A Topical Review of Evaluation Instruments for Computing Education

Adrienne Decker
Department of Engineering Education
University at Buffalo
Buffalo, NY, USA
adrienne@buffalo.edu

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

ABSTRACT

As computing education research continues to grow and mature as a field, it becomes more important to focus on the quality and rigor of our research studies. One important aspect of any research study is its formal evaluation. Using standardized and validated instruments relevant to computer science education to perform evaluations can increase the quality of the study and the value of its results. However, researchers often create their own instruments rather than using existing ones, perhaps due to their lack of knowledge of the value of using an existing instrument or due to the challenge of finding such instruments. Through a review of relevant computing education literature, this paper presents a listing of 47 evaluation instruments specifically designed for measuring programs or constructs related to computing that can influence student achievement and learning. Analysis of purpose, target audience, reliability, and validity of the instruments is also presented. The paper ends with a call for the community to begin to make more regular use of validated instruments in their studies when possible and to develop and validate additional instruments in areas where few exist.

CCS CONCEPTS

• **Social and professional topics~Computing education programs** • **Social and professional topics~Computer science education** • **Social and professional topics~Student assessment**

KEYWORDS

Evaluation instruments, validation, assessment

ACM Reference format:
Adrienne Decker and Monica M. McGill. 2019. A Topical Review of Evaluation Instruments for Computing Education In *Proceedings of 50th ACM Technical Symposium on Computer Science Education (SIGCSE '19), February 27-March 2, 2019, Minneapolis, MN, USA.ACM, New York, NY, USA, 7 pages.* https://doi.org/10.1145/3287324.3287393

SIGCSE '19, February 27-March 2, 2019, Minneapolis, MN, USA

ACM ISBN 978-1-4503-5890-3/19/02.
https://doi.org/10.1145/3287324.3287393

1 INTRODUCTION

Since the publication of Computer Science Education Research in 2004 [9], there has been a handbook for conducting research in computing education and an increased scrutiny of the methods used in such studies. Many have argued that as a community we need to a better job of both designing and reporting on the research done in this space [1, 13, 19, 21].

One key aspect of doing a better job at our research is in measurement and evaluation. Recent studies that looked at data across multiple other studies have noted that there is a lack of both standardization in variables that are measured and how they are measured [5, 6, 10, 16].

It is critically important to adopt validated measures so that the variables that are measured are being done reliably. Using standardized instruments will also help us as a community to be able to better compare results. However, in order to achieve this goal, researchers need to be aware of what instruments are available to them and choose to use them when appropriate for their research. One problem that faces researchers is the lack of ability to easily find such instruments, which often results in individual studies developing instruments for their own use that are not able to be validated.

Therefore, this study was undertaken to provide a taxonomy of existing research-based evaluation instruments that are available in publications and publicly-accessible websites and databases. For the collection and analysis, our overarching research question is: *What evaluation and assessment instruments are available in accessible computing education venues and online databases?*

It is our hope that by providing such a list, researchers will use these instruments in their studies. In addition, researchers and evaluators can potentially re-validate these instruments as well as creating new instruments in areas where there are few. The long-term goal of this research is to promote the creation and usage of standardized measures within the computer science education community for commonly measured constructs across a wide-range of learners.

2 METHODOLOGY

To determine the set of evaluation instruments that have been designed for computing education and recently used in research

studies, we began a systematic search of the computing education literature and other related databases. To do so, we:

- Examined 297 articles in the https://csedresearch.org database (2012-2016), noting any instruments used within studies as well as articles related to instrument validation, [15]
- Examined the table of contents for the proceedings of the ACM ICER conference, the ACM journal Transactions on Computing Education, and Taylor and Francis' Computer Science Education in the years 2012-2016 to determine if additional instrument validation studies were published that were not covered by the database above (as it is specialized to research in pre-college computing activities),
- Searched several online databases for instruments, including American Evaluation Association [2], STELAR [7] The Pear Institute [18], Institute for the Integration of Technology into Teaching and Learning [11], MSPNet [4], Engineering is Elementary [3], and
- Used search engines to perform searches based on relevant keywords (e.g., computer science, inventory, survey, instrument, evaluation, interest, self-efficacy, etc.).

