INFACT: Include Neurodiversity in Foundational and Applied Computational Thinking

1. Significance

1.1 National Priority

"*The long-term strength of our workforce requires that the full range of STEM and non-STEM career pathways be available to all Americans. This imperative is undergirded by two foundational principles: first, that every individual in the United States is afforded the opportunity to reap the benefits of advancements in science and technology; second, that our ability to respond to national needs and remain globally competitive will require the capabilities and ingenuity of individuals of diverse backgrounds.*" National Science Board (2015, p. 22)

1.2 Absolute Priorities and Competitive Preference Priority

INFACT meets **Absolute Priority 1** by designing a comprehensive STEM program that addresses the great cognitive variability among learners. INFACT will incorporate innovative strategies to address issues related to many commonly labeled disabilities such as specific learning disabilities (SLD), dyslexia, dyscalcula, ADD, and autism (hereafter called *neurodiverse learners*). INFACT will embed flexible supports that support a wide variety of learners' executive function and sensory needs that are often associated with neurodiversity.

INFACT is a field-initiated innovation that meets **Absolute Priority 3** and the **Competitive Preference Priority** to strengthen CS and STEM education. INFACT infuses fundamental practices of CT within STEM teaching and learning to leverage the strengths and to support the challenges of learner variability in today's classrooms.

1.3 Rationale

The future workforce needs computational thinkers. Growth in STEM employment is outpacing the growth of employment at large (NSF, 2015), and computer occupations are projected to yield half a million new jobs between 2014 and 2024 (NSF, 2018). Further, the nature of STEM jobs is

becoming increasingly computational (Kaczmarczyk, Dopplick, & EP Committee, 2014), meaning that part of preparing learners for STEM fields includes computational thinking and computer science instruction. The Computer Science workforce, however, represents a conspicuously narrow demographic. Neurodiverse students are underrepresented in STEM university programs and the STEM workforce despite an apparent inclination towards STEM for many neurodiverse learners, such as those with autism (NSF, 2015; Wei, Yu, Shattuck, & Blackorby, 2017).

Learner variability may present an opportunity to build a strong computational workforce. Considering neurodiversity as learner variability, as opposed to learning disability, leads researchers to interesting questions about how we unleash the potential in learners for tasks needed in our future workforce (Rose, Rouhani, & Fischer, 2013). Many neurodiverse learners have particular areas of strength in tasks related to CT. For example, some learners with autism exhibit hyper-attention and hyper-systemizing behaviors (Baron-Cohen, Ashwin, Ashwin, Tavassoli, & Chakrabarti, 2009). Some individuals with dyslexia show strengths in the ability to integrate information and build global perspectives (Schneps, Brockmole, Sonnert, & Pomplun, 2012; von Károlyi, 2001), which may lead to better collaborative problem-solving.

These diverse cognitive functions may translate to valuable skills such as CT and coding, and thus the recognition and nurturing of these talents may be crucial for the development of our future workforce (Martinuzzi & Krumay, 2013). For this reason, many large companies, including Microsoft, SAP, and Google, now have established programs or are developing programs specifically designed to recruit neurodiverse individuals (Wang, 2014). To prepare for this future workforce, educators are building materials that focus on CT within STEM (e.g., Weintrop et al., 2016; Grover, 2017; Wing, 2011; Barr & Stephenson, 2011) to describe the ways

of thinking and practices used by those who design computational systems. The K–12 education community is actively studying how this new and critical set of practices can be infused into current educational programs.

1.4 Building Upon Promising Strategies

There are several promising lines of research investigating how to address learner variability in STEM education (e.g., Israel, et al., 2018; Ke & Lee, 2016; Ray, Israel, Lee, & Do, 2018) and how to infuse CT within STEM teaching and learning (e.g., Weintrop et al., 2016). For INFACT, a consortium of these STEM education researchers, computer scientists, cognitive psychologists, and educational designers from universities, non-profits, and industry come together to build a comprehensive program that is greater than the sum of its parts.

In particular, INFACT will build upon promising approaches to address the growing learning variability in today's classrooms such as differentiation—providing unique experiences for each learner based upon their strengths, struggles, and choices (Tomlinson & Moon, 2013)—and meeting the individual cognitive, emotional, and social needs of each child and their unique preferences and abilities (Immordino-Yang, Darling-Hammond, & Krone, 2018). Because differentiated classroom instruction requires that teachers understand and monitor what each learner knows, what they are ready to learn, and what resources can meet these needs (Tomlinson & Moon, 2013), INFACT materials will not only include cognitive supports for different learners, but innovative forms of formative assessment to be able to measure what learners really know. Many neurodiverse learners have cognitive differences that impact their ability to listen, think, speak, read, write, spell, or do mathematical calculations (USDoE, 2017), rendering testing ineffective at measuring learning outcomes (Thiede et al., 2015; Brendefeur et al., 2015). INFACT will use innovative formative assessments that use information about what

learners *do* in an activity, rather than what they can say on a test, which show promising evidence for assessment of neurodiverse learners (Ke & Lee, 2016, Grover, 2019; Rowe, Asbell-Clarke, & Baker, 2015; Rowe et al., 2017; Rowe et al., 2018; Rowe et al., 2019).

1.5 Exceptional Approaches to Absolute and Competitive Priorities

INFACT builds upon promising evidence in learning materials that infuse CT into STEM by bringing together several strong lines of research into a comprehensive program. The exceptional approaches that INFACT deploys are listed in Table 1 and described in the next section.

Table 1: List of Exceptional Approaches to EIR Absolute and Competitive Priorities

Exceptional Approaches to EIR Absolute and Competitive Priorities
Curation and integration of a wealth of CT learning materials with promising evidence to infuse CT into STEM for grades 3–8.
Careful design of a coherent CT progression for grades 3–8 that lays a conceptual foundation for CT practices and prepare learners for application of CT practices in a variety of STEM projects.
Embedded supports that leverage the intersection between CT practices and executive function.
Innovative and accessible assessments that provide formative information about CT practices for an audience with high learner variability.
A curated and correlated set of accessible CT outcome measures for an audience with high learner variability.
A system of teacher tools and professional development that support flexible and effective implementation of INFACT in a broad array of grade 3–8 classes.

2. Description of Exceptional Approaches

2.1 Integrating of Existing Promising Learning, Assessment, and Teaching Materials

INFACT brings together a group of leading researchers as members of the consortium who will design the comprehensive program elements. Together, we will curate existing materials and design sequences of activities within and across grades 3–8 to build foundational and applied CT within STEM teaching and learning. INFACT will also include contributions from colleagues

who have materials with promising evidence to infuse CT into STEM learning and assessment

(e.g., CT assessments from James Lester from NCSU). These colleagues will serve as advisors.

