Evaluation and Assessment Needs of Computing Education in Primary Grades

Rebecca Vivian
The University of Adelaide
Adelaide, Australia
rebecca.vivian@adelaide.edu.au

Diana Franklin
University of Chicago
Chicago, IL, USA
dmfranklin@uchicago.edu

Dave Frye
North Carolina State University
Raleigh, NC, USA
dafrye@ncsu.edu

Alan Peterfreund
SageFox Consulting Group
MA, USA
apeterfreund@sagefoxgroup.com

Jason Ravitz
Evaluation by Design LLC
jason@evaluationbydesign.com

Florence Sullivan
University of Massachusetts
MA, USA
fsullivan@educ.umass.edu

Melissa Zeitz
Liberty Elementary School
Springfield Public Schools
Springfield, MA, USA
zeitzm@springfieldpublicschools.com

Monica M. McGill
Knox College & CSEdResearch.org
Galesburg, IL, USA
monica@csedresearch.org

ABSTRACT

Until recently, computer science (CS) has been predominantly taught at upper-secondary or tertiary levels. Lately, however, CS curricula have been introduced into school systems from the very first year of school. In this paper, we undertake a participatory research approach, using focus group discussions between a group of experts in the field of evaluation and assessment at the primary level (K-5). The group considered the evaluation and assessment measures they have used, what their current needs are and how the CS education community can move towards meeting those needs. We present the discussion results as a position paper, situated in the context of broader education research. The experts identified three key priorities for the education research community: creating a universal taxonomy of assessment in the primary grades (K-5), creating measurements of student progression and growth over time, and creating culturally relevant evaluations and assessments. Through identifying key priorities, this work provides direction for urgently needed resource development and research directions for K-5 evaluation and assessment.

KEYWORDS

Primary education, K12, K-12, teachers, pre-service teachers, evaluation, assessment

ACM Reference Format:
Rebecca Vivian, Diana Franklin, Dave Frye, Alan Peterfreund, Jason Ravitz, Florence Sullivan, Melissa Zeitz, and Monica M. McGill. 2020. Evaluation and Assessment Needs of Computing Education in Primary Grades. In *Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE '20), June 15–19, 2020, Trondheim, Norway.* ACM, New York, NY, USA, Article 4, 7 pages. https://doi.org/10.1145/3341525.3387371

ITiCSE '20, June 15–19, 2020, Trondheim, Norway

ACM ISBN 978-1-4503-6874-2/20/06...$15.00
https://doi.org/10.1145/3341525.3387371

1 INTRODUCTION

Until recently, Computer science (CS) has predominantly been taught at senior secondary or tertiary levels, if at all [15, 25], with more recent advancements seeing CS being introduced as a new subject from the primary years (K-5) in a number of countries [19]. Successful implementation of CS curricula will involve the role of quality assessment practices [21]. Teachers require sophisticated assessment skills if they are to identify what their students know and can do in relation to valued learning outcomes and how they can best assist their students in their learning [55]. The development of assessment practices for a new area is a complex endeavour with assessment being intertwined with pedagogical content knowledge [44, 55], which is still developing for K-5 CS. Collaborative efforts that bring stakeholders together around a focused agenda can help to accelerate the process of curriculum implementation through shared expertise and resources. Although there are challenges associated with developing evaluations and assessments across contexts due to variations in CS curricula [21], high-level strategic research can inform local and contextualised work.

During the 2019 International Computing Education Research (ICER) conference, a workshop was held with 26 experienced evaluators, educators, researchers, and other stakeholders to investigate evaluation and assessment needs across CS education [42]. Participants were grouped in one of their selected areas based on expertise and interest: Primary, Post-Secondary, and Teacher. A participatory research approach [5, 29] was adopted, utilising focus group discussions [2, 40] between participants, and engaging participants in the analysis and writing of results. This paper focuses on the findings generated by the authors as participants of the Primary Evaluation and Assessment Needs group (K-5). Our goal was to consider and explore the following questions:

- What are the current gaps and challenges in evaluating and assessing academic achievement of K-5 students in CS?
- What are ways the CS education community can rise to meet those needs?

We discussed two primary paradigms: the evaluation of student learning and observing what is happening in the classroom. We consider *assessment* as encompassing 1) in-class ongoing formative and summative assessment about student learning that are assessed by teachers or researchers [41] and 2) external assessments, such as standardised examinations designed and marked outside individual schools. We also consider *evaluation* of student learning, teachers, processes, and programs to determine impact [50, p. 37]. We consider evaluation and assessment to include cognitive and non-cognitive factors, since there is a large body of evidence that shows that non-cognitive factors impact academic achievement [20, 32]. There are multiple perspectives from which assessment would be of interest depending on the stakeholder group. These are not limited to classroom teachers concerned with their students' progress and impact of teaching, school leadership, education and CS researchers, and government bodies.