After creating the initial list, we solicited the CS education community for additional instruments using the SIGCSE-Members listserv and computing education social media sites. At the end, we had a total of 47 evaluation instruments.

Demographics for each instrument were initially coded by a student assistant, with instructions and training from the research team, followed by verification by an independent coder (one of the researchers on the project).

The following data was collected for each instrument:

- Title of instrument
- Authors
- Year of publication
- Brief description of instrument
- Cost to use instrument
- Number of questions
- Type of questions
- Time required to complete instrument (or time limit)
- Target demographic
- Constructs assessed
- Reliability evidence presented
- Validity evidence presented
- URL to relevant article explaining instrument
- URL of instrument

3 RESULTS

Table 3 presents a list of these instruments and gives an indication of what type of construct it measures (cognitive, noncognitive, or program assessment). References for each instrument are given in Appendix A.

3.1 Cost and Availability

The cost to use the instrument and its availability seem to be intrinsically linked in our field. Of the 47 instruments studied, 39 (83%) were available, usually in the article that first used the instrument in a research study or a separate article that focused on the validation of the instrument. Many of those articles are not open access, but would be available through university libraries. Once a researcher had access to the article, the instrument does not have a cost to use it.

Of those instruments that were not available from the aforementioned articles, four were obtained from the authors who agreed to allow them to be housed in https://csedresearch.org database for free usage by anyone interested. Due to its nature, SCS1 [A-31] is available to anyone who requests access and is free to use, but is not publicly available to protect those using it for research purposes. That is, if the questions on the instrument become available to anyone, then it its efficacy to show knowledge of programming concepts is diminished.

In the end, 44 (94%) of the instruments are available to other researchers and do not require the researcher to pay a fee to administer. We attempted to reach out to the authors of the remaining instruments, and if the instrument is provided at some point in the future, it can be accessed in the https://csedresearch.org database.

3.2 Constructs Assessed

We divided the constructs assessed by the instruments into two main categories, cognitive and noncognitive constructs. For the purposes of this analysis, we considered cognitive constructs to be content or domain knowledge of computing or computer science, and noncognitive constructs are everything else.

There were 13 instruments (28%) that measured cognitive constructs. Of these, the types of knowledge measured were: computational thinking (6), CS1 concepts (3), CS2 concepts (1), digital logic (1), algorithmic analysis (1), and research skills (1).

There were 31 instruments (66%) that measured noncognitive constructs which included constructs such as self-efficacy, anxiety, confidence, enjoyment, sense of belonging, intent to persist, and perceptions.

There were four instruments that were categorized as measuring both cognitive and noncognitive constructs.

During analysis, it became clear that there was actually a third category of construct that was coded separately from cognitive and noncognitive. That third category is program evaluation. Program evaluation instruments seek to measure the effectiveness of a specific program and have items that address the issues pertinent to determining problems with delivery or execution of the program/intervention as well as ways to improve. Eight of the 47 instruments (17%) were categorized as having some or all of their items concerned with program evaluation.

3.3 Number of Items

The number of items on an instrument can generally give an idea of how long it would take to administer such an instrument. The amount of time needed to complete the survey or interview

for the instruments ranges from 5 minutes to 1 hour. Table 1 shows the breakdown of number of items on the instruments.

The largest segment of the instruments (40%) had between 11 and 30 items. It is important to note that the instruments with over 50 items generally have subscales of smaller numbers of items that can be administered independently of each other. One of the instruments [A-42] is an observational protocol and item count does not make sense in that context. Project Quantum [A-8] is a contribution-driven question bank for creating evaluations and currently has over 8000 contributed questions, but is specifically designed so that a user only gives questions with topics of interest to their assessment needs.

Table 1: Number of items in studied instruments

Number of items	Number of instruments (%)
1 – 10	11 (23%)
11 – 30	19 (40%)
31 – 50	6 (13%)
> 50	9 (19%)

3.4 Item Type

In analyzing the collected demographic data, we noticed three main categories of item types, Likert-type prompts, multiple choice, and open-ended questions. Several of the instruments (29%) combined these three types of items and few had item types that did not fit into these categories. Table 2 shows the breakdown of types of items for the instruments studied.

