was added to a referencing repository and collaborative documents (in this case we used Mendeley and GSuite). This expedited the work once the group met in person. Additionally, the use of a real-time collaboration LaTeX environment (in this case, Overleaf) enabled members to easily and concurrently work on the report. The referencing repository also linked into the real time collaboration tool, allowing real time updates of the the bibliography.

9.2.4 Closing the Survey Instrument. When the Working Group closed the survey, the tool used to collect the data, allowed participants to continue. This is not an issue, except when comparing numbers of participants from our downloaded data and the data in

the tool, they differed. Consideration of a cut-off and data download time may be of value when working across multiple contexts and time zones.

9.2.5 A Single Survey. This Working Group developed three instruments as mentioned previously, one for Australia, one for the US and one for International participants. There were very minor differences between the surveys (for example, the landing page and in the case of the US survey compared to the International survey, it had an additional breakdown of state/region). This in itself was not an issue or constraint, however, the tool used produced three separate data-sets. This again was not inherently an issue, but took considerable time to combine, while validating the combined data-sets. A fork or conditional in the survey could have been more efficient. This could be alleviated with a coordinated HREC approval effort. A second, but again minor consideration, was that several participants took a survey that was not their intended survey (from a different region). This perhaps was due to the international profile and reach of some of the Working Group members, where if they promoted the survey, their reach would have included participants from other intended survey jurisdictions. Perhaps if the surveys contained some specific details (which was not the case in this survey), this would have posed more of a problem.

9.2.6 Survey Early Exit. It was noted throughout the data preprocessing that multiple participants exited the survey at varying stages. A number of reasons may have caused this but one being that a survey that takes an hour or more is too long for most people, even when provided the option to only answer compulsory questions. Those who did complete the survey typically completed all questions. Another issue was the way in which the survey collected data, that meant that exiting the survey at any time would exclude the participants' data from being used in the survey collection. This again is a valid method in survey "opt-out" implementation. The tool that the Working Group used saved the data after a minimum of one page was passed. This is one of the main reasons, why the sample size reduced so rapidly. This is not an issue for cleaning the data, but it does highlight that many participants started the survey and exited without completing. This was perhaps a lost opportunity. While the Working Group was able to examine what block the participants exited the survey on, the question that they exited on was not available. For instrument validation, this metric would have had significant value, to focus efforts on particular questions, that may have had a high exit rate. It is acknowledged that there could also have been other reasons why participants exited, such as time constraints, but this data could have had value in investigating this.

An addition (which the tool this working used did not have as a feature or was not easily implemented), which may have to developed as a bespoke survey instrument, was capturing timing between questions and questions blocks as outlined in Table 2. This may also have had value for identifying questions that require further investigation.

9.3 Recommendations for Future Use

The survey instrument can be used across different countries. In order to be able to compare the differences between the prescribed and enacted curriculum, the first step to using the survey instrument would be to complete the country report template as described in section 6.1. The country report is designed to capture the prescribed curriculum in the country where the research will be conducted.

A researcher may choose to use different sections of the survey rather than the complete instrument or apply it to subsets of the K-12 population such as secondary school teachers. If this approach is adopted then population validity (see Section 7.2.2) would need to be conducted in order to determine the representation of the targeted sample population to ensure its validity across these groups.

The survey might need to be translated to the native language of the population it will be administered to. Face validity (see section 7.2.1) would need to be conducted by language speaker experts in order to ensure that the translated survey remains valid.

10 FUTURE RESEARCH

In this Working Group report, we present a pilot survey instrument that has evidence of validity to capture the enacted K-12 CS curriculum across countries, with the addition of associated templates to record intended curriculum in each country. There are a number of avenues for future work that we are exploring as a group.

10.1 Enacted and Intended Curriculum

Our key area of interest in future work is to further explore the data to answer our key research question: What are the similarities and differences in K-12 teachers' perspectives regarding their pedagogy, practice, assessment, resources, and experiences? Once finalised, we will conduct a further survey period within the context of the countries represented by the Working Group members, to look at alignment of enacted and intended curriculum country by country, and the differences and similarities across countries and how these change over time.

10.2 Self-Esteem and Motivation

The two constructs measured as part of this survey, self-esteem and motivation, were only partially analysed as part of this working group study. The time required to a full analysis on these constructs was deemed beyond the scope of this part of the study. However, we acknowledge that further analysis of this data can be conducted. We include some of this future work here.

10.2.1 Self-Esteem. Previous work on exploring related constructs in self-efficacy (a closely related construct to self-esteem) within CS K-12 curriculum has been conducted at individual country levels (e.g., Australia [115]). This was also highlighted in studies mentioned in Section 7, but with CS students. The Self-Esteem construct returned a Cronbach α of 0.89, compared to the previous Bergin CS1 implementation [90], with an α of 0.91.

An interesting process outlined in the development of the Bergin model was the reduction techniques applied to the self-esteem data [14, 90]. Principle Component Analysis (PCA) was used to reduce the Bergin Self-Esteem questions (there were ten within this question) to one value which accounted for as much of the variance in the multiple data points as possible [14, 87]. This would allow for the comparison of Self-Esteem within the Working Group instrument with other factors/data points and additional external data-sets (such as the Bergin CS1 study). Therefore, future work

includes conducting the following studies:

- Correlate Self-esteem with multiple other data points within this instrument, for example:
  - Resources Used
  - PD attended
  - Demographics
  - Country, and
  - Grade Level.
- Compare and contrast teacher self-esteem with that of undergraduate CS1 students.