Table 2: List of Consortium Members, Contributions, Research Sites, and Citations

Consortium Members	Contribution	Gr.	Research School(s)	Cited Research
TERC PI: Asbell-Clarke	CT-infused STEM teaching and learning materials; *Zoombinis* CT learning game; Adaptive formative assessments	3-8	Braintree Public Schools, National *Zoombinis* Teacher Network	Rowe et al., 2015; Rowe et al., 2017; Rowe et al., 2018
UMD PI: Weintrop	*SpheroMath* curriculum using robotics	3-4	Washington DC Public Schools	Orton et al., 2016; Pei et al, 2018; Weintrop et al., 2016
Looking Glass PI: Grover	VELA CT learning activities and assessments	5-8	SF & Palo Alto Unified School Districts	Grover et al., 2015; Grover, 2017; Grover et al., 2019
FSU PI: Ke	Mixed Reality adaptive assessments for representation flexibility	5-8	FSU Center for Autism and Related Disabilities	Ke & Lee, 2016; Ke & Moon, 2018
UFLA PI: Israel	Metacognitive supports in CT	3-7	PK Yonge Developmental Research School	Israel et al., 2018; Ray et al., 2018
Digital Promise, PI: Burke	Teacher PD models for CT; Learner Variability Project design tools	5-8	League of Innovative Schools	Burke, 2016; Kafai & Burke, 2014

The INFACT learning materials will include a variety of digital and non-digital activities

that embed CT practices within STEM contexts. For example, TERC's CT-learning game,

Zoombinis (Rowe et al., 2018), has been used along with supporting offline activities in a

nationwide study of over 50 classrooms. *Zoombinis* classroom materials support the development

of Problem Decomposition, Pattern Recognition, Abstraction, and Algorithm Design during

gameplay and help teachers connect those CT practices to mathematical problem-solving and

scientific investigations. Researchers at University of Maryland (UMD) use activities with

Sphero robots to support grade 4 math learning throughout the Washington, DC Public School

district. Researchers at Looking Glass Ventures (in collaboration with Stanford University and

SRI International) have designed and empirically investigated learning activities and assessments for key CT elements in middle schools in the SF Bay Area to examine learning of algorithmic thinking and its key elements such as variables, expressions, loops, abstraction that are valuable foundations for CT in any STEM setting. Ke at Florida State University (FSU) and Israel at University of Florida (UFLA) are both researching innovative methods to support learner variability in CT and STEM using mixed reality interfaces and games. Digital Promise has been studying CT implementation within schools and large school districts, as well as providing design models in their Learner Variability Project, and is bringing both lines of work to INFACT.

2.2 Designing a Coherent CT Learning Progression for grades 3–8

The consortium will build from these research-based learning materials to co-design a coherent CT progression for grades 3–8 that can be infused into STEM curriculum (Table 3). Although the progression may be iterative, with a back-and-forth between foundation building and application of CT in STEM, we will support a gradual transition from foundational practices to application in STEM.

Table 3: Focus Points of CT Progression

CT Progression Focus Points		
Grades 3–4	**Grades 5–6**	**Grades 7–8**
Conceptual Foundation	Tool Building	Application
e.g., Zoombinis, Puzzles, Sorting Activities, Spheros	e.g., Scratch, BlocksCAD, Code.Org, AppInventor	e.g., Modeling, Designing, Simulations, Data Visualizations

TERC has been exploring the infusion of CT into differentiated STEM teaching and learning in grades 3–8 during a Research-Practice Partnership with Braintree Public Schools (BPS) in Braintree, MA. Based on the promising evidence from TERC's RPP, the progression will begin in grade 3 with everyday activities that build foundational concepts of problem

decomposition, pattern recognition and abstraction, and algorithm design. Elementary teachers use games, puzzles, and sorting activities to build a conceptual foundation for these CT practices. Concurrently, technology teachers gradually introduce coding, modeling, and design tools that employ these foundational practices and prepare students to use them in specialized STEM classes in middle school. STEM teachers in middle school can then engage students in CT-infused projects without have to take time to teach the required tools. INFACT will build upon this evidence to design a flexible progression of CT-infused STEM teaching and learning for grades 3–8.

2.3 Leveraging CT to Address Learning Variability

A key innovation of INFACT is leveraging the natural intersection between CT and learner variability. Much of learning variability revolves around executive function, sensory needs, and cognitive load. Executive function (EF) tasks are involved in planning and organizing goal-oriented tasks, and invoke the neural processes of attention, working memory, and information processing (Denkla, 1994, Semenov & Zelazo, 2018). Learners with strong EF skills are able to pay attention and engage their thoughts and emotions in a goal-directed way that sets them up for classroom success, and avoids situations where learners are derailed by distractions and emotional dysregulation (Semenov & Selazo, 2018). To help support EF strategy development, playful activities may increase persistence, and learners may benefit from experiences with choice, symbol manipulation, and hands-on opportunities to construct and navigate among multiple representations of a challenge (Moran & Gardner, 2018).

Interestingly, many of the cognitive tasks required for high-level executive functioning overlap with CT practices, and thus may present as strengths and/or weaknesses for many neurodiverse learners (Table 4). This intersection between CT and executive function will guide

the selection of INFACT materials and design of supports and be a focus of the design research. Embedded supports within INFACT will be designed to support these CT practices and the related EF for a broad range of learner variability.

Table 4: Intersection of CT Practices and Executive Function

CT Practice	Description	Related Executive Function
Problem Decomposition	Breaking up large problems into smaller problems	Identifying manageable components of a goal-oriented
Pattern Recognition	Noticing repeating and salient trends	Attention to details and systematicity
Abstraction	Generalizing patterns to rules or categories	Metacognition and rule-shifting
Algorithm Design	Designing generalized solutions or procedures to solve problem	Metacognition, systematicity, and planning

2.4 Materials for Learning Variability

INFACT will draw from research in special education, cognitive psychology, and educational neuroscience to embed supports for learner variability in select INFACT materials. The Universal Design community, focusing on inclusion of physically and neurodiverse populations, emphasizes the need for multiple representations of a concept (e.g., Meyer, Rose & Gordon, 2014), which is consistent with research in early mathematics education research (e.g., Fuchs et al., 2011). However, the increased cognitive load associated with multiple elements presented simultaneously may be overwhelming, particularly for those students lacking prior knowledge (Lee, Plass & Homer, 2006). There is no one-size fits all solution, which is why differentiation is so powerful (Tomlinson & Moon, 2013).