These discussions reflect what the group yielded based on their experiences and knowledge. This paper summarizes and reflects upon the discussions about the state of K-5 CS evaluation and assessment and contextualizes these discussions in light of education research. This work is important for CS education researchers and evaluators who want to work towards meeting those needs. By defining the current needs of the community, this information can shape funding organizations and others with vested interest in gathering empirical evidence for improving the state of the field.

2 METHODS

A participatory research approach [5, 29] with in-depth focus group discussions [54] was adopted to engage purposely selected experts (n=7) as data-generators and analyzers based on their experiences and knowledge of K-5 evaluation and assessment in CS. Participants were invited by facilitators based on their discipline expertise (as researchers and/or educators) and availability to attend. Though not a formal study, these processes were followed to add strength to the results of the discussions.

The facilitators designed the research agenda as part of ICER submission requirements; however, participants had control over the research processes (e.g. leading discussions, recording methods, discussion direction). As with participatory research, participants were involved in the analysis of data generated and the synthesis of findings of the research process [5].

Focus group discussion is a participatory method that allows for the discussion to go in any direction [2]. Researchers adopt the role of facilitator, allowing discussions to emerge between participants with researchers on the peripheral [40]. Participants can exchange viewpoints and experiences, collectively identify and organize priority points or concepts, leading to clarification and refinement results [54]. Ethical challenges associated with focus groups concern *consent* due to the unpredictability of the discussion, *confidentiality* due to the nature of the small group setting, and *risk of harm* for sensitive topics [51]. Participants could remove discussion notes from the shared document at any point during the workshop and reporting of discussions did not identify individuals.

The facilitators commenced the full-day workshop with introductions and a presentation on the research agenda, background literature and terminology, providing contextualisation and scope for discussions. One participant at each table was assigned to lead discussions using questions designed by the facilitators. One scribe was assigned to record discussions in an online collaborative document with all members having access to add and refine notes. Facilitators provided resources for collecting and organising data (e.g. paper, post-its, pre-arranged online documents) with each table having choice over discussion processes. The K-5 group used a combination of post-it note and collaborative document brainstorming, in combination with voting to determine priorities and solutions. At regular intervals, each table shared a summary of discussions prompting cross-pollination of insights and feedback between tables of experts. Following the workshop, the group Table Lead synthesized the discussion data into a paper with one facilitator collaborating on the contextualisation of the results in literature. All members had an opportunity to review and contribute to the final position paper. We present the discussion findings below as a position paper, contextualised in education literature.

3 ASSESSMENT & EVALUATION MEASURES

Assessment and evaluation have been formalized for many topics at the primary level and many formal methods and frameworks exist [38, 41, 56]. In our discussions, some members said that their country, state, or workplace was not in the process of formally assessing or evaluating primary CS education, while others such as Australia were undergoing a period of transition [19, 58]. However, there were a number of current formal and informal methods of evaluation and assessment being adopted by the group or that members were aware others were adopting for teaching and research, as shown in Table 1.

Relating to research, various approaches to assessing both students and teachers' capacity to develop assessments were identified that partially align with recommended classroom assessment strategies for K-12 [27, 30] and CS education assessment practices identified in the literature and used for a survey of K-12 teachers' CS assessment practices [58]. With the exception of strategies specifically designed for CS education (e.g. Bebras Computational Thinking Challenges [57]), many types of student assessment measures are well-used across primary education, including familiar methods such as observation, quizzes/tests [1, 3], rubrics and cognitive interviewing [23].

4 GAPS IN ASSESSMENT & EVALUATION

Assessment and feedback strategies discussed earlier by the group served as a backdrop for identifying gaps. Discussions about current evaluation and assessment practices in CS were very "programming" focused, with an expressed need to expand work in primary years to include a broader focus of CS curricula. Several related topics were raised that were deemed important to explore, including:

- Trajectories after interventions to determine changes (e.g. in performance, interest, identity) over time.
- Knowledge of intermediate programming concepts.