10.2.2 Motivation for Selected Environment. The Motivation construct (question 46) for teacher selection of particular programming environments and languages consisted of a set of items that were curated by the Working Group prior to releasing the survey. This construct would be interesting to study further and conduct a factor analysis on it. Factor analysis could help determine if any questions are redundant or contains multicollinearity with other questions within this construct. The Cronbach α was 0.78, and perhaps with factor analysis, this could be improved to a value greater than 0.8. Future work could also conduct a a focus group with teachers, using this construct to determine if there are other items that should be added.

Similar to self-esteem, future work includes:

- Conducting a factor analysis on this construct to improve internal reliability and further measure validity
- Correlate Motivation with multiple other data points within this instrument, for example:
  - PD attended
  - CS Teaching Experience
  - Demographics
  - Country
  - Grade Level
  - Self-Esteem

10.3 Data and Further Analysis

Future data analysis, contains a multitude of directions. This can vary from correlation analysis, to perhaps even prediction analysis. In addition, the analysis can be within the data collected with this instrument, or with external data, or a combination of both. Some proposed examples could be as follows:

- Correlations and regression models of data points collected in this instrument.
- Factor analysis of construct's questions
- Demographic Analysis.
- Correlate student CS grades with national tests (such as AP CS exams in the US, or local jurisdiction specific exams), with the enacted curriculum statistical constructs and/or teacher self-esteem ratings.
- A deeper analysis of country specific differences.
- Additional collection of Country Reports, and the analysis of differences that exist between them.
- Analysis of intended curriculum, with the enacted curriculum.
- Analysis of the enacted curriculum with teacher self-esteem, and multiple other factors.
- Additional face validity (and the analysis) to be conducted on the revised survey by sending the survey to 6-8 experts that did not participate in the Working Group to get their feedback.

With the dataset that has been captured there is a significant amount of additional analysis that could be conducted. In addition, this is a snapshot of the current state. As this instrument is in its pilot phase, the following iteration of the instrument will capture additional data. This will allow a further analysis to be conducted. Coupled with this data, the Working Group envision this study to be longitudinal, thus yearly snapshot data will also be collected. This is another avenue to analyse the data, and will perhaps be of value to the CS education community, for tracking changes over time within the K-12 intended and enacted curriculum landscape.

11 CONCLUSIONS

This report describes a detailed collaborative international process for the design, pilot and evaluation of an international survey instrument for Measuring Teacher Enacted Computing Curriculum (the METRECC), in addition to an accompanying country overview report for survey administrators.

We present preliminary pilot results that demonstrate similarities and differences in intended and enacted curriculum implementation across seven countries with 244 teacher participants from seven countries. Early analysis reveals similarities between the intended curricular of countries with a state plan, but differences in the enacted curricular, for example in the area of robotics where there are differences between classroom practice across countries in our study. The variation in the years taught by teachers from different countries has also meant that, given the sample size, only limited meaningful comparisons can be identified at this stage. However as we continue to use the survey instrument, more finely grained findings are expected to emerge.

This report documents a process for checking the survey for evidence of validity and reliability, with recommendations for future survey administers and those interested in expanding the survey for their own contexts. Through our analysis we demonstrate that the survey has evidence of validity and reliability, however, we present a revised version of the METRECC based on our focus group evaluation that was based on a combination of results from the pilot survey, evidence of validity and reliability, instrument question item analysis and participant exit points. Despite research having found teachers to report reasonably accurately in surveys about classroom practice and student behaviour [25], there are recognised nuances associated with self-report measures [55]. As recommended by researchers [20, 21, 25, 85] there is a need for future work to investigate further evidence of reliability and validity with survey instruments in combination with the collection and comparison of alternative data sources, such as teacher observation studies, teacher logs and examples of student work. It would be worth comparing teachers‘ self-reported enacted curriculum with observations of classroom practice to determine how closely they align.

A limitation of this report is that, although the sample size is sufficient for a pilot analysis, it is based on a small sample size in total as well as small sample sizes within each country. However, the primary focus of this Working Group was to focus on the development of a robust survey instrument. This report presents the descriptive results of the pilot survey, however, further work is required to analyse the dataset in more detail. Future work of the Working Group will involve re-launching the revised survey to capture larger quantities of data and to re-check altered questions for evidence of validity and reliability. With a larger dataset, the group will be able to interrogate the data in more detail to determine factors that impact on enacted CS curriculum in schools. With the availability of the study instruments as open source, there are opportunities for researchers to continue to develop, refine and evaluate the survey for more countries.