The largest segment of the instruments (38%) had only Likert-type prompts. Including those with the instruments that combine Likert-type prompts with other types of items that percentage raises to 64%. Of the seven remaining instruments not covered by the table two were interview protocols; one was an observation protocol; two asked for answers on a 100-point scale; one had Likert-type prompts and semantic differential items; one had questions that asked the participant to choose the level of use of technology that best describes their level; and one was an autograding/machine grading assessment.

Table 2: Types of items in studied instruments

Type	No. of instruments (%)
Likert-type items only	18 (38%)
Multiple choice items only	6 (13%)
Open-ended items only	2 (4%)
Likert and Multiple Choice	6 (13%)
Likert and Open Ended	2 (4%)
Multiple Choice and Open Ended	2 (4%)
All three types	4 (9%)

3.5 Target Demographic

Figure 1 shows the breakdown of the target demographics for the instruments. The grade levels indicated are based on the United States system. For purposes of the reporting, an instrument can be included for more than one demographic category (e.g., it targets all K-12 students). The target of undergraduates has the largest percentage of instruments (26%), but if you consider instruments targeting K-12 (and remove duplicated counting), there are actually 20 instruments (43%) that target that demographic. Pre-service and in-service teachers are targets of 15% of the instruments.

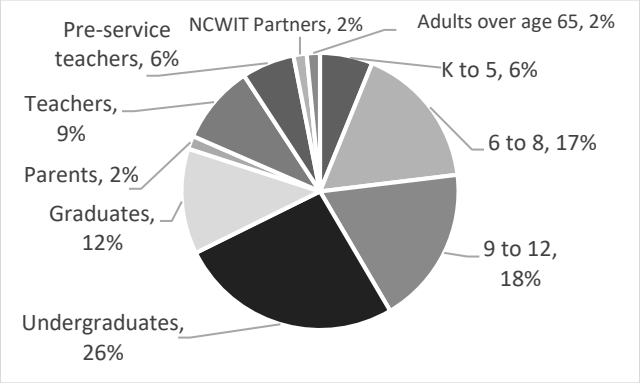

Figure 1: Target Demographics of Instruments Studied

3.6 Reliability

Reliability of an instrument is a statement about its consistency in measurement. Ideally, an instrument will provide the same results or score for the same answers each time it is given. Reliability information was found for 28 (60%) of the instruments. The most common form of reliability information was Cronbach's alpha or some other measure of internal consistency. There were two instruments that showed evidence of test-retest reliability within the data set.

3.7 Validity

Validity of an instrument helps us determine if we are measuring what we think we are measuring. Reliability is a necessary, but not sufficient condition to establish an instrument's validity. Reliability and/or validity have been checked for the specified particular demographic in a particular setting. Using a validated, reliable instrument does not necessarily mean that the instrument will be reliable and/or valid in different settings. It can provide, however, a greater measure of confidence than an instrument that has not been validated or determined to be reliable.

Validity can come in many forms and we see primarily two forms reported in the data. Of the 47 instruments studied, 24 (51%) provide some evidence for their validity. It was common in the data to see expert opinion give evidence of face validity and several instruments reported on construct validity evidence through various measures.

4 DISCUSSION

The instruments found are able to be used for free, with the exception of a few that require access to the publication in which they are included. This is a limitation for some, but often not those in the research community that have access to the most common journals and conference proceedings through college and university libraries. One positive step that could be taken for the instruments is to make the instrument freely accessible by

making it available outside the article which describes it. That way, those interested in using it do not need access to the article to get the items from the instrument.

A majority of the instruments had under 30 items, which means that they should be relatively quick to administer. Couple that with the fact that the most common type of item was Likert scale and we have a collection of instruments that are not onerous on a participant's time. As pointed out, even those with a large number of items have subscales that allow the researcher to give a smaller number of items that best suit their needs. Multiple choice is also another popular option for these instruments. Aside from time to administer, another advantage of these item types is their ability to be scored quickly and consistently.

The instruments tend to strongly favor measuring noncognitive constructs (ie., those that are not domain or content knowledge). However, it was observed, although not studied rigorously for this analysis, that several of the constructs appeared across many of the instruments. Further analysis is needed to uncover duplication of constructs being measured or gaps where the instruments fail to cover factors that have been shown to impact student achievement and learning [8, 12, 14]. The beginnings of this gap analysis is available in [17].