Area	Assessment and Evaluation Strategies
Programming or CS Project work	• Rubrics • Automated systems • Analyzing products of coding activities, including automated analysis of projects (e.g. Dr Scratch [37]) • Think-aloud interviews with students for elaboration/thinking process/understanding • Programming assignments (e.g. nifty.stanford.edu) • Reading/explaining code or blocks • Drawing/flowchart prediction of output • Student activity and progress in the online learning management system (LMS) • Student reflections • Sequencing activities (e.g. ordering blocks correctly)
Problem Solving in CS	• Bebras Computational Thinking Challenges • Adapted CS Unplugged [4] scenarios to test knowledge transfer (e.g. [45])
Written Work	• Interviews • Writing samples, written products, responses to open-ended prompts, essay analysis • Rubrics
General	• Checklists • Surveys (short check-points or longer) • Tests/exams/quizzes (e.g. Project Quantum, community.computingatschool.org.uk/resources/4382) • Student activity and progress in the online learning management system (LMS)
Research	• Draw a computer scientist [34] • Observations • Interviews and surveys (students, teachers and parents) for cognitive and non-cognitive factors • Focus groups • Classroom observation (field notes) • Artefact analysis of programming projects, including creativity, modifications, use of blocks (sometimes paired with artefact-based interviews) • Analysis of teacher-created assessments • Qualitative analysis of various data sources, collected ethnographically over time

Table 1: General adopted CS education assessment and evaluation strategies at the primary level.

- Self-efficacy and non-cognitive factors at the early years, particularly exploring pre-literate or early-literacy approaches.
- Impact of primary CS curricula on non-cognitive factors such as biases (e.g. CSTA, CSforAll, Code.org) [9, 11, 13].
- Alternative approaches to assessment for different age groups (e.g. CS Principles uses essays and rubrics, but essays are not suitable for early years).
- Exploring whether students experience curriculum in culturally relevant ways.
- Assessment of unplugged activities and transfer of knowledge and skills.
- Impact of user experience/interface in software/hardware learning environments and unplugged activities.

Reflecting on these gaps and challenges involved in evaluation and assessment, they are categorized as those concerning 1) research, 2) learning and teaching, or 3) research, learning and teaching (see Table 2). It was apparent that there is a clear need to bring together educators and researchers, with systematic and infrastructure support, to work on solving these challenges.

Prior research [61] indicates teachers require support in evaluating student thinking processes, addressing problems related to student programming errors, and misconceptions in programming. One strategy recommended is to help teachers make student thinking visible in assessment. We discussed issues that arise with teachers' lack of confidence or knowledge/skills in CS as impacting the implementation of assessments and the depth of feedback. This aligns with research showing that as teachers are required to start formally teaching, their self-efficacy increased [58], suggesting exposure and practice is one key element of building capacity. Further, we noted the need to provide opportunities for teachers to build confidence in designing, implementing and providing feedback in assessment that is appropriate, accurate and addresses misconceptions. This aligns with research in general assessment literacy which suggests that PD should be offered in assessment and should be targeted and differentiated to teacher's experience levels and career stage [17]. Similarly, pre-service teachers need directed PD and ongoing support in assessment methods [31].

Through brainstorming and sorting contributions, we formed a number of key areas of priority for work in primary evaluation and assessment, including culturally relevant evaluation and assessment, improved alignment of instruments to intended goals, and strategies and efforts to collect, share, remix, and gather evidence

Research	Learning and Teaching	Both
Lack of funding to validate instruments. Difficulty in measuring K-2 (literacy levels and community understanding). Capturing outcomes in terms of qualitative stories of "impact" on student learning or lives.	Standards are distributed across all teachers in schools (unclear how to manage assessment across multiple classrooms and disciplines). Teachers lack skills and knowledge with new CS curriculum.	Pre-literate or early literacy students. Duplicated efforts (e.g. instruments, research) Tracking students longitudinally. Starting point is unpredictable by grade level, never knowing when first exposure is. Mitigating circumstances. School level performance vs. individual students (e.g. structural, access, infrastructure differences). How to assess prior knowledge to interpret gains. Assessment of work across platforms.

Table 2: Identified gaps and challenges in assessment and evaluation

of validity for instruments (Table 3). We separate priorities into two focus areas: work focusing on *construct* research and development, and work focusing on harnessing or developing *methods and tools* for data collection, analysis and monitoring.

The group had a cautionary note in that teachers engaging in evaluation and assessment approaches need to be confident that data about student performance is used to improve student outcomes and teacher growth (e.g. informing PD) rather than to judge teachers individually. Respect for teachers during evaluation is critical for its success, and judging teachers individually may cause serious risk of backlash. This aligns with previous findings that teachers resist and fear summative evaluations, subjective, inaccurate, ineffective, and inconsistent evaluations, and evaluations that do not lead to growth from PD [10, 12, 16, 24, 26].