REFERENCES

[1] 2018. *2018 State of Computer Science Education.* Technical Report. https://code.org/files/2018{_}state{_}of{_}cs.pdf

[2] John Ainley and Ralph Carstens. 2018. Teaching and Learning Int'l. Survey (TALIS) 2018 Conceptual Framework. 187 (2018). https://doi.org/https://doi.org/10.1787/799337c2-en

[3] Mansour M. Al-Sulaiman. 1999. A Computing Curriculum for Technical High Schools in the Kingdom of Saudi Arabia. *J. King Saud Univ. Comput. Inf. Sci.* 11 (Jan. 1999), 85–104. https://doi.org/10.1016/S1319-1578(99)80005-6

[4] Giovanni Apolone and Paola Mosconi. 1998. The Italian SF-36 Health Survey: translation, validation and norming. *Journal of clinical epidemiology* 51, 11 (1998), 1025–1036.

[5] Australian Curriculum, Assessment and Reporting Authority (ACARA). 2015. Australian Curriculum: Digital Technologies. http://www.australiancurriculum.edu.au/. (2015).

[6] Anja Balanskat and Katja Engelhardt. 2014. *Computing our future Computer programming and coding - Priorities, school curricula and initiatives across Europe.* Technical Report. https://doi.org/10.1111/j.1465-7295.200

[7] A. Bandura. 2006. *Guide for Constructing Self-Efficacy Scales.* Age Information Publishing, Greenwich. 307–337 pages.

[8] Albert Bandura. 2006. Guide for constructing self-efficacy scales. In *Self-efficacy beliefs of adolescents.* 307–337. https://doi.org/10.1017/CBO9781107415324.004

[9] Erik Barendsen, NataÅąa Grgurina, and Jos Tolboom. 2016. A new informatics curriculum for secondary education in The Netherlands. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).* https://doi.org/10.1007/978-3-319-46747-4{_}9

[10] Erik Barendsen, Linda Mannila, Barbara Demo, NataÅąa Grgurina, Cruz Izu, Claudio Mirolo, Sue Sentance, Amber Settle, and GabrielÄŰ Stupurienė. 2015. Concepts in K-9 Computer Science Education. In *Proceedings of the 2015 ITiCSE on Working Group Reports - ITICSE-WGR '15.* https://doi.org/10.1145/2858796.2858800

[11] T. Bell, P. Andreae, and A. Robins. 2014. A case study of the introduction of Computer Science in NZ schools. *ACM Transactions on Computing Education* 14 (2014), 1–31.

[12] Carlo Bellettini, Violetta Lonati, Dario Malchiodi, Mattia Monga, Anna Morpurgo, Mauro Torelli, and Luisa Zecca. 2014. Informatics education in Italian secondary schools. *ACM Transactions on Computing Education (TOCE)* 14, 2 (2014), 15.

[13] Susan Bergin. 2006. *A computational model to predict programming performance.* Ph.D. Dissertation. Maynooth University.

[14] Susan Bergin. 2006. *A computational model to predict programming performance.* Ph.D. Dissertation. Maynooth University.

[15] Marie Bienkowski and Eric Snow. 2017. Studying Implementation of Secondary Introductory Computer Science: Pilot Results (Abstract Only). In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education (SIGCSE '17).* ACM, NY, NY, USA, 703–703. https://doi.org/10.1145/3017680.3022432

[16] Jonathan Black, Jo Brodie, Paul Curzon, Chrhttps://www.overleaf.com/project/5d2843014a4c0d57d252435bystie Myketiak, Peter Mcowan, and Laura R. Meagher. 2013. Making computing interesting to school students: Teachers' perspectives. *Annual Conference on Innovation and Technology in Computer Science Education, ITiCSE*, 255–260. https://doi.org/10.1145/2462476.2466519

[17] Rolf K. Blank, Andrew Porter, and John Smithson. 2001. *New Tools for Analysing Teaching, Curriculum and Standards. Results from Survey of Enacted Curriculum Project.* Technical Report. CCSO, Washington, DC.

[18] N Brown, M Kölling, T Crick, S Peyton Jones, S Humphreys, and S Sentance. 2013. Bringing computer science back into schools: lessons from the UK. In *44th ACM technical symposium on Computer Science education.* ACM, Denver, Colorado, USA, 269–274.

[19] Neil C. C. Brown, Sue Sentance, Tom Crick, and Simon Humphreys. 2014. Restart: The Resurgence of Computer Science in UK Schools. *ACM Trans. Comput. Educ.* 14, 2, Article 9 (June 2014), 22 pages. https://doi.org/10.1145/2602484

[20] L. Burstein, L.M. McDonnell, J. Van Winkle, T. Ormseth, J. Mirocha, and G.Guitton. 1995. Validating national curriculum indicators. (1995).

[21] Eric M. Camburn, Seong Won Han, and James Sebastian. 2017. Assessing the Validity of an Annual Survey for Measuring the Enacted Literacy Curriculum. *Educational Policy* 31, 1 (2017), 73–107. https://doi.org/10.1177/0895904815586848 arXiv:https://doi.org/10.1177/0895904815586848

[22] Michael E Caspersen, Judith Gal-Ezer, Andrew McGettrick, and Enrico Nardelli. 2018. Informatics for All The strategy. (2018).