We will design a flexible set of scaffolds for interactive INFACT materials that detect differences in how learners interpret words and symbols (e.g., dyslexia and dyscalcula), their sensory needs (e.g., sensitivity to stimuli), and their executive function needs (e.g., organization

and metacognition) to provide an adaptive, customized learning experience. INFACT materials will provide flexibility in how representations (e.g., text and numbers) are used by offering a variety of manipulatives and graphical representations and gradually transition from one symbolic notation to another. For example, TERC's RPP is co-designing an elementary math interactive that uses CT to support place value understandings, with unit construction and coordination for 1s, 10s, and 100s. Embedded graphical organizers support learners as they sort through objects with different units, and also offers multiple representations of information and gradually move learners from their preferred representation. Graphical overlays can also provide flexible and versatile charts and tables that may help emphasize the salient STEM concept where CT is being applied.

Supports for learner variability will also draw from research on the role of metacognition and reflection in EF (Israel et al., 2016; Semenov & Selazo, 2018). Learners are required to engage their working memory to remember new information at the same time that they reflect on what they already know. Scaffolds such as graphical organizers, milestones, visual timelines, and other supports will help learners keep track of their goal and progress towards that goal.

INFACT supports will also address variability in learners' sensory needs. When working with learners with autism, we will draw from the promising evidence of Benton et al. (2014) who offered quiet and familiar environments, visual organizers, and consistent session structure and routines to learners with autism. When they adapted a similar design model for learners with dyslexia, they decreased distractions and added multi-sensory activities, short and focused activities, and easier-to-read fonts. Because of the frequent co-morbidity of multiple conditions related to neurodiversity (Mayes & Calhoun, 2006; Willcutt, 2014), we will emphasize flexibility in the design of INFACT materials to support each learner's unique set of needs and preferences.

2.5 Assessments and Adaptive Interactives

INFACT is framed in the learning science framework that learning materials and assessment are inextricably linked (NRC, 2001) and that adequate assessment is critical for differentiated learning (Tomlinson & Moon, 2014). Alongside the INFACT CT Progression, we will build a progression of competencies that can be used in a variety of formative assessment models. These formative CT assessments will be embedded within a select set of INFACT learning activities and will guide the adaptation and customization of those activities. INFACT will leverage several innovative methods that show promising evidence to assess implicit learning—learning that can be may be masked by tests that rely on text, coding, or other symbolic formalisms that may prevent accurate measurement of STEM knowledge (Baroody, Bajwa & Eiland, 2009; Chong & Seigel, 2008; Rowe et al., 2015). We will use embedded assessments within digital learning experiences, such as games and VR simulations (e.g., Ke, Shute, Clark, & Erlebacher, 2019), and within coding and modeling activities (e.g., Grover et al., 2017), as well as process-based assessments for project-based learning (Asbell-Clarke & Bradbury, 2017).

Each of these assessment models has different rigorous research and validation models to support claims of evidence of learning outcomes. Building from game-based assessment using an Evidence-Centered Design (ECD) model (Shute & Kim, 2014; Mislevy & Riconscente, 2005), TERC's current research deploys educational data mining (EDM) methods such as automated detectors of strategies and CT practices in gameplay logs. Starting with extensive human analysis (Rowe et al., 2015), researchers have identified CT practices within gameplay behaviors in the logic puzzle game, *Zoombinis* (Rowe et al., 2017). The datalogs containing each "click" in the game, along with a timestamp and unique PlayerID, are collected and organized in our existing game-based, learning research architecture called *DataArcade*. *DataArcade* allows efficient

human labeling of CT during simulated "video replays" generated from the data logs that eventually leads to the building of automated detectors of CT practices during gameplay. In another example, FSU's current research in virtual reality and game-based training integrates individualized training of representational flexibility for learners on the autism spectrum into CT-infused STEM activity (Ke & Lee, 2016; Ke & Moon, 2018). Bayesian networks (BNs) and other machine learning methods (e.g., random forest and support vector machine) are used to accumulate these measures and make a diagnostic estimate on an individual's task-relevant representational competence level.

These and other novel methods will be used to design innovative formative assessments for INFACT that both inform the teacher with actionable suggestions of next steps, and inform the interactive itself on how to adapt to best suit the unique learner. TERC's *DataArcade* already has this real-time adaptive functionality available. By monitoring the a) sensory choices the learner makes (e.g., color and sound); b) the learner's mastery as evidenced by the automated detectors; and other features such as c) the learner's productive and/or unproductive persistence (Beck & Gong, 2013; Kai et al., 2017), an adaptive system can then customize the interface, the hints, and the pacing of the puzzles for individual learners.

2.6 Teacher Tools and Professional Development

One of the greatest challenges of CT education in elementary and middle school is that the subject area is new to most teachers and there are no existing models of learning, teaching, and assessment for teachers to have learned or used previously. CT also must fit within the existing, overstuffed curriculum and busy lives of teachers and staff. To tackle this challenge, we will build from promising evidence for teacher professional development in TERC's RPP with BPS

as well as the researchers of Digital Promise who work with teachers nationwide, including the League of Innovative Schools, on the infusion of CT throughout their existing curriculum.

As learned from our RPP, INFACT PD will respond to teacher and administrators' goals to supplement professional development (PD) efforts already underway in any school or district. INFACT will use a variety of modes of PD, created through an iterative onsite process with consortium members, and focusing on scalable implementation. We will design a scalable suite of resources that will include a series of short videos for teachers and online virtual workshops to extend existing PD efforts at a district level. Examples include: video resources with short explanations and examples of CT in everyday life; an easy-to-navigate teacher interface and system that follows the CT progression and links to materials for each grade and subject area; virtual working sessions on INFACT teacher tools to ask questions and get support; and materials and virtual support for on-site teacher professional learning communities (PLCs).

3. Project Design and Management Plan

3.1 Goal

The long-term goal of INFACT is to broaden participation in STEM, particularly by leveraging the intersection between learning variability and CT. This goal will provide CS opportunities more equitably for all learners, and has the potential to strengthen our future workforce by including learners with critical talents that may otherwise go untapped.

3.2 Objective

The objective of INFACT is to design, develop, implement, and study a comprehensive program that promotes the building of a conceptual foundation for CT within STEM in elementary school and the application of CT into STEM projects throughout middle school. INFACT will be designed to prepare all upper elementary- and middle-school learners, with respect to learning

variability, by aligning differentiated learning materials, learning assessments, and teacher PD.

To achieve this objective, we have outlined a 4-year timeline of major activities (Table 5).