It is very promising to see such a large number of these instruments providing either reliability (60%) or validity evidence (51%). There are actually 20 instruments (43%) that provide both. Assessment and evaluation specialists will tell you that there are many types of both reliability and validity and that more is always better [20]. While we did not specifically analyze what type of validity evidence was presented for this study, content or face validity was very commonly cited as the only validity evidence. While that is a start, for sure, more is needed. Further, many of the instruments rely on self-report, which can be a threat to construct validity. Self-report threat can occur because the participant wants to make themselves "look better" to the researcher/research project [20].

4.1 Limitations

There are a number of limitations of this study. First, the search space for instruments, while quite large in some aspects, is still limited. It is possible that there instruments in the literature outside our date range for the searched databases that exist, particularly newer instruments that were not revealed through our searches or calls to the community for input. Future work to expand the collection of instruments and to add to this taxonomy is already planned and will include additional years, venues, and targeted searches.

Further, we are often limited in the evidence presented about a particular instrument to one paper or website entry. It is quite possible that other papers exist that were not found during this search that would provide additional information about the instrument. One of our next steps is to contact the authors to verify the data we curated for each instrument and to provide us with additional evidence or links that describe the evidence about the instruments. This will be time-consuming work, but will provide greater assurance that the publicly available collection of instruments and their summaries is accurate and useful for the community.

4.2 Future Work

The next steps for this taxonomy are to make it more accessible to those in the community. To that end, each one of the instruments identified here and the data extracted from each is presented in searchable form at https://csedresearch.org.

As identified previously, we believe more instruments are available than what we were able to identify for this taxonomy at this time. New instruments will be added to the website as soon as we are made aware of them. The website will provide a way for instrument authors to contribute information about an instrument not listed and to provide additions and corrections to entries already there.

Additionally, we continue work with a group of evaluators who are also collecting a large set of evaluation instruments for both STEM and computing education. We will continue to share our collection of instruments with them in order to promote their usage far and wide.

5 CONCLUSION

We take this work as a starting point for the community. Not only do measures of many constructs and programs already exist, many of them have already been shown to be reliable and/or validated. We encourage those conducting research in computing education and wanting to measure specific constructs to first look to see if there is an instrument already available that measures those constructs.

There is also a need to validate existing instruments that have not yet been validated as well as validate instruments in different contexts. Doing so will help improve the results of the initial research as well as providing a validated measure of the construct for others.

We recognize that the instruments available may not measure exactly what is needed for a particular research study. However, adaptation of an existing instrument is favorable to creating a brand new instrument, especially if constructs are similar. When faced with the challenge of an instrument not being available, we as a community should also look outside our own literature to see if other educational/learning sciences literature provides a starting point. If not, that truly represents a need in the community to create and/or adapt a measure for use in future research.

ACKNOWLEDGMENTS

This material is based upon work supported by the U.S. National Science Foundation under Grant Nos. 1625335 and 1757402.