[23] Charalambos Y. Charalambous and George N. Philippou. 2010. Teachers' concerns and efficacy beliefs about implementing a mathematics curriculum reform: integrating two lines of inquiry. *Educational Studies in Mathematics* 75, 1 (01 Sep 2010), 1–21. https://doi.org/10.1007/s10649-010-9238-5

[24] Vanea Chiprianov and Laurent Gallon. 2016. Introducing Computational Thinking to K-5 in a French Context. In *Proceedings of the 2016 ACM Conference on Innovation and Technology in Computer Science Education - ITiCSE '16.* https://doi.org/10.1145/2899415.2899439

[25] Penny CluniesâĂŘRoss, Emma Little, and Mandy Kienhuis. 2008. SelfâĂŘreported and actual use of proactive and reactive classroom management strategies and their relationship with teacher stress and student behaviour. *Educational Psychology* 28, 6 (10 2008), 693–710. https://doi.org/10.1080/01443410802206700

[26] Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement* 20, 1 (1960), 37–46.

[27] Computer Science for ALL Students (CSForALL). 2019. Common Data Collection Survey and Support Document. (2019). https://www.csforall.org/projects_and_programs/rppforcs-resources-for-projects

[28] Jacqueline Corricelli, Seth Freeman, and Chinma Uche. [n. d.]. *CS for CT: Examining the Landscape of Computer Science in Connecticut.* Technical Report. ECEP Alliance.

[29] Lee J Cronbach. 1951. Coefficient alpha and the internal structure of tests. *psychometrika* 16, 3 (1951), 297–334.

[30] Lee J Cronbach and Paul E Meehl. 1955. Construct validity in psychological tests. *Psychological bulletin* 52, 4 (1955), 281–302.

[31] Oracle Academy CSTA. 2014. Results from the CSTA—Oracle Academy 2014 U.S. High School CS Survey:. (2014). https://www.csteachers.org/documents/en-us/51ef16d4-1bfe-44b6-81aa-0f3682baeb79/1

[32] CSTA Research Committee. 2015. *CSTA National Secondary School Computer Science Survey 2015.* Technical Report. https://www.csteachers.org/page/high-school-surveys

[33] P. Curzon, T. Bell, J. Waite, and M. Dorling. 2019. *Computational thinking.* Cambridge University Press. 513–546 pages.

[34] Valentina Dagienė. 2008. Teaching Information Technology and Elements of Informatics in Lower Secondary Schools: Curricula, Didactic Provision and Implementation. In *Proceedings of the 3rd Int'l. Conf. on Informatics in Secondary Schools - Evolution and Perspectives: Informatics Education - Supporting Computational Thinking (ISSEP '08).* Springer-Verlag, Berlin, Heidelberg, 293–304. https://doi.org/10.1007/978-3-540-69924-8_27

[35] Valentina Dagiene, T Jevsikova, Carsten Schulte, Sue Sentance, and N Thota. 2013. A comparison of current trends within Computer Science teaching in school in Germany and the UK. In *International Conference on Informatics in Schools (ISSEP)*, Ira Diethelm (Ed.). Oldenburg, Germany, 63–75.

[36] Joy-Anne D'Anca. 2017. *Mindset and Resilience: An Analysis and Intervention for School Administrators.* Ph.D. Dissertation. St. John's University (New York), School of Education and Human Services.

[37] Adrienne Decker and Monica M. McGill. 2019. A Topical Review of Evaluation Instruments for Computing Education. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education - SIGCSE '19.* ACM Press, New York, New York, USA, 558–564. https://doi.org/10.1145/3287324.3287393

[38] Peter J. Denning, Douglas E Comer, David Gries, Michael C. Mulder, Allen Tucker, A. Joe Turner, and Paul R Young. 1989. Computing as a discipline. *Computer* 22, 2 (1989), 63–70.

[39] Department for Education. 2013. *The National Curriculum in England.* Department for Education Government of UK, Crown, Cheshire.

[40] Marie desJardins and Susan Martin. [n. d.]. CE21—Maryland: The State of Computer Science Education in Maryland High Schools. *Proceedings of the 44th ACM Technical Symposium on Computer Science Education* ([n. d.]), 711–716.

[41] Directorate for Learning and Assessment Programmes. 2019. SEC Syllabus (2019): Computing. https://www.um.edu.mt/__data/assets/pdf_file/0017/292310/SEC09.pdf. (2019).

[42] Ellen A Drost et al. 2011. Validity and reliability in social science research. *Education Research and perspectives* 38, 1 (2011), 105.

[43] Caitlin Duncan and Tim Bell. 2015. A pilot computer science and programming course for primary school students. In *Proceedings of the Workshop in Primary and Secondary Computing Education*. ACM, 39–48.

[44] Carol S. Dweck. 2008. *Mindset: The new psychology of success*. Random House Digital, Inc.

[45] Education Scotland. 2017. Benchmark Technologies. education.gov.scot/improvement/documents/technologiesbenchmarkspdf.pdf. (2017).

[46] Katrina Falkner, Sue Sentance, Rebecca Vivian, Sarah Barksdale, Leonard Busuttil, Monica M Mcgill, and Keith Quille. 2019. An International Comparison of K-12 Computer Science Education Intended and Enacted Curricula. In *Koli Calling*. Koli, Finland, (in publication).