Table 5: Timeline of Major Activities to Achieve INFACT Objective

Year 1				Year 2				Year 3				Year 4			
10/19	1/20	4/20	7/20	10/20	1/21	4/21	7/21	10/21	1/22	4/22	7/22	10/22	1/23	4/23	7/23
Design Research						Evaluation Research					Analysis	Dissemination			
Mtg 1				Mtg 2		Impact Study 1		Impact Study 2				Mtg 3			
CT Progression Design						Data Collection for Outcome Measures						Prepare for Scale Up			
Learning Materials and Assessment Curation and Integration							Data Collection on Implementation					Publish Findings and Models			
PD Design															

3.3 Anticipated Learning Outcomes

INFACT is intended to improve learners' mastery of CT practices; their self-efficacy as CT problem solvers; their self-efficacy as STEM problem-solvers; and their desire to participate in more STEM activities. In particular, we intend for INFACT to narrow the gap between neurodiverse learners and the general population in these areas. We anticipate that learners in STEM classes who use INFACT (treatment condition) will show more improvement in these outcomes than classes that use other materials for the same amount of time to teach CT in non-INFACT STEM classes (control condition). We also anticipate that learners who typically struggle academically (e.g., are identified by an IEP) may benefit the most from INFACT, and that they may reveal CT practices in the INFACT learning assessments that are not revealed in current forms of CT assessment.

3.4. Outcome Measures

All participants in the study will take grade-appropriate pre- and post-tests on INFACT outcome measures including a) **Student Mastery of CT Practices** (Rowe et al., 2019), and b) **Student**

Disposition and Self-Efficacy in CT and STEM problem solving (e.g., Meluso et al., 2012). Student Mastery of CT practices will be measured through items that focus on fundamental practices of CT. We will use online logic puzzles previously designed and validated for grades 3–8 used in TERC's prior research (Rowe et al., 2019) to measure learners' performance in Problem Decomposition, Pattern Recognition, Abstraction, and Algorithm Design. Figure 1 shows an example of an item assessing whether or not a student can abstract (or generalize) a rule from the patterns.
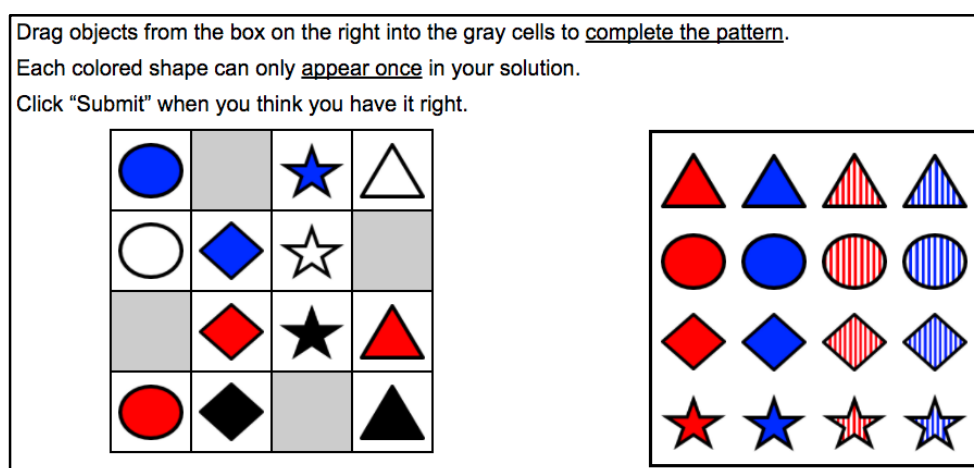

Figure 1: Example CT Mastery Outcome Measure: Abstraction

To expand and refine these items, while retaining their non-reliance on heavy text or coding nomenclature, the consortium will draw from an active body of research currently underway in CT assessment (e.g., Basu et al., 2019; Grover, 2017, 2019; Rowe et al., 2019; Tissenbaum et al., 2018). Many CT assessments emerging in the field rely on the construction or analysis of coding artifacts (Moreno-León & Robles, 2015; Ota, Morimoto, & Kato, 2016) or textual instruments (e.g., Dagienė, Stupurien, & Vinikien, 2016; Izu et. al, 2017) or a combination (Dagiene & Futschek, 2008; Wiebe et. al., 2019). The INFACT consortium will curate promising items and adapt their delivery to accommodate for learner variability. By curating, adapting, and correlating a well-chosen selection of these CT outcome measures, we will also contribute to the

field the first comprehensive set of standardizable outcome measures of CT practices that are suitable for a broad range of learner variability.

Since neurodiverse learners may struggle with text-based questionnaires, we will also adapt existing textual instruments for self-efficacy (e.g., de Cássia Martinelli et al., 2008) and STEM dispositions (e.g., Knezek, et al., 2012) into an interactive format, based on guidance from Bandura (2006). By having a simulated student give voice to low self-efficacy or low disposition, adaptations may lessen social desirability bias (cf., Nederhof, 1985; Fisher, 1993). The re-validation of these instruments is part of the Impact Study 1 design.
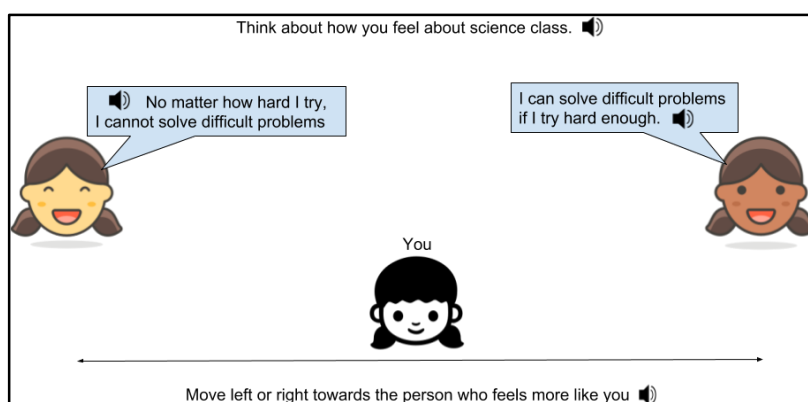

Figure 2: Example of Modified Self-Efficacy Item

In addition to these overarching pre-post measures for all participants, when available we will include information from each participant's interactions with any INFACT adaptive interactive formative assessments, as well as information from teachers about other assessments used during the implementation period to build a broader measure of each learner's outcomes.

4. Research and Evaluation Plan

To achieve the research objectives (Table 6), the consortium will conduct nearly two years of iterative design research and development of the comprehensive. Impact Study 1 near the end of year 2 will serve as a pilot test of the program and research instruments to prepare for Impact

Study 2 in year 3, which is a quasi-experimental study of at least 960 students to measure the impact of INFACT on their CT practices and their self-efficacy in STEM and CT.