Table 3: List of Evaluation Instruments

Program Evaluation	Cog-nitive	Noncog-nitive	Evaluation Instrument Title	Appendix Citation
	X		Algorithm Analysis Concept Inventory	[A-11]
		X	BASICS Study Student Implementation Questionnaire	[A-32]
		X	BASICS Study Teacher Implementation Questionnaire	[A-33]
X			BASICS Study Teacher Interview Guide	[A-34]
X	X	X	CISE REU A La Carte Student Survey	[A-37]
		X	Cognitive Load Component Survey	[A-21]
	X		Commutative Assessment	[A-40]
	X		Computational Thinking Pattern Analysis (CTPA)	[A-17]
	X	X	Computational Thinking Survey	[A-44]
	X		Computational Thinking Test	[A-36]
		X	Computer Anxiety Scale	[A-43]
		X	Computer Attitude Questionnaire (CAQ 5.14)	[A-15]
		X	Computer Attitude Questionnaire (CAQ 5.22)	[A-15]
		X	Computer Attitude Scale	[A-19]
		X	Computer Programming Self-Efficacy Scale	[A-31]
		X	Computer Science Attitude and Identity Survey (CSAIS)	[A-38]
		X	Computer Science Attitude Survey (Hoegh and Moskal)	[A-14]
		X	Computer Science Attitude Survey (Weibe et al)	[A-42]
		X	Computer Science Interest Survey	[A-2]
		X	Computing Attitudes Survey	[A-10]
	X		Digital Logic Concept Inventory (DLCI) (Form A)	[A-13]
		X	Draw-a-Computer-Scientist Test	[A-12]
		X	Effectiveness of Technology Outreach Survey	[A-20]
X			Evaluation of Faculty Participation in POSSE	[A-7]
	X	X	Evaluation of Student Participation in HFOSS	[A-7]
	X		Fairy Assessment	[A-41]
		X	Microcomputer Beliefs Inventory for Middle School Students	[A-35]
X			NCWIT Computing Major Pace and Workload Experience	[A-22]
		X	NCWIT Incoming Student Survey	[A-23]
X			NCWIT Overall Computing Major Satisfaction: Student Survey	[A-24]
X			NCWIT Pair Programming Student Final Assessment	[A-25]
		X	NCWIT Parent Survey	[A-26]
X			NCWIT Program Partner Survey	[A-27]
		X	New Computer Game Attitude Scale	[A-18]
	X		Project Quantum	[A-8]
	X		Project TREES Survey	[A-3]
		X	Robotics Activities Attitudes Scale	[A-9]
	X		SCS1	[A-30]
		X	Self-Efficacy for Computational Thinking (SECT)	[A-1]
		X	Stages of Adoption of Technology (SA)	[A-4]
	X	X	Student Centered Observation Protocol for computer-science Education (SCOPE)	[A-39]
		X	Student Computing Interest: Post Survey for Outreach programs	[A-28]
X			Student Outreach Experience: Interview Protocol for Students	[A-29]
		X	Teacher Attitudes toward Computers (TAC)	[A-5]
		X	Teachers' Attitudes Toward Information Technology (TAT)	[A-16]
		X	Teachers' Self-Efficacy in Computational Thinking (TSECT)	[A-1]
		X	Technology in Education Competency Survey (TECS)	[A-6]

APPENDIX A. INSTRUMENT REFERENCES

Please see https://csedresearch.org/evaluation-instruments/ for more information about instruments.

[A-1] Nathan Bean, Joshua Weese, Russell Feldhausen, and R Scott Bell. 2015. Starting from scratch: Developing a pre-service teacher training program in computational thinking. In Frontiers in Education Conference (FIE), 2015 IEEE. IEEE, 1–8.

[A-2] Janet Seeley Blouin. 2011. High school seniors' computer self-efficacy and interest in computer science careers. Ph.D. Dissertation. University of Georgia.

[A-3] Guanhua Chen, Ji Shen, Lauren Barth-Cohen, Shiyan Jiang, Xiaoting Huang, and Moataz Eltoukhy. 2017. Assessing elementary studentsâĂŹ computational thinking in everyday reasoning and robotics programming. Computers & Education 109 (2017), 162–175.

[A-4] Rhonda Christensen. 1997. Effect of technology integration education on the attitudes of teachers and their students. Unpublished doctoral dissertation, University of North Texas. Available: https://digital.library.unt.edu/ark:/67531/metadc277676/?q=christensen.

[A-5] Rhonda Christensen and Gerald Knezek. 1996. Constructing the Teachers' Attitudes Toward Computers (TAC) Questionnaire. (1996).

[A-6] Rhonda Christensen and Gerald Knezek. 2001. The Technology in Education Competency Survey (TECS): A self-appraisal instrument for NCATE standards. In Society for Information Technology & Teacher Education International Conference. Association for the Advancement of Computing in Education (AACE), 2288–2289.

[A-7] HFOSS Community. Foss2Serve. Evaluation Instruments. Retrived from: http://foss2serve.org/index.php/Evaluation_Instruments

[A-8] Computing at School. Project Quantum: tests worth teaching to. Retrieved from: https: //community.computingatschool.org.uk/resources/4382/single

[A-9] Jennifer Cross, Emily Hamner, Lauren Zito, Illah Nourbakhshh, and Debra Bernstein. 2016. Development of an assessment for measuring middle school student attitudes towards robotics activities. In Frontiers in Education Conference (FIE), 2016 IEEE. IEEE, 1–8.

[A-10] Brian Dorn and Allison Elliott Tew. 2015. Empirical validation and application of the computing attitudes surve. Computer Science Education, 25, 1, 1-36,

[A-11] Mohammed F Farghally, Kyu Han Koh, Jeremy V Ernst, and Clifford A Shaffer. 2017. Towards a concept inventory for algorithm analysis topics. In Proceedings of the 2017 ACM SIGCSE. ACM, 207–212.