The group agreed that with the development of assessments, there is a need to focus on aligning what we are assessing with why. This is reflected in curriculum alignment research that suggests student achievement can be improved when there is alignment between what is written (e.g. standards), taught and assessed [53].

It has been previously been shown that assessment approaches used by teachers differ, with new teachers focusing more on standardized assessments that are more summative than formative, and more experienced teachers choosing formative and more equitable methods [17]. With this, teachers need continuing PD on contemporary methods and perspectives on assessment [17, 31]. Measurements of teacher assessment literacy should include factors shown to influence student assessment, particularly career stage, teaching division and previous assessment education [17].

These challenges are not necessarily new and exist at various levels in K-12 education in general, with broader research in education by the Organisation for Economic Co-operation and Development (OECD) [41] noting challenges across governance and implementation, student assessment, and teachers, schools, and systems evaluation.

Key Priority Areas
Construct Areas
Culturally relevant evaluation and assessment (how, who, what and norming assessments for sub-populations).
Capturing and understanding culturally relevant aspects of CS activities.
Advancements in understanding and measurement required for K-2, including pre- or early literacy assessment practices.
Approaches that capture the breadth of CS (e.g. society impact, user design) beyond the scope of programming.
Methods and Tools
Establishment of a universal taxonomy of assessment at K-5 (e.g. non-cognitive and cognitive factors, process vs product, growth vs mastery).
Strategies and efforts to nationally and internationally collect, share, remix and gather evidence of validity for instruments.
Reuse of common measures/constructs to share findings across projects.
Capturing contextual data (e.g. prior knowledge, resources, demographics) as part of assessment and evaluation.
Improved alignment of instruments to intended goals/purposes.
Alignment of teacher and student assessments and resources to better understand the impact of teaching practices and targeted areas for teacher PD.
Development of automated tools to support analysis.
Approaches to measuring student growth/progression over time.

Table 3: Key priority areas for primary evaluation and assessment.

5 FUTURE DIRECTIONS

We decided on a deep-dive into solutions for three key priorities. From the number of critical priorities in Table 3, three were identified as top priorities through member voting. Each member could allocate three votes to a priority, with the priorities receiving the highest number of votes being selected for further discussion. Our position is that these are the key priority areas that must be investigated thoroughly and within the next two-three years to establish their foundation in K-12 computing education research. These areas can provide a solid foundation for future work and were perceived as most urgently needed to advance K-5 assessment and evaluation.

5.1 Universal taxonomy for K-5 CS assessment

The first key priority is to establish a taxonomy of assessment, including both cognitive and non-cognitive factors, for K-5 learners that could be used across contexts and countries. This taxonomy would address key elements, including knowledge and skills, process and product assessment, growth and mastery. A synthesis of existing instruments, frameworks and learning sequences used by both teachers and researchers could lay the groundwork for informing and developing tools that takes into account current practice. For example, Marzano and Kendall (2008) defined the New Taxonomy [35] for assessment that addresses weaknesses in the Bloom's Taxonomy [6]. Some systems to support assessment and evaluation, such as [60], and work by Grover [22, 23] on K-12 CS assessment should be considered as part of the support structure.

To achieve the development of a robust and universal taxonomy, it is important to bring together a diverse, multidisciplinary, international team/committee, representative of a number of countries and year level experts for a design charrette around the different angles that could be assessed in K-5, including developmental and structural considerations. Opportunities provided through working group models (e.g. ICER, ITiCSE, Computer Science Teachers Association (CSTA)) could support this effort [14].

We propose that the group undertake scoping work to build a thorough taxonomy with supporting structures:

- Identify existing taxonomies, frameworks or instruments, building upon categories being assessed/evaluated, including key concepts for K-12 CS.
- Identify cognitive and social developmental considerations for K-5.
- Identify and refine categories and subcategories for the taxonomy that are universally relevant.
- Recommendations for appropriate assessment types for purposes, topic, technologies and guidance on how to remix and build upon existing assessments.
- Recommendations about data privacy and use (e.g. who is able to view/access data from which categories and what level of granularity).
- Encouragement of replication studies (and funding of) to further review and refine assessment instruments.
- Creation of a central resource site where data or tools can be easily assessed and shared (i.e. CSEdResearch.org [36]).
- Creation of ways to promote discussion and review of these resources as they are shared.

This work should result in the update of existing primary school resources to include assessments that are inclusive and culturally relevant. This taxonomy can involve primary school teachers and inform teacher PD. A central site for storing the framework could be on a platform (e.g. CSEdResearch.org [36]).