[47] Katrina Falkner and Rebecca Vivian. 2015. A review of Computer Science resources for learning and teaching with K-12 computing curricula: an Australian case study. *Computer Science Education* 25, 4 (2015), 390–429. https://doi.org/10.1080/08993408.2016.1140410

[48] Katrina Falkner, Rebecca Vivian, and Sally-Ann Williams. 2018. An ecosystem approach to teacher professional development within computer science. *Computer Science Education* 28, 4 (10 2018), 303–344. https://doi.org/10.1080/08993408.2018.1522858

[49] Cheri Fancsali, Linda Tigani, Paulina Toro Isaza, and Rachel Cole. 2018. A Landscape Study of Computer Science Education in NYC. Association for Computing Machinery (ACM), 44–49. https://doi.org/10.1145/3159450.3159467

[50] Italo Fiorin et al. 2018. Indicazioni nazionali e nuovi scenari. Documento a cura del Comitato Scientifico Nazionale per le Indicazioni Nazionali per il curricolo della scuola dellâĂŹinfanzia e del primo ciclo di istruzione. (2018).

[51] Luca Forlizzi, Michael Lodi, Violetta Lonati, Claudio Mirolo, Mattia Monga, Alberto Montresor, Anna Morpurgo, and Enrico Nardelli. 2018. A core informatics curriculum for italian compulsory education. In *Int'l. Conf. on Informatics in Schools: Situation, Evolution, and Perspectives*. Springer, 141–153.

[52] Judith Gal-Ezer, Catriel Beeri, David Harel, and Amiram Yehudai. 1995. A High School Program in Computer Science. *Computer* 28, 10 (Oct. 1995), 73–80. https://doi.org/10.1109/2.467599

[53] Adam Gamoran, Andrew C. Porter, John Smithson, and Paula A. White. 1997. Upgrading High School Mathematics Instruction: Improving Learning Opportunities for Low-Achieving, Low-Income Youth. *Educational Evaluation and Policy Analysis* 19, 4 (1997), 325–338. https://doi.org/10.3102/01623737019004325 arXiv:https://doi.org/10.3102/01623737019004325

[54] Walter Gander, Antoine Petit, GÃľrard Berry, Barbara Demo, Jan Vahrenhold, Andrew McGettrick, Roger Boyle, MichÃĺle Drechsler, Avi Mendelson, Chris Stephenson, Carlo Ghezzi, and Bertrand Meyer. 2013. *Informatics Education: Europe Cannot Afford to Miss the Boat*. Technical Report. Association for Computing Machinery &, Joint Informatics Europe ACM Europe Working Group on Informatics Education, New York. 1–21 pages.

[55] Gerald J. Haeffel and George S. Howard. 2010. Self-Report: PsychologyâĂŹs Four-Letter Word. *The American Journal of Psychology* 123, 2 (2010), 181. https://doi.org/10.5406/amerjpsyc.123.2.0181

[56] Paul Gilster and Paul Glister. 1997. *Digital literacy*. Wiley Computer Pub. New York.

[57] Hai Hong, Jennifer Wang, and Sepehr Hejazi Moghadam. 2016. K-12 Computer Science Education Across the U.S. In *Lecture Notes in Computer Science*. 142–154. https://doi.org/10.1007/978-3-319-46747-4{_}12

[58] Anita M. Hubley and Bruno D. Zumbo. 2013. Psychometric characteristics of assessment procedures: An overview. In *APA handbook of testing and assessment in psychology, Vol. 1: Test theory and testing and assessment in industrial and organizational psychology*. American Psychological Association, Washington, 3–19. https://doi.org/10.1037/14047-001

[59] Peter Hubwieser, Michal Armoni, and Michail N. Giannakos. 2015. How to Implement Rigorous Computer Science Education in K-12 Schools? Some Answers and Many Questions. *ACM Transactions on Computing Education* (2015). https://doi.org/10.1145/2729983

[60] Peter Hubwieser, Michail N. Giannakos, Marc Berges, Torsten Brinda, Ira Diethelm, Johannes Magenheim, Yogendra Pal, Jana Jackova, and Egle Jasute. 2015. A Global Snapshot of Computer Science Education in K-12 Schools. In *ITiCSE Working Group Reports*. ACM, Vilnius, Lithuani, 65–83. https://doi.org/10.1145/2858796.2858799

[61] Peter Hubwieser, Sigrid Schubert, Michal Armoni, Torsten Brinda, Valentina Dagiene, Ira Diethelm, Michail N. Giannakos, Maria Knobelsdorf, Johannes Magenheim, and Roland Mittermeir. 2011. Computer science/informatics in secondary education. In *Proceedings of the 16th annual conference reports on Innovation and technology in computer science education - working group reports - ITiCSE-WGR '11*. https://doi.org/10.1145/2078856.2078859

[62] Petri Ihantola, Arto Vihavainen, Alireza Ahadi, Matthew Butler, JÃijrgen BÃűrstler, Stephen H. Edwards, Essi Isohanni, Ari Korhonen, Andrew Petersen, Kelly Rivers, Miguel ÃĄngel Rubio, Judy Sheard, Bronius Skupas, Jaime Spacco, Claudia Szabo, and Daniel Toll. 2015. Educational data mining and learning analytics in programming: Literature review and case studies. *ITiCSE-WGP 2015 - Proceedings of the 2015 ITiCSE Conference on Working Group Reports* (2015), 41–63.