[A-12] Alexandria K Hansen, Hilary A Dwyer, Ashley Iveland, Mia Talesfore, Lacy Wright, Danielle B Harlow, and Diana Franklin. 2017. Assessing children's understanding of the work of computer scientists: the draw-a-computer-scientist test. In Proceedings of the 2017 ACM SIGCSE. ACM, 279–284.

[A-13] Geoffrey L Herman, Michael C Loui, and Craig Zilles. 2010. Creating the digital logic concept inventory. In Proceedings of the 41st ACM technical symposium on Computer science education. ACM, 102–106.

[A-14] Andrew Hoegh and Barbara M Moskal. 2009. Examining science and engineering students' attitudes toward computer science. In Frontiers in Education Conference, 2009. FIE'09. 39th IEEE. IEEE, 1–6.

[A-15] Gerald Knezek and Rhonda Christensen. 1996. Validating the Computer Attitude Questionnaire (CAQ). (1996).

[A-16] Gerald Knezek and Rhonda Christensen. 1998. Internal consistency reliability for the teachers' attitudes toward information technology (TAT) questionnaire. In Proceedings of the society for information technology in teacher education annual Conference. 831–836.

[A-17] Kyu Han Koh, Hilarie Nickerson, Ashok Basawapatna, and Alexander Repenning. 2014. Early Validation of Computational Thinking Pattern Analysis. In Proceedings of ITiCSE '14. ACM, New York, NY, USA, 213–218. https://doi.org/10.1145/2591708.2591724

[A-18] Eric Zhi-Feng Liu, Chun-Yi Lee, and Jen-Huang Chen. 2013. Developing a New Computer Game Attitude Scale for Taiwanese Early Adolescents. Journal of Educational Technology & Society 16, 1 (2013).

[A-19] Brenda H Loyd and Clarice Gressard. 1984. Reliability and factorial validity of computer attitude scales. Educational and Psychological measurement 44, 2 (1984), 501–505.

[A-20] Monica M McGill, Adrienne Decker, and Amber Settle. 2016. Undergraduate students' perceptions of the impact of pre-college computing activities on choices of major. ACM Transactions on Computing Education (TOCE) 16, 4 (2016), 15.

[A-21] Briana B Morrison, Brian Dorn, and Mark Guzdial. 2014. Measuring cognitive load in introductory CS: adaptation of an instrument. In Proceedings of the tenth annual conference on International computing education research. ACM, 131–138.

[A-22] National Center for Women & Information Technology. NCWIT Computing Major Pace and Workload: Student Survey. Retrieved from https://www.ncwit.org/file/computing-major-pace-and-workload-student-survey

[A-23] National Center for Women & Information Technology. NCWIT Incoming Student Survey. Retrieved from https://www.ncwit.org/file/incoming-student-survey

[A-24] National Center for Women & Information Technology. NCWIT Overall Computing Major Satisfaction: Student Survey. Retrieved from https://www.ncwit.org/file/incoming-student-survey

[A-25] National Center for Women & Information Technology. NCWIT Pair Programming Student Final Assessment. Retrieved from https://www.ncwit.org/file/pair-programming-student-final-assessment

[A-26] National Center for Women & Information Technology. NCWIT Parent Survey. Retrieved from https://www.ncwit.org/file/parent-survey

[A-27] National Center for Women & Information Technology. NCWIT Program Partner Survey. Retrieved from https://www.ncwit.org/file/program-partner-survey

[A-28] National Center for Women & Information Technology. NCWIT Student Computing Interest: Post Survey for Outreach Programs. Retrieved from https://www.ncwit.org/file/student-computing-interest-post-survey-outreach-programs

[A-29] National Center for Women & Information Technology. NCWIT Student Outreach Experience Interview Protocol for Students. Retrieved from https://www.ncwit.org/file/student-outreach-experience-interview-protocol-students

[A-30] Miranda C Parker, Mark Guzdial, and Shelly Engleman. 2016. Replication, validation, and use of a language independent CS1 knowledge assessment. In Proceedings of the 2016 ACM conference on international computing education research. ACM, 93–101.