Assessment and evaluations should not only be customized and appropriate for different contexts, but it also is essential that they take into account the diversity of learners. One area includes incorporating language diversity and addressing how different literacies inform learning and assessment of computing outcomes, including in English, multi-lingual and bilingual learners. [59].

5.2 Culturally relevant evaluation and assessment

We perceived the need to address culturally relevant assessment and evaluation in K-5 CS education as a high priority and over the development of a "one size fits all" approach. This supports the many initiatives in the CS for All movement to be inclusive and diverse [13]. Assessment and evaluations should not only be customized and appropriate for different contexts, but it also is essential that they take into account the diversity of learners. Decades of research in education acknowledges and supports this need in other subject areas at the primary level, including work that considers its overall need across various subcultures (e.g. race/ethnic groups) [7, 28, 43, 47, 52].

For this priority area, we recommend bringing together international experts in equity, culturally and linguistically relevant teaching to promote positive identity/perceptions of CS in K-2 and 3-5 teaching, CS education, and assessment development for various subpopulations. The following activities by the experts would lay a sound foundation for work in this space:

- Collect examples of culturally relevant CS assessments.
- Create guidance documents around strategies for culturally relevant assessment.
- Evaluate/modify existing assessments to include culturally relevant considerations.
- Consider a guided process/tool for assessment development for teachers and/or researchers/evaluators.

This expert group could build on existing efforts in culturally relevant CS and provide an extension to culturally relevant researcher-practitioner partnerships (RPP) for CS assessment and evaluation by forming a K-5 network on the topic (e.g. modelled from a recently funded U.S. National Science Foundation (NSF) project #1923136).

The outcomes of this group would be a report, inclusive of:

- Determine what culturally relevant sustaining pedagogy evaluation and assessment looks like at K-2 and 3-5, with examples.
- Raise awareness of equity issues and establish a shared value around culturally relevant assessment and evaluation.
- Determine methods to measure bias in children for CS (over time) and the impact culturally relevant activities and environments can have on biases across ages.
- Determine the dimensions of cultural relevance that we care about assessing (in curriculum, teaching, and how that is reflected in student experiences) (e.g. reviewing the NYC framework for culturally sustaining pedagogy [39] and culturally relevant scorecard).
- Implement research into understanding and designing effective assessments, and leverage research in other closely-aligned learning areas (e.g. STEM).

To further expand work by the expert group, the funding of a design charrette series bringing together researchers, content creators, community groups, educators, to create culturally responsive curriculum/pedagogies and assessments for CS would be valuable.

This work would also consider the development of activities and instruments to measure impact of culturally relevant CS interventions on students in K-2 and 3-5 (e.g. biases, perceptions) and where changes occur. To add strength to the output of this group, work developed should be reviewed by teams with knowledge of cultural relevance, including special education, subpopulations, etc.

5.3 Measurement of student progression and growth longitudinally

As countries establish new CS curricula for K-5, there is a need to understand what learning progressions look like for different learners at different stages in their education, as well as how to measure growth over time. This can extend upon work Seiter and Foreman (2013), Seiter (2015), DeLyser (2016), and Grover (2014, 2017) already started in computational thinking and programming [18, 22, 23, 48, 49], as well as frameworks and tools used in other subject areas like art, mathematics, and life sciences [8, 33, 46].

To address this key area, we recommend separating K-2 from years 3-5 due to differences in students' literacy and ability. In particular, developing assessments and evaluations suitable for K-2 students is an area requiring significant work. Permissions systems and infrastructure for data privacy and policy are needed so teachers can assess without risk and not be penalised for what their students might be lacking in knowledge or skills (e.g. teacher sees individual student performance, but school leader sees aggregate).

To start, we recommend establishing an international network or communication system between researchers and practitioners in K-5 CS education focused around assessment and evaluation. In the first stages of scoping and development, we suggest:

- Consider how K-2 differs from 3-5 (e.g. CS literacy, biases, cognitive development).
- Determine how to measure and evaluate the early years levels (e.g. K-2), particularly pre-literate, skills, practices and content knowledge in CS.
- Unpack and flag longitudinal metrics/measures, including existing instruments.
- Build trajectories for growth, organized around 1) specific concepts/practices in CS, and 2) CS learning opportunities (informal and formal, content-integrated and standalone).
- Provide a path for low-stakes assessment of student learning without judgment or connection to teacher, school, or student performance to facilitate greater use of shared tools.