Table 6: Research Objectives with Outcome, Process, Sample, and Methods

Participatory Design	Impact Study 1	Impact Study 2	Implementation Study
CT Progression with aligned learning materials, assessment, and PD	Baseline measurements of CT Practice	Potential evidence of impact on broad range of learners in grades 3–8	Lessons on potential issues of scale-up
Iterative Design Sessions with INFACT activity sequences and supports	Re-validation of modified instruments	Implementation for up to 10 hours of CT-infused STEM instruction in treatment and control classes	
Informal groups of learners in grades 3–8 with at least 50% with IEP status	8 Classes with 160 learners in grades 3–8 with at least 20% with IEP status	40 Classes with 960 learners in grades 3–8 with at least 20% with IEP status	
Observations and Think Aloud Interviews	Mastery of CT Practices Self-Efficacy and Disposition in STEM	Mastery of CT Practices Self-Efficacy and Disposition in STEM	Classroom Observations and Teacher Interviews

4.1 Design Research

Guided by the Diversity for Design (D4D) framework (Benton et al., 2014), we will conduct an extensive participatory design process of the program elements (Druin, 1999) in the first 2 years of the project.

4.1.1 Sample. We will work with groups of diverse learners in grades 3–8, drawing samples from each of the schools and districts where the consortium of researchers is already working (Table 1). These audiences include a broad range of learner variability and leverage longstanding research partnerships from the consortium's prior work (see letters of collaboration).

4.1.2 Design Research Questions. During the Participatory Design Process, we will test versions of selected INFACT materials appropriate for the given audience. For example, in elementary classes we might use a sequence or combination of a) the Zoombinis game with

associated offline activities to build foundational knowledge of CT practices; b) Sphero activities in mathematics; and c) a puzzle station in classes where learners solve CT problems collaboratively. Researchers at each setting will observe, collect recordings and artifacts, and informally interview learners and teachers during design testing. The dyad or triad of research teams working on any one sequence will meet frequently through the design process to exchange and interpret findings. The questions guiding the design research include:

DRQ1: Which features of the INFACT activity sequences support a CT progression? How can those features be generalized and expanded upon with other INFACT activities?

DRQ2: What types of features of INFACT leverage the intersection between CT and EF? How can the materials be optimized to serve a broad range of learner variability?

4.1.3 Design Research Methods. The design research process will begin with a 3-day meeting at TERC for the consortium to lay out all our existing materials and find natural sequences and combinations. By the end of the meeting, we will have the skeleton of a CT progression for grades 3–8, and will have identified gaps that require new development. Throughout year 1 and 2, teams whose projects are complementary in a sequence will work together in small groups to refine a set of activities that focus on one conceptual thread of CT or one type of CT application. We will conduct approximately 2–3 design sessions for each of about 5 activity sequences throughout the first year, selecting participant groups to ensure broad learner variability (as identified by classroom teachers) in the overall sample. During design sessions, we will introduce an iterative sequence of paper prototypes and functional working prototypes for the sequences of INFACT activities. Because we are working with young, neurodiverse learners who may have trouble expressing their understandings in writing or on a test, we will use a think-aloud protocol along with manipulatives to reveal their CT and STEM thinking as they proceed

through the INFACT activities. When necessary and available, we will work with learners'
regular learning specialists to communicate in the most effective manner during the think-aloud.

4.2 Project Evaluation

New Knowledge Organization Ltd. (NKO) will serve as external evaluators for this project,
conducting a series of quantitative and qualitative studies to provide valid and reliable data on
relevant outcomes, as well as guidance for effective strategies.

INFACT focuses in narrowing the gap that exists because of learner variability. To best
represent this audience, evaluators will disaggregate data by students' IEP status. To assess the
impacts of INFACT, evaluators will implement **Impact Study 1: Baseline** to establish baseline
measurements for the target outcomes and validate modifications to standardized instruments;
and **Impact Study 2: Aggregate Performance Improvement** to examine overall and specific
improvements for target populations across the INFACT implementation.

4.2.1 Overall Quasi-Experimental Design. The INFACT research design incorporates the
student-level variable of learning needs (2 levels: students with IEPs and students without such
plans), as well as the classroom-level variables of grade (2 levels: elementary and middle
school), student access to technology (2 levels: high and low), student activity choice (2 levels:
high and low), and treatment condition (2 levels: treatment and control classes). Crossing these
factors yields a design with two types of learning needs, nested in eight types of classrooms, and
divided in two treatment conditions (Table 7). Because the design assigns groups at the
classroom level but is interested in individual student outcomes, evaluators will treat our design
as a cluster-level assignment adhering to the What Works Clearinghouse (WWC) criteria
(USDoE, 2017, p. 19ff.). As per WWC (pp. 9–14) evaluators will track overall and differential
attrition in design conditions.

Table 7: Research Design for Impact Study 2

		Treatment								Control							
		Elementary				Middle School				Elementary				Middle School			
		High Student Choice		Low Student Choice		High Student Choice		Low Student Choice		High Student Choice		Low Student Choice		High Student Choice		Low Student Choice	
		High Tech Access	Low Tech Access	High Tech Access	Low Tech Access	High Tech Access	Low Tech Access	High Tech Access	Low Tech Access	High Tech Access	Low Tech Access	High Tech Access	Low Tech Access	High Tech Access	Low Tech Access	High Tech Access	Low Tech Access
IEP Status	Yes	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
	No	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45

The Core Dependent Measures include Mastery of CT practices as well as STEM disposition and self-efficacy as CT and STEM problem-solving. Measurements of Mastery of CT practices will derive from the CT learning outcomes described in section 3.4, specifically the pre- and post-test items on Problem Decomposition, Pattern Recognition, Abstraction, and Algorithm Design. Measurements of STEM disposition and self-efficacy as CT and STEM problem solvers will derive from the interactive instruments that are modified from standard items.

4.2.2 Impact Study 1: Baseline and Validation. During Year 2, evaluators will (1) modify and re-validate interactive versions of previously validated text-based instruments of self-efficacy and STEM dispositions, and (2) work with the consortium to establish baseline measurements for the curated and adapted CT learning outcome instruments, as per WWC (2017, pp. 27).

4.2.2.1 Impact Study 1 Sample. Evaluators will implement Impact Study 1 in 8 classrooms with at least one per grade level in grades 3–8. We will draw from our networks to build a prospective sample for both Impact Studies, ensuring a 20% representation of students with IEPs. We will select 8 of these classes that are using CT in Spring of year 2 for Impact Study 1.

4.2.2.2 Re-Validating Self-Efficacy Metrics. Evaluators will validate the interactive self-efficacy scale and STEM disposition for the pilot audience against previously validated text-based instruments with an item factor analysis, examination of communalities and measures of association, and reliability and stability testing.

4.2.2.3 Baseline Measurements of Target Outcomes. Researchers will establish baseline measurements for the Core Dependent Measures to: 1) assess whether the baseline classrooms are representative of the entire implementation and if not, make statistical adjustments as per WWC (2017); 2) derive the prior probabilities needed for the Markov chain Monte Carlo (MCMC) methods for multi-response general linear mixed models (MGLMM) used in later studies (cf. Hadfield, 2010; Berridge & Crouchley, 2011); and 3) test TERC's data architecture, Data Arcade, to collect baseline data and match anonymous learner IDs to data streams.