[63] Indiana ECEP Alliance. 2017. *Status of K-14 Computer Science Education in Indiana: Landscape Report*. Technical Report. https://ecepalliance.org/resources/landscape-reports

[64] Barbara Jones and Suzanne L Flannigan. 2006. Connecting the digital dots: Literacy of the 21st century. *Educause Quarterly* 29, 2 (2006), 8–10.

[65] Ilkka Jormanainen. 2018. On Computer Science Major Students' Motivation in a Practically Oriented Robotics Course. In *Proceedings of the 18th Koli Calling Int'l. Conf. on Computing Education Research (Koli Calling '18)*. ACM, NY, NY, USA, Article 29, 2 pages. https://doi.org/10.1145/3279720.3279749

[66] Caroline Kearney and Àgueda Gras-Velázquez. 2015. eTwinning Ten Years On: Impact on teachersâĂŹ practice, skills, and professional development opportunities, as reported by eTwinners. *Central Support Service of eTwinning-European Schoolnet, Brussels. Disponible en:[En línea].[Consulta: 10 de octubre de 2017]* (2015).

[67] Alexander Kurz, Stephen Elliott, Joseph H. Wehby, and John Smithson. 2010. Alignment of the Intended, Planned, and Enacted Curriculum in General and Special Education and Its Relation to Student Achievement. *The Journal of Special Education* 44 (11 2010), 131–145. https://doi.org/10.1177/0022466909341196

[68] Laura R. Larke. 2019. Agentic neglect: Teachers as gatekeepers of EnglandâĂŹs national computing curriculum. *British Journal of Educational Technology* 50, 3 (2019), 1137–1150.

[69] Clifford H Lee and Elisabeth Soep. 2016. None but ourselves can free our minds: Critical computational literacy as a pedagogy of resistance. *Equity & Excellence in Education* 49, 4 (2016), 480–492.

[70] A.H Licht, E. Tasiopoulou, and P. Wastiau. 2017. *Open Book of Educational Innovation*. European Schoolnet, Brussels.

[71] Francesco Maiorana. 2019. Interdisciplinary Computing for STE(A)M: a low Floor high ceiling curriculum. In *Proceedings of the Association for Teacher Education in Europe (ATEE) Spring Conference*. University of Latvia, 37–52.

[72] Linda Mannila, Amber Settle, Valentina Dagiene, Barbara Demo, Natasa Grgurina, Claudio Mirolo, and Lennart Rolandsson. 2014. *Computational Thinking in K-9 Education*. https://doi.org/10.1145/2713609.2713610

[73] Robert J Marzano and John S Kendall. 2006. *The new taxonomy of educational objectives*. Corwin Press.

[74] Mary L McHugh. 2012. Interrater reliability: the kappa statistic. *Biochemia medica: Biochemia medica* 22, 3 (2012), 276–282.

[75] Orni Meerbaum-Salant, Michal Armoni, and Mordechai (Moti) Ben-Ari. 2010. Learning Computer Science Concepts with Scratch. In *Proceedings of the Sixth Int'l Workshop on Computing Education Research (ICER '10)*. ACM, NY, NY, USA, 69–76. https://doi.org/10.1145/1839594.1839607

[76] F. Moller and T. Crick. 2016. A National Engagement Model for Developing Computer Science Education in Wales. In *The 9th International Conference on Informatics in Schools,*. Munster, Germany, 1–13.

[77] Jacqui Murray. 2018. Technology in the Classroom: What is Digital Literacy. TeachHub. com, www. teachhub. com/technology-classroom-what-digital-literacy. (2018).

[78] National Council for Curriculum and Assessment. 2018. Computer Science Curriculum Specification. https://www.curriculumonline.ie/Senior-cycle/Senior-Cycle-Subjects/Computer-Science. (2018).

[79] National Network of Education Research Practice Partnerships. 2019. Research-Practice Partnerships (RPP) for CS common data collection 9-12. Personal communication. (2019).

[80] Adina Nistor, Jesus Clemente-Gallardo, Takis Angelopoulos, Karolina Chodzinska, Maria Clemente Gallardo, Agata Gozdzik, Agueda Gras-Velazquez, A Grizelj, Katrien Kolenberg, Despoina Mitropoulou, et al. 2019. Bringing Research into the Classroom–The Citizen Science approach in schools. *Scientix Observatory* (2019).

[81] Victo Nolet and Margaret J. McLaughlin. 2000. *Accessing the general curriculum: Including students with disabilities in standards-based reform*. Thousand Oaks, CA: Corwin Press, Inc.

[82] Outlier Research & Evaluation. 2017. BASICS Study ECS Teacher Implementation and Contextual Factor Questionnaire Measures [Measurement scales]. (2017). http://outlier.uchicago.edu/basics/resources/MeasuresTeacherImplementation/

[83] Birgit Pepin, Ghislaine Gueudet, and Luc Trouche. 2013. Investigating textbooks as crucial interfaces between culture, policy and teacher curricular practice: two contrasted case studies in France and Norway. *ZDM* 45, 5 (2013), 685–698.