Data collection and tool development would be a logical next step, with the following aspects needing to be addressed:

- Determine ways to collect data about K-2 students who might be using tactile/tangible modes of CS.
- Develop tools for capturing the role of context in assessment (teacher self-efficacy and characteristics, resources, prior knowledge, etc).
- Determine metacognitive strategies young learners adopt in CS and which ones can lead to success, particularly through self-assessment.
- Identify the best pre- and post-evaluation methods to measure growth over time in K-5.

Through research in these areas, a broader framework can be established that includes the measurement of student progression in the various areas of CS education, from exposure onset in Kindergarten through the end of primary grades.

6 CONCLUSION

This position paper engaged seven experts in participatory research to discuss primary years (K-5) CS assessment and evaluation, generating discussion findings relating to practices, challenges, needs and solutions. CS education in the primary levels is new for many teachers and students across the world. As a CS education community we have a limited scope of knowledge of how to teach and assess at these year levels, and many of the priority areas and recommendations discussed by the group centred around establishing a strong foundation in K-5 assessment from which to build. This work is limited to a small group of individuals with only one external to the US; however, the benefits of a participatory focus discussion group afforded in-depth discourse about the current state, challenges, organisation of priorities and possible solutions.

By starting with the fundamentals such as establishing learning progressions, key assessment and evaluation categories, and leveraging work that already exists, the CS education community can start to form an idea of what assessment and evaluation at these year levels looks like in order to build a platform from which a variety of support resources and research projects can emerge.

7 ACKNOWLEDGEMENTS

This material is based upon work supported by the U.S. National Science Foundation under Grant Nos. 1757402, 1745199, and 1933671.

REFERENCES

[1] Tuba Acar-Erdol and Hülya Yıldızlı. 2018. Classroom Assessment Practices of Teachers in Turkey. *International Journal of Instruction* (2018). https://files.eric.ed.gov/fulltext/EJ1183343.pdf
[2] Michael Agar and James MacDonald. 1995. Focus Groups and Ethnography. *Human Organization* 54, 1 (3 1995), 78–86. https://doi.org/10.17730/humo.54.1.x102372362631282
[3] Vjollca Ahmedi. 2019. Teachers' Attitudes and Practices Towards Formative Assessment in Primary Schools. *Journal of Social Studies Education Research* (2019). https://files.eric.ed.gov/fulltext/EJ1229397.pdf
[4] Tim Bell, Jason Alexander, Isaac Freeman, and Mick Grimley. 2009. Computer Science Unplugged: school students doing real computing without computers. *The NZ Journal of Applied Computing and Information Technology* 13, 1 (2009), 20–29.
[5] Jarg Bergold and Stefan Thomas. 2012. Participatory research methods: A methodological approach in motion. *Historical Social Research* 13, 1 (2012). https://doi.org/10.17169/fqs-13.1.1801
[6] Benjamin S Bloom, MD Englehart, Edward J Furst, Walker H Hill, and David R Krathwohl. 1956. Taxonomy of educational objectives: Handbook I. *Cognitive domain. New York: David McKay* (1956).
[7] Río Piedras Campus. 2009. Systematic Evaluation of UPR Libraries: Reaffirms Collaboration, Promotes Evaluation and Assessment Culture. In *Qualitative and Quantitative Methods in Libraries.*
[8] Kefyn Catley, Richard Lehrer, and Brian Reiser. 2005. Tracing a prospective learning progression for developing understanding of evolution. *Paper Commissioned by the National Academies Committee on test design for K-12 Science achievement* (2005), 67.
[9] Code.org. 2019. Code.org. Retrieved June 21, 2019 from https://code.org
[10] MNCT Coimbra. 2013. Supervision and evaluation: Teachers perspectives. *International Journal of Humanities and Social Science* 3, 5 (2013), 65–71.
[11] Computer Science Teachers Association. 2017. CSTA K-12 Standards. https://www.csteachers.org/page/standards.
[12] Sharon Conley and Naftaly S Glasman. 2008. Fear, the school organization, and teacher evaluation. *Educational policy* 22, 1 (2008), 63–85.
[13] CSforAll. 2019. CSforAll. Retrieved June 20, 2019 from https://csforall.org