4.2.3 Impact Study 2: Aggregate Performance Improvement. During Year 3 of the proposed project, the evaluators will answer the following research questions:

> **IS2.RQ:** To what extent does INFACT build a conceptual foundation for CT within STEM in elementary school, and application of CT in STEM projects in middle school?

> **IS2.RQa:** To what extent do improvements differ between the target outcomes; in other words, how broad are the effects of INFACT on CT improvements?

4.2.3.1 Impact Study 2 Sample. INFACT will be used in approximately 40 classes in participating schools. Digital Promise will recruit the bulk of these teachers from the League of Innovative Schools, which includes over 100 schools districts across 33 states serving over 3 million K–12 students, over 40% of which have started offering CT and introductory CS instruction. TERC and other consortium members will supplement with the network of teachers who have engaged in CT research in prior research (e.g., TERC's recent national study of over 50 classes using Zoombinis to promote CT in grades 3–8) to ensure a representative sample with a broad range of learner variability.

Evaluators will select classes in schools that have at least two equivalent classes eligible for the study to allow for parity between treatment and control classes. When possible, evaluators

will use two sections taught by the same teacher to minimize extraneous variability. Treatment class teachers will implement INFACT content for at least 10 classroom hours. The extent to which each component of INFACT is used during implementation will be recorded by the teacher. Control teachers will use their regular curriculum for CT for a similar amount of time. Both types of classes will undergo data collection for all measures. Evaluators will use TERC's *DataArcade* to collect the data and track anonymous learner IDs across pre-INFACT and post-INFACT data collection on learner outcomes.

4.2.3.2 Impact Study 2 Design. The repeated measures (pre-INFACT versus post-INFACT) quasi-experimental design will account for learning needs, classroom types, and treatment condition (Table 7). This design will require at least 30 students for each of the 32 factor cells, totaling at least 960 students. The design will control for the socio-economics of the schools, time spent on INFACT at the class level, and teachers' self-efficacy in teaching STEM (FIEI, 2012). If possible, the design will also control for pre-INFACT student grades and STEM-related standardized test scores.

4.2.3.3 Impact Study 2 Analysis. MCMC-MGLMM analysis will allow evaluators to test the effects of INFACT on the Core Dependent Measures while accounting for group-level differences in the nested quasi-experimental factors. Evaluators will use the baseline measurements obtained during Impact Study 1 (see 4.2.2.2 above) to establish the prior (naive) probabilities of measurements without the random effects of nested factors.

To ensure parity, evaluators will compare the quasi-experimental design model and a second model based on propensity score analyses (cf. WWC 2017, pp. 31–32) of the nesting factors (learning needs, grade levels, and school types) and the school-level covariates (average socio-economic status, STEM grades, and test scores) to describe the effects of INFACT. In

addition to testing the multivariate effects of INFACT on the dependent measures (answering IS2.RQ), evaluators will also contrast improvements on the four CT practices (answering IS2.RQa).

4.2.4 Implementation Research. In parallel with the aggregate learning assessment (Impact Study 2), the implementation study will qualitatively characterize and assess the validity of our classroom-level variables (see section 4.2.1), seeking to answer the following research question:

IMPRQ1: How does INFACT implementation vary in different types of classrooms?

In Year 3, two evaluators will conduct one full week of observational research in eight classrooms representing maximum variation among the implementing classes. Observation will take place when the teacher first introduces INFACT activities in the classroom. One evaluator will focus on individual student engagement using the Baker Rodrigo Ocumpaugh Monitoring Protocol (BROMP) (Ocumpaugh, Baker, & Rodrigo, 2015), and the other will focus on the whole classroom by using an ecological observation protocol (cf. van Lier, 1997). These observations will be time-aligned with each other and with time-stamped *DataArcade* logs for analysis, enabling us to draw connections between classroom-level phenomena and individual engagement. In addition, classroom interactions will be audio-recorded and CT-focused activities will be fully transcribed, allowing evaluators to use discourse-analytic methods to compare teacher and student framing of computational thinking, following other research on student socialization into subject-specific language use (e.g., Kobayashi, Zappa-Hollman, & Duff, 2017; Duff, 2010; O'Connor, 2015; Chrisomalis, 2015; etc.).

5. Project Team and Responsibilities

A team of PIs from a consortium of organizations will lead this project under the direction of TERC's PI, Asbell-Clarke. In the first 2 years, the consortium teams will meet in person once per

year and bi-weekly by videoconference to progress on the iterative design research of the INFACT program. Consortium members have ample experience in distributed collaborations and are comfortable using cloud-based tools for design and production. Underlying the consortium's design efforts, each team will conduct local design research sessions that will feed into the collective vision and development. In year 3, NKO will conduct the impact and implementation studies. A skeletal team from the consortium will be on call for support needed during the implementation, but this will be minimal so that future scale up of the resulting model is feasible. In year 4, the consortium will resume in person and bi-weekly virtual meetings to disseminate findings and prepare the model for future scale-up.

Dr. Jodi Asbell-Clarke will oversee the TERC leadership of all aspects of the INFACT project. She has led numerous federally-funded projects with national impact in game-based learning and CT education, including an international consortium. She will facilitate design meetings to efficiently and effectively share information from distributed design research sessions to provide a coherent and clear vision of the emergent CT progression and the workflow on how to achieve project outcomes. She will be supported at TERC by a project director (Kelly Paulson) as well as a team of researchers and developers including Drs. Elizabeth Rowe and Mia Almeda, learning scientists and EDM specialists; Dr. Ibrahim Dahlstrom-Hakki, cognitive psychologist and learner variability specialist; and Teon Edwards, curriculum and technology designer.

Dr. Quinn Burke is Senior Research Scientist at Digital Promise Global (DPG) and is a member of the Learning Sciences Research division. Dr. Burke will lead development of teacher (grades 3–8) professional development modules based on the alignment of CT competencies with learner variability factors; he will also lead the recruitment of League of Innovative Schools

districts for the impact research and evaluation. Mr. Viv Vuchic, a specialist in learner variability, and Dr. Judi Fusco, a teacher education specialist, will support DPG's work.

Dr. Shuchi Grover is a Senior Research Scientist at Looking Glass Ventures and Research Scholar at Stanford University's Human Sciences and Technologies Advanced Research (H-STAR) Institute. Her NSF-funded projects study CT learning and assessment in varied PK–12 contexts including introductory CS education as well as STEM classrooms that integrate STEM and CT. Conducting design research in middle schools in the San Francisco Bay Area, Dr. Grover's team will investigate the use of learning activities and assessment targeting key concepts for CT and STEM such as variables, algorithmic thinking, and abstraction in grade 6–8 CS and STEM classrooms.