[A-31] Vennila Ramalingam and Susan Wiedenbeck. 1998. Development and validation of scores on a computer programming self-efficacy scale and group analyses of novice programmer self-efficacy. Journal of Educational Computing Research 19, 4 (1998), 367–381.

[A-32] Outlier Research & Evaluation. 2017. BASICS Study ECS Student Implementation and Contextual Factor Questionnaire Measures [Measurement scales]. Technical Report. Outlier Research & Evaluation at UChicago STEM Education | University of Chicago; Chicago, IL. https://s3.amazonaws.com/cemse/basics/files/findings/BASICS_SQ_Measures_FINAL.pdf

[A-33] Outlier Research & Evaluation. 2017. BASICS Study ECS Teacher Implementation and Contextual Factor Questionnaire Measures [Measurement scales]. Technical Report. Outlier Research & Evaluation at UChicago STEM Education | University of Chicago; Chicago, IL. https://s3.amazonaws.com/cemse/basics/files/findings/BASICS_TQ_Measures_FINAL.pdf

[A-34] Outlier Research & Evaluation. 2017. BASICS Study ECS Teacher Interview Guide [Measurement scales]. Technical Report. Outlier Research & Evaluation at UChicago STEM Education | University of Chicago; Chicago, IL. https://s3.amazonaws.com/cemse/basics/files/findings/BASICS_TI_Guide_FINAL.pdf

[A-35] Iris M Riggs and Larry G Enochs. 1993. A microcomputer beliefs inventory for middle school students: Scale development and validation. Journal of Research on Computing in Education 25, 3 (1993), 383–390.

[A-36] Marcos Román-González, Juan-Carlos Pérez-González, Jesús Moreno-León, and Gregorio Robles. 2018. Can computational talent be detected? Predictive validity of the Computational Thinking Test. International Journal of Child-Computer Interaction (2018).

[A-37] Audrey S Rorrer. 2016. An evaluation capacity building toolkit for principal investigators of undergraduate research experiences: A demonstration of transforming theory into practice. Evaluation and program planning 55 (2016), 103–111.

[A-38] Alicia N Washington, Shaefny Grays, and Sudhipta Dasmohapatra. 2016. The Computer Science Attitude and Identity Survey (CSAIS): A Novel Tool for Measuring the Impact of Ethnic Identity in Underrepresented Computer Science Students. In Proceedings of the ASEE's 123rd Annual Conference & Exposition 2016.

[A-39] DC Webb, SB Miller, H Nickerson, R Grover, and K Gutiérrez. 2014. Student Centered Observation Protocol for computer-science Education (SCOPE). University of Colorado at Boulder (2014).

[A-40] David Weintrop. 2015. Comparing Text-based, Blocks-based, and Hybrid Blocks/Text Programming Tools. In Proceedings of the Eleventh Annual International Conference on International Computing Education Research (ICER '15). ACM, New York, NY, USA, 283–284. https://doi.org/10.1145/2787622.2787752

[A-41] Linda Werner, Jill Denner, Shannon Campe, and Damon Chizuru Kawamoto. 2012. The fairy performance assessment: measuring computational thinking in middle school. In Proceedings of the 43rd ACM technical symposium on Computer Science Education. ACM, 215–220.

[A-42] Eric Wiebe, Laurie Williams, Kai Yang, and Carol Miller. 2003. Computer science attitude survey. computer science 14, 25 (2003), 0–86.

[A-43] Katherine V Wild, Nora C Mattek, Shoshana A Maxwell, Hiroko H Dodge, Holly B Jimison, and Jeffrey A Kaye. 2012. Computer-related self-efficacy and anxiety in older adults with and without mild cognitive impairment. Alzheimer's & Dementia 8, 6 (2012), 544–552.

[A-44] Aman Yadav, Ninger Zhou, Chris Mayfield, Susanne Hambrusch, and John T Korb. 2011. Introducing computational thinking in education courses. In Proceedings of the 42nd ACM technical symposium on Computer science education. ACM, 465–470.