[14] csteachers.org. 2019. Computer Science Teachers Association. Retrieved August 29, 2019 from https://csteachers.org
[15] 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.
[16] Linda Darling-Hammond, Audrey Amrein-Beardsley, Edward Haertel, and Jesse Rothstein. 2012. Evaluating teacher evaluation. *Phi Delta Kappan* 93, 6 (2012), 8–15.
[17] Christopher DeLuca, Adelina Valiquette, Andrew Coombs, Danielle LaPointe-McEwan, and Ulemu Luhanga. 2018. Teachers' approaches to classroom assessment: A large-scale survey. *Assessment in Education: Principles, Policy & Practice* 25, 4 (2018), 355–375.
[18] Leigh Ann DeLyser, Bryan Mascio, and Kelsey Finkel. 2016. Introducing Student Assessments with Evidence of Validity for NYC's CS4All. In *Proceedings of the 11th Workshop in Primary and Secondary Computing Education*. ACM, 17–26.
[19] 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*. ACM, 111–142.
[20] Camille A Farrington, Melissa Roderick, Elaine Allensworth, Jenny Nagaoka, Tasha Seneca Keyes, David W Johnson, and Nicole O Beechum. 2012. *Teaching Adolescents to Become Learners: The Role of Noncognitive Factors in Shaping School Performance–A Critical Literature Review.* ERIC.
[21] D. Giordano, F. Maiorana, A. Csizmadia, S. Marsden, C. Riedesel, and S. Mishra. 2015. New horizons in the assessment of computer science at school and beyond: Leveraging on the ViVA platform. In *ITiCSE-WGP 2015 - Proceedings of the 2015 ITiCSE Conference on Working Group Reports.* https://doi.org/10.1145/2858796.2858801
[22] Shuchi Grover and Satabdi Basu. 2017. Measuring Student Learning in Introductory Block-Based Programming: Examining Misconceptions of Loops, Variables, and Boolean Logic. *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education* (2017). https://doi.org/10.1145/3017680.3017723
[23] Shuchi Grover, Stephen Cooper, and Roy Pea. 2014. Assessing computational learning in K-12. In *Proceedings of the 2014 conference on Innovation & technology in computer science education - ITiCSE '14*. 57–62.
[24] Heather Hill and Pam Grossman. 2013. Learning from teacher observations: Challenges and opportunities posed by new teacher evaluation systems. *Harvard educational review* 83, 2 (2013), 371–384.
[25] 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
[26] Brian A Jacob and Lars Lefgren. 2008. Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of labor Economics* 26, 1 (2008), 101–136.
[27] Joint Committee for Standards on Educational Evaluation [JCSEE]. 2015. Classroom assessment standards: Practices for PK-12 teachers.
[28] Robin Jones, Mary Masters, Alison Griffiths, and Nicole Moulday. 2002. Culturally relevant assessment of Indigenous offenders: A literature review. *Australian psychologist* 37, 3 (2002), 187–197.
[29] S. Kindon, R. Pain, and M. Kesby. 2009. *Participatory Action Research: connecting people, participation and place* (routledge ed.). Routledge, United Kingdom. 1–260 pages. https://doi.org/10.1016/B978-008044910-4.00490-9
[30] DA Klinger, PR McDivitt, BB Howard, MA Munoz, WT Rogers, and EC Wylie. 2015. The classroom assessment standards for preK-12 teachers.
[31] Clare Lee and Dylan Wiliam. 2005. Studying changes in the practice of two teachers developing assessment for learning. *Teacher Development* 9, 2 (2005), 265–283.
[32] Jihyun Lee and Valerie J Shute. 2010. Personal and social-contextual factors in K–12 academic performance: An integrative perspective on student learning. *Educational Psychologist* 45, 3 (2010), 185–202.
[33] Bill Lucas, Guy Claxton, and Ellen Spencer. 2013. Progression in student creativity in school: First steps towards new forms of formative assessments. (2013).
[34] C. Dianne Martin. 2004. Draw a computer scientist. In *Proceedings of the Conference on Integrating Technology into Computer Science Education, ITiCSE.* https://doi.org/10.1145/1044550.1041628
[35] Robert J Marzano and John S Kendall. 2008. *Designing and assessing educational objectives: Applying the new taxonomy.* Corwin Press.
[36] M. McGill and A. Decker. 2017. Computer Science Education Resource Center. https://csedresearch.org
[37] Jesús Moreno-León, Gregorio Robles, and Marcos Román-González. 2015. Dr. Scratch: Automatic Analysis of Scratch Projects to Assess and Foster Computational Thinking. *RED. Revista de Educación a Distancia* 15, 46 (2015), 1–23. https://doi.org/10.6018/red/46/10
[38] National Council for Curriculum and Assessment. 2018. Assessment in the Primary School Curriculum. https://www.curriculumonline.ie/getmedia/2b3eaa53-cb4b-4053-9d71-2d28d9d6c734/Assessment-Guidelines.pdf.