Dr. Maya Israel is is an Associate Professor of Educational Technology in the School of Teaching and Learning at the University of Florida (UFLA). Her NSF funded research focuses on supporting learner variability in STEM. Her team will contribute models of metacognitive scaffolding for CT and STEM and will conduct related design research.

Dr. Fengfeng Ke is an Associate Professor of Educational Psychology and Learning Systems at Florida State University (FSU). She leads grants from NSF and foundations to study inclusive design with interactive digital technologies. Dr. Ke's team will contribute work on VR assessments for learners with Autism and conduct related design research.

Dr. David Weintrop is an Assistant Professor at the University of Maryland with a joint appointment between the College of Education and College of Information Studies. His research team will contribute Sphero.Math curriculum which investigates ways to use robotics as a mechanism for integrating CT into elementary (3rd and 4th grade) mathematics classes. His prior research has looked extensively at the potential for situating CT learning within STEM contexts.

Dr. John Voiklis of NKO will lead the design, analysis, and publication of impact evaluation results. Dr. Voiklis is a social-cognitive scientist who uses data science methods to study problem solving and reasoning about concepts and social norms. He has taught applied statistics to researchers, business students, and prospective teachers at Teachers College, Columbia University.

INFACT will also benefit from advisors who lend expertise and materials but will not participate in the design research efforts of the consortium. These include James Lester (NCSU) and José Blackorby (CAST), who will attend design meetings and help review materials for their alignment with promising evidence in the areas of CT-infused STEM learning and accessibility for a broad range of learner variability.

6. Dissemination Plan

During the design research, and as a key part of the final year of the project, the consortium will position the program for distribution through an educational scale-up program. We will attend salient conferences (e.g., NSTA, CSTA, and ISTE) to promote the materials to teachers as well as to districts and materials distributors. During the final in-person consortium meeting in year 4, the team will put appropriate IP and commercialization agreements in place to distribute the comprehensive program in an scalable manner. The consortium will consult with current distributors of educational programs to help place the program for widespread dissemination in a future scale-up endeavor. Finally, the consortium members will co-present the models for CT-infused STEM teaching and learning, and models for embedded scaffolds for learner variability that stem from this project work at relevant research conferences such as AERA, IMBES, and SIGCSE, and they will co-publish papers on this interdisciplinary work.

References

Asbell-Clarke, J. & Bradbury, S. (2017). Project-based Learning and Assessment for At-Risk Youth. Paper presented at the Future of Education Technology Conference (FETC), Orlando.

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan (Eds.), Self-efficacy beliefs in adolescents (pp. 307–337). Greenwich, CT: Information Age.

Baron-Cohen, S., Ashwin, E., Ashwin, C., Tavassoli, T., & Chakrabarti, B. (2009). Talent in autism: hyper-systemizing, hyper-attention to detail and sensory hypersensitivity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1522), 1377–1383.

Baroody, A. J., Bajwa, N. P., & Eiland, M. (2009). Why can't Johnny remember the basic facts?. *Developmental Disabilities Research Reviews*, *15*(1), 69–79. doi:10.1002/ddrr.45).

Barr, V., & Stephenson, C. (2011). Bringing computational thinking to K-12: what is Involved and what is the role of the computer science education community?. *Inroads*, *2*(1), 48-54.

Basu, S., McKlin, T., Rutstein, D., Weintrop, D., Yadav, A., & Burke, Q. (2019). Assessing Computational Thinking: A "Landscape" symposium about where we are at & where to go. Symposium to be held at the Annual Meeting of the American Educational Research Association (AERA) 2019. Toronto, Ontario, Canada.

Beck, J. E., & Gong, Y. (2013, July). Wheel-spinning: Students who fail to master a skill. In International Conference on Artificial Intelligence in Education (pp. 431–440). Springer, Berlin, Heidelberg.

Benton, L., Vasalou, A., Khaled, R., Johnson, H., & Gooch, D. (2014, April). Diversity for design: a framework for involving neurodiverse children in the technology design process. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3747–3756). ACM.

Berridge, D. M. and Crouchley, R. (2011). Multivariate generalized linear mixed models using R. CRC press, Boca Raton, FL.

Brendefur, J., Johnson, E. S., Thiede, K. W., Smith, E. V., Strother, S., Severson, H. H., & Beaulieu, J. (2015). Developing a comprehensive mathematical assessment tool to improve mathematics intervention for at-risk students. *International Journal for Research in Learning Disabilities, 2*(2), 65–90.

Burke, Q. (2016). Mind the metaphor: Charting the rhetoric about introductory programming in K-12 schools. On the Horizon. 24(3).

Chong, S. L., & Siegel, L. S. (2008). Stability of computational deficits in math learning disability from second through fifth grades. *Developmental Neuropsychology*, *33*(3), 300–317.

Chrisomalis, S. (2015). What's so improper about fractions? Prescriptivism and language socialization at Math Corps. *Language in Society, 44*(1), 63–85.

Dagiene, V. and Futschek, G. 2008. Bebras international contest on informatics and computer literacy: Criteria for good tasks. In International Conference on Informatics in Secondary Schools-Evolution and Perspectives. Springer, 19–30.

Dagiene, V., Stupurien, G. and Vinikien, L. 2016. Promoting Inclusive Informatics Education Through the Bebras Challenge to All K-12 Students. In Proceedings of the Proceedings of the 17th International Conference on Computer Systems and Technologies 2016. ACM, 407–414.

de Cássia Martinelli, S., Bartholomeu, D., Gakyia Caliatto, S., and de Grecci Sassi, A. (2008). Children's Self-Efficacy scale: Initial psychometric studies. *Journal of Psychoeducational Assessment*, 27(2):145–156.

Denkla, M. B. (1994). *Measurement of executive function*. Baltimore, MD, US: Paul H Brookes Publishing Co.

Duff, P.A. (2010). Language socialization into academic discourse communities. *Annual Review of Applied Linguistics*, *30*, 169–192.

Druin, A. (1999). *The Role of Children in the Design Technology*.

Fisher, R. J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of Consumer Research*, 20(2):303–315.

Friday Institute for Educational Innovation (FIEI) (2012). Teacher Efficacy and Beliefs toward STEM Survey. Raleigh, NC.

Fuchs, L.S., Powell, S.R., Seethaler, P.M., Cirino, P.T., Fletcher, J.M., Fuchs, D., & Hamlett, C.L. (2011). The development of arithmetic and word-problem skill among students with mathematics disability. In J.M. Kauffman & D.P. Hallahan (Eds.), Handbook of special education. NY: Routledge.

Grover, S. (2017). Assessing algorithmic and computational thinking in K-12: Lessons from a middle school classroom. In Emerging research, practice, and policy on computational thinking (pp. 269–288). Springer, Cham.

Grover (2019). An Assessment for Introductory Computational Concepts for Middle School. Paper to be presented at the National Council for Measurement in Education (NCME). Toronto.