[84] Paul R. Pintrich, David A. F. Smith, Teresa Garcia, and Wilbert J. McKeachie. 1991. A Manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ). *Mediterranean Journal of Social Sciences* 6, 1 (1991), 156—164.

[85] Andrew C. Porter and John L. Smithson. 2001. Defining, Developing and Using Curriculum Indicators. CPRE Research Reports, 12-2001. (2001).

[86] Paige Prescott, Irene A. Lee, and Kersti Tyson. 2019. Teacher Beliefs in Student Capabilities As a Mediating Factor in a Novel Understanding of Enactment of CT Curriculum. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education (SIGCSE '19)*. ACM, NY, NY, USA, 1277–1277. https://doi.org/10.1145/3287324.3293841

[87] Keith Quille and Susan Bergin. 2016. Programming: Further Factors that Influence Success. In *Psychology of Programming Interest Group (PPIG), 7th to 10th Spetember, University of Cambridge*. University of Cambridge.

[88] Keith Quille and Susan Bergin. 2018. Programming : Predicting Student Success Early in CS1 . A Re-validation and Replication Study. In *23rd Annual ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE'18)*. 15—20.

[89] Keith Quille and Susan Bergin. 2019. CS1 : how will they do ? How can we help ? A decade of research and practice research and practice. *Computer Science Education* 29 (2019), 254–282. https://doi.org/10.1080/08993408.2019.1612679

[90] Keith Quille, Susan Bergin, and Aidan Mooney. 2015. PreSS #, A Web-Based Educational System to Predict Programming Performance. *International Journal of Computer Science and Software Engineering (IJCSSE)* 4 (2015), 178–189.

[91] Keith Quille, Natalie Culligan, and Susan Bergin. 2017. Insights on Gender Differences in CS1 : A Multi- institutional , Multi-variate Study . ACM, New York, NY, USA, 263–268. https://doi.org/10.1145/3059009.3059048

[92] Janine Remillard and Daniel Heck. 2014. Conceptualizing the curriculum enactment process in mathematics education. *ZDM* 46 (10 2014), 705–718. https://doi.org/10.1007/s11858-014-0600-4

[93] Morris Rosenberg. 1965. *Society and the adolescent self-image*. Princeton University Press, Princeton, New Jersey.

[94] Morris Rosenberg. 2015. *Society and the adolescent self-image*. Princeton university press.

[95] Daisy W. Rutstein, Yuning Xu, Kevin McElhaney, and Marie Bienkowski. 2019. Developing Implementation Measures for K-12 Computer Science Curriculum Materials. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education (SIGCSE '19)*. ACM, NY, NY, USA, 321–327. https://doi.org/10.1145/3287324.3287424

[96] Alan Schoenfield. 2011. *How We Think*. New York: Routledge. https://doi.org/10.4324/9780203843000

[97] S Schrire and D Levy. 2012. Troubleshooting MOOCs: the case of a Massive Open Online Course at a college of education. (2012), 761–766 pages. http://www.editlib.org/p/40832

[98] Carsten Schulte, Malte Hornung, Sue Sentance, Valentina Dagiene, Tatjana Jevsikova, Neena Thota, Anna Eckerdal, and Anne-Kathrin Peters. 2012. Computer science at school/CS teacher education: Koli working-group report on CS at school. In *Proceedings of the 12th Koli Calling International Conference on Computing Education Research*. ACM, 29–38.

[99] Deborah Seehorn, Stephen Carey, Brian Fuschetto, Irene Lee, Daniel Moix, Dianne O'Grady-Cunniff, Barbara Boucher Owens, Chris Stephenson, and Anita Verno. 2011. *CSTA K-12 Computer Science Standards, Computer Science Teachers Association*. ACM, New York.

[100] Cynthia Selby and John Woollard. 2013. Computational thinking: the developing definition. (2013).

[101] Sue Sentance, Simon Humphreys, and Mark Dorling. 2014. The network of teaching excellence in computer science and master teachers. In *Proceedings of the 9th Workshop in Primary and Secondary Computing Education*. ACM, 80–88.

[102] John Smithson and Rolf K. Blank. 2006. *Indicators of Quality of Teacher Professional Development and Instructional Change Using Data from Surveys of Enacted Curriculum: Findings from NSF MSP-RETA Project*. CCSSO.

[103] Robert H. Stupnisky, Allison BrckaLorenz, Bridget Yuhas, and Frédéric Guay. 2018. Faculty members' motivation for teaching and best practices: Testing a model based on self-determination theory across institution types. *Contemporary Educational Psychology* 53 (2018), 15–26.