REFERENCES

[1] Ahmed Al-Zubidy, Jeffrey C. Carver, Sarah Heckman, and Mark Sherriff. 2016. A (Updated) Review of Empiricism at the SIGCSE Technical Symposium. In Proceedings of the 47th ACM Technical Symposium on Computing Science Education (SIGCSE '16). ACM, New York, NY, USA, 120-125. DOI: https://doi.org/10.1145/2839509.2844601

[2] American Evaluation Association. N.d. American Evaluation Association. Retrieved from: http://www.eval.org/

[3] Boston Museum of Science. Research Instruments. Retrieved from: https://www.eie.org/engineering-elementary/research/research-instruments

[4] Center for School Reform at TERC. MSPNet Resource List. Retrieved from: http://hub.mspnet.org/index.cfm/msp_tools

[5] Adrienne Decker, Monica M. McGill, and Amber Settle. 2016. Towards a Common Framework for Evaluating Computing Outreach Activities. In Proceedings of the 47th ACM Technical Symposium on Computing Science Education (SIGCSE '16). ACM, New York, NY, USA, 627-632. DOI: https://doi.org/10.1145/2839509.2844567

[6] Adrienne Decker and Monica M. McGill. 2017. Pre-College Computing Outreach Research: Towards Improving the Practice. In Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education (SIGCSE '17). ACM, New York, NY, USA, 153-158. DOI: https://doi.org/10.1145/3017680.3017744

[7] Education Development Center, Inc. STELAR – STEM Learning and Research Center. Retrieved from: http://stelar.edc.org/resources

[8] 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

[9] S. Fincher and M. Petre, eds., Computer Science Education Research, Taylor & Francis, The Netherlands, Lisse, 2004.

[10] Petri Ihantola, Arto Vihavainen, Alireza Ahadi, Matthew Butler, Jürgen Börstler, Stephen H. Edwards, Essi Isohanni, Ari Korhonen, Andrew Petersen, Kelly Rivers, Miguel Ángel 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. In Proceedings of the 2015 ITiCSE on Working Group Reports (ITICSE-WGR '15). ACM, New York, NY, USA, 41-63. DOI: https://doi.org/10.1145/2858796.2858798

[11] Institute for the Integration of Technology into Teaching and Learning. UNT Institute for the Integration of Technology. Retrieved from: https://iittl.unt.edu/content/instruments

[12] 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.

[13] Alex Lishinski, Jon Good, Phil Sands, and Aman Yadav. 2016. Methodological Rigor and Theoretical Foundations of CS Education Research. In Proceedings of the 2016 ACM Conference on International Computing Education Research (ICER '16). ACM, New York, NY, USA, 161-169. DOI: https://doi.org/10.1145/2960310.2960328

[14] Robert J Marzano. 2003. What works in schools: Translating research into action. ASCD.

[15] Monica McGill and Adrienne Decker. (2017). Computer Science Education Repository. Available online: https://csedresearch.org.

[16] Monica M. McGill, Adrienne Decker, and Zachary Abbott. 2018. Improving Research and Experience Reports of Pre-College Computing Activities: A Gap Analysis. In Proceedings of the 49th ACM Technical Symposium on Computer Science Education (SIGCSE '18). ACM, New York, NY, USA, 964-969. DOI: https://doi.org/10.1145/3159450.3159481

[17] Monica M. McGill, Adrienne Decker, Tom McKlin, and Kathy Haynie. 2019. A Gap Analysis of Noncognitive Constructs in Evaluation Instruments Designed for Computing Education. In Proceedings of the 50th ACM Technical Symposium on Computer Science Education (SIGCSE '19), ACM, New York, NY, USA. 7 pages. DOI: https://doi.org/10.1145/3287324.3287362

[18] Pear Institute. PEAR (Program in Education, Afterschool and Resiliency). Retrieved from: http://www.pearweb.org/atis/tools/jump

[19] J. Randolph, G. Julnes, E. Sutinen, S. Lehman, "A Methodological Review of Computer Science Education Resarch," Journal of Information Technology Education, vol. 7, 2008, pp. 135-162

[20] Trochim, W. M. 2006. The Research Methods Knowledge Base, 2nd Edition. Retrieved from http://www.socialresearchmethods.net/kb/.

[21] D. W. Valentine, "CS Education Research: A Meta-Analysis of SIGCSE Technical Symposium Proceedings," Proc. of the 35th SIGCSE Technical Symposium on Computer Science Education, Norfolk, VA, USA, March 3-7, 2004, pp. 255-259.