[39] New York State Education Department. 2019. NYSED Culturally Responsive-Sustaining Education Framework. http://www.nysed.gov/bilingual-ed/culturally-responsive-sustaining-education-framework
[40] Tobias O.Nyumba, Kerrie Wilson, Christina J. Derrick, and Nibedita Mukherjee. 2018. The use of focus group discussion methodology: Insights from two decades of application in conservation. *Methods in Ecology and Evolution* 9, 1 (1 2018), 20–32. https://doi.org/10.1111/2041-210X.12860
[41] Organisation for Economic Co-operation and Development (OECD). 2019. Evaluation and Assessment Frameworks for Improving School Outcomes: Common Policy Challenges.
[42] Proceedings of the ACM International Computing Education Research (ICER) conference. 2019. Proceedings of the ACM International Computing Education Research (ICER) conference (Toronto, Canada, August 2019).
[43] Lesley Kay Rameka. 2011. Being Māori: Culturally relevant assessment in early childhood education. *Early Years* 31, 3 (2011), 245–256.
[44] Janine T. Remillard. 2000. Can Curriculum Materials Support Teachers' Learning? Two Fourth-Grade Teachers' Use of a New Mathematics Text. *The Elementary School Journal* 100, 4 (3 2000), 331. https://doi.org/10.1086/499645
[45] Brandon R. Rodriguez. 2015. *Assessing Computational Thinking in Computer Science Unplugged Activities.* Ph.D. Dissertation. https://doi.org/10.1017/CBO9781107415324.004
[46] Julie Sarama and Douglas H Clements. 2009. *Early childhood mathematics education research: Learning trajectories for young children.* Routledge.
[47] James E Savage and Alvis W Adair. 1977. Testing minorities: Developing more culturally relevant assessment systems. *The Negro Educational Review* 28, 3 (1977), 219.
[48] Linda Seiter. 2015. Using SOLO to classify the programming responses of primary grade students. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education.* ACM, 540–545.
[49] Linda Seiter and Brendan Foreman. 2013. Modeling the learning progressions of computational thinking of primary grade students. In *Proceedings of the ninth annual international ACM conference on International computing education research.* ACM, 59–66.
[50] K. Shakman and S. M. Rodriguez. 2015. Logic models for program design, implementation, and evaluation: Workshop toolkit (REL 2015–057). http://ies.ed.gov/ncee/edlabs
[51] Julius Sim and Jackie Waterfield. 2019. Focus group methodology: some ethical challenges. *Quality & Quantity* 53, 6 (11 2019), 3003–3022. https://doi.org/10.1007/s11135-019-00914-5
[52] Renee Smith-Maddox. 1998. Defining culture as a dimension of academic achievement: Implications for culturally responsive curriculum, instruction, and assessment. *Journal of Negro Education* (1998), 302–317.
[53] David Squires. 2012. Curriculum Alignment Research Suggests That Alignment Can Improve Student Achievement. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas* 85, 4 (5 2012), 129–135. https://doi.org/10.1080/00098655.2012.657723
[54] Anja P. Tausch and Natalja Menold. 2016. Methodological Aspects of Focus Groups in Health Research. *Global Qualitative Nursing Research* 3 (1 2016), 1–12. https://doi.org/10.1177/2333393616630466
[55] Helen Timperley. 2008. Teacher professional learning and development. *The International Academy of Education* 1, 18 (2008), 1–30. https://doi.org/10.1002/hrm
[56] Sverre Tveit. 2013. Educational assessment in Norway. *Profiles of Education Assessment Systems Worldwide* (2013), 221–237.
[57] USA Bebras® Computing Challenge. 2019. Bebras Challenge. https://www.bebraschallenge.org/
[58] 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 on - WiPSCE '18.* ACM Press, New York, New York, USA, 1–10. https://doi.org/10.1145/3265757.3265762
[59] Sara Vogel, Christopher Hoadley, Laura Ascenzi-Moreno, and Kate Menken. 2019. The Role of Translanguaging in Computational Literacies: Documenting Middle School Bilinguals' Practices in Computer Science Integrated Units. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education.* 1164–1170.
[60] David Weintrop, Elham Beheshti, Michael S Horn, Kai Orton, Laura Trouille, Kemi Jona, and Uri Wilensky. 2014. Interactive assessment tools for computational thinking in High School STEM classrooms. In *International Conference on Intelligent Technologies for Interactive Entertainment.* Springer, 22–25.
[61] Aman Yadav, Ninger Zhou, Chris Mayfield, Susanne Hambrusch, and John T Korb. 2011. Introducing computational thinking in education courses. In *SIGCSE.* ACM Press, Dallas, Texas, 465. https://doi.org/10.1145/1953163.1953297