[104] Maciej M. Sysło and Anna Beata Kwiatkowska. 2015. Introducing a New Computer Science Curriculum for All School Levels in Poland. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 141–154. https://doi.org/10.1007/978-3-319-25396-1{_}13

[105] The Royal Society. 2012. *Shut down or restart? The way forward for computing in UK schools*. Technical Report. London. https://royalsociety.org/~/media/education/computing-in-schools/2012-01-12-computing-in-schools.pdf

[106] The Royal Society. 2017. *After the reboot: Computing education in UK schools*. Technical Report. The Royal Society, London, United Kingdom. 1–116 pages. royalsociety.org/computing-education

[107] David Thompson and Tim Bell. 2013. Adoption of New Computer Science High School Standards by New Zealand Teachers. In *Proceedings of the 8th Workshop in Primary and Secondary Computing Education (WiPSE '13)*. ACM, NY, NY, USA, 87–90. https://doi.org/10.1145/2532748.2532759

[108] David Thompson, Tim Bell, Peter Andreae, and Anthony Robins. 2013. The Role of Teachers in Implementing Curriculum Changes. In *Proceeding of the 44th ACM Technical Symposium on Computer Science Education (SIGCSE '13)*. ACM, NY, NY, USA, 245–250. https://doi.org/10.1145/2445196.2445272

[109] William Trochim. 2006. Web Center for Social Research Methods. https://www.socialresearchmethods.net. (2006).

[110] Siny Tsang, Colin F Royse, and Abdullah Sulieman Terkawi. 2017. Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. *Saudi journal of anaesthesia* 11, Suppl 1 (2017), S80.

[111] Sepehr Vakil. 2014. A critical pedagogy approach for engaging urban youth in mobile app development in an after-school program. *Equity & Excellence in Education* 47, 1 (2014), 31–45.

[112] Gilbert A Valverde, Leonard J Bianchi, Richard G Wolfe, William H Schmidt, and Richard T Houang. 2002. *According to the book: Using TIMSS to investigate the translation of policy into practice through the world of textbooks*. Springer Science & Business Media.

[113] Jan van den Akker, Wilmad Kuiper, and Uwe Hameyer. 2004. Curriculum Perspectives: An Introduction. In *Curriculum Landscapes and Trends*. Springer Dordrechtds, Dordrecht, Chapter 1, 1–10. https://doi.org/10.1007/978-94-017-1205-7

[114] Maarten van Veen, Fred Mulder, and Karel Lemmen. 2004. What is Lacking in Curriculum Schemes for Computing/Informatics?. In *Proceedings of the 9th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education (ITiCSE '04)*. ACM, NY, NY, USA, 186–190. https://doi.org/10.1145/1007996.1008046

[115] Rebecca Vivian and Katrina Falkner. 2018. A Survey of Australian Teachers' Self-efficacy and Assessment Approaches for the K-12 Digital Technologies Curriculum. In *Proceedings of the 13th Workshop in Primary and Secondary Computing Education (WiPSCE '18)*. ACM, NY, NY, USA, Article 5, 10 pages. https://doi.org/10.1145/3265757.3265762

[116] Rebecca Vivian, Katrina Falkner, and Nickolas Falkner. 2014. Addressing the challenges of a new digital technologies curriculum: MOOCs as a scalable solution for teacher professional development. *Research in Learning Technology* 22 (Aug. 2014). https://doi.org/10.3402/rlt.v22.24691

[117] Jennifer Wang, Hai Hong, Jason Ravitz, and Sepehr Hejazi Moghadam. 2016. Landscape of K-12 computer science education in the US: Perceptions, access, and barriers. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education*. ACM, 645–650.

[118] Chris Watkins and Peter Mortimore. 1999. Pedagogy: What do we know. *Understanding pedagogy and its impact on learning* (1999), 1–19.

[119] Mary Webb, Niki Davis, Tim Bell, Yaacov J Katz, Nicholas Reynolds, Dianne P Chambers, and Maciej M Sysło. 2017. Computer science in K-12 school curricula of the 2lst century: Why, what and when? *Education and Information Technologies* 22, 2 (3 2017), 445–468. https://doi.org/10.1007/s10639-016-9493-x

[120] Drew Westen and Robert Rosenthal. 2003. Quantifying construct validity: two simple measures. *Journal of personality and social psychology* 84, 3 (2003), 608.

[121] Jeannette M Wing. 2008. Computational thinking and thinking about computing. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 366, 1881 (2008), 3717–3725.

[122] Gary KW Wong, Ho Yin Cheung, Edwin CC Ching, and John MH Huen. 2015. School perceptions of coding education in K-12: A large scale quantitative study to inform innovative practices. In *2015 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*. IEEE, 5–10.

[123] Aman Yadav, Sarah Gretter, Susanne Hambrusch, and Phil Sands. 2016. Expanding computer science education in schools: understanding teacher experiences and challenges. *Computer Science Education* 26, 4 (2016), 235–254.

[124] Andreas Zendler, O William McClung, and Dieter Klaudt. 2015. A cross-cultural comparison of concepts in computer science education: The US-Germany experience. *The International Journal of Information and Learning Technology* 32, 4 (2015), 235–256.

[125] Andreas Zendler, Christian Spannagel, and Dieter Klaudt. 2010. Marrying content and process in computer science education. *IEEE Transactions on Education* 54, 3 (2010), 387–397.

Appendices

A PILOT STUDY MATERIALS

We have provided information about the METRECC instrument and have made the following items publicly available at https://csedresearch.org/tool/?id=185:

- a downloadable PDF of our original pilot METRECC survey
- a downloadable PDF of our revised METRECC survey
- our country report template
- our pilot dataset (n=244, with participant approval)