Grover, S., Jackiw, N., & Lundh, P. (2019). Concepts before coding: non-programming interactives to advance learning of introductory programming concepts in middle school. *Computer Science Education*, 1–30.

Hadfield, J. D. (2010). MCMC methods for Multi-Response generalized linear mixed models: The MCMCglmmR package. Journal of Statistical Software, 33(2).

Immordino-Yang, M. H., Darling-Hammond, L., & Krone, C. (2018). The Brain Basis for Integrated Social, Emotional, and Academic Development. *National Commission on Social, Emotional, and Academic Development*.

Israel, M., Ray, M. J., Maa, W. C., Jeong, G., Lee, C., Lash, T., & Do, V. (2018). School embedded and district-wide instructional coaching in K-8 computer science: Implications

for including students with disabilities. *Journal of Technology and Teacher Education*, 26(3), 471–501.

Izu, C., Mirolo, C., Settle, A., Mannila, L. and Stupuriene, G. Exploring Bebras Tasks Content and Performance: A Multinational Study. Informatics in Education, 16, 1 (2017), 39–59.

Kaczmarczyk, L., Dopplick, R., & EP Committee. (2014). Rebooting the pathway to success: Preparing students for computing workforce needs in the United States. *Education Policy Committee, Association for Computing Machinery.(ACM, New York, 2014). Retrieved from http://pathways. acm. org/ACM_pathways_report. pdf*.

Kafai, Y.B. & Burke, Q. (2014). Connected code: Why children need to learn programming. Cambridge, MA: MIT Press.

Kai, S., Almeda, V.M., Baker, R. S., Shechtman, N., Heffernan, C., & Heffernan, N. T. (2017). Modeling Wheel-spinning and Productive Persistence in Skill Builders. In EDM.

Ke, F., & Lee, S. (2016). Virtual reality based collaborative design by children with high-functioning autism: Design-based flexibility, identity, and norm construction. *Interactive Learning Environments*, *24(7)*, 1511–1533.

Ke, F., & Moon, J. (2018). Virtual collaborative gaming as social skills training for high-functioning autistic children. *British Journal of Educational Technology*, *49(4)*, 728–741.

Ke, F., Shute, V., Clark, K., & Erlebacher, G. (2019). *Interdisciplinary design of the game-based learning platform: A phenomenological examination of the integrative design of game, learning, and assessment*. New York, NY: Springer.

Knezek, G., Christensen, R., Tyler-Wood, T., Periathiruvaldi, S., Alexander, C., Mayes, G., Owens, C. & Magoun, D. (2012). Measurement of STEM Dispositions in Elementary School Students. In P. Resta (Ed.), Proceedings of SITE 2012–Society for Information Technology & Teacher Education International Conference (pp. 1052–1059). Austin, Texas, USA: Association for the Advancement of Computing in Education (AACE).

Kobayashi, M., Zappa-Hollman, S., & Duff, P. (2017). Academic discourse socialization. In P. Duff & S. May (Eds.), *Language socialization* (volume). *Encyclopedia of language and education*, 3rd ed. (pp. 239–254). Cham, Switzerland: Springer International.

Lee, H., Plass, J. L., & Homer, B. D. (2006). Optimizing cognitive load for learning from computer-based science simulations. *Journal of educational psychology*, *98*(4), 902.

Martinuzzi, A., & Krumay, B. (2013). The good, the bad, and the successful–how corporate social responsibility leads to competitive advantage and organizational transformation. *Journal of change management*, *13*(4), 424–443.

Mayes, S. D., & Calhoun, S. L. (2006). Frequency of reading, math, and writing disabilities in children with clinical disorders. *Learning and Individual Differences*, *16*(2), 145–157. https://doi.org/10.1016/j.lindif.2005.07.004Mesibov, G. B., Shea, V., & Schopler, E. (2005). *The TEACCH approach to autism spectrum disorders*. Springer Science & Business Media.

Meluso, A., Zheng, M., Spires, H. A., & Lester, J. (2012). Enhancing 5th graders' science content knowledge and self-efficacy through game-based learning. *Computers & Education*, *59*(2), 497–504.

Meyer, A., Rose, D.H., & Gordon, D. (2014). *Universal design for learning: Theory and Practice*. Wakefield, MA:CAST Professional Publishing.

Mislevy, R., & Riconscente, M. (2005). Evidence-centered Assessment Design: Layers, Structures, and Terminology (PADI Technical Report 9).

Moran, S., & Gardner, H. (2018). Hill, skill, and will: executive function from a multiple-intelligences perspective. In Meltzer, L. (Ed.). *Executive function in education: From theory to practice*. Guilford Publications.

Moreno-León, J., Robles, G., & Román-González, M. (2015). Dr. Scratch: Automatic analysis of scratch projects to assess and foster computational thinking. *RED. Revista de Educación a Distancia*, (46), 1–23.

National Research Council (2001). *Knowing What Students Know.: The Science and Design of Educational Assessment*. National Academies Press: Washington.

National Science Board. (2015). *Revisiting the STEM workforce: A companion to science and engineering indicators 2014*. Arlington.

National Science Foundation, National Center for Science and Engineering Statistics (2015). Women, minorities, and persons with disabilities in science and engineering. Special Report NSF 15-311. Arlington, VA. Retrieved from https://www.nsf.gov/statistics/2015/nsf15311/

National Science Foundation, National Center for Science and Engineering Statistics, Scientists and Engineers Statistical Data System (SESTAT), https://www.nsf.gov/statistics/sestat/, and the National Survey of College Graduates (NSCG) (1993, 2015), https://www.nsf.gov/statistics/srvygrads/. From the National Science Board Science and Engineering Indicators 2018.

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology*, 15(3):263–280.

O'Connor, B.H. (2015). "I heard it wasn't really a myth": Enacting and contesting expertise in an Arizona science classroom. Linguistics & Education, 31, 30–43.

Ocumpaugh, J., Baker, R.S. & Rodrigo, M.M.T. (2015). *Baker Rodrigo Ocumpaugh Monitoring Protocol (BROMP) 2.0 Technical and Training Manual*.

Orton, K., Weintrop, D., Beheshti, E., Horn, M., Jona, K., & Wilensky, U. (2016). Bringing Computational Thinking Into High School Mathematics and Science Classrooms. In Proceedings of ICLS 2016 (pp. 705–712). Singapore.

Ota, G., Morimoto, Y., & Kato, H. (2016, September). Ninja code village for scratch: Function samples/function analyser and automatic assessment of computational thinking concepts. In *2016 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)* (pp. 238–239). IEEE.

Pei, C. (Yu), Weintrop, D., & Wilensky, U. (2018). Cultivating Computational Thinking Practices and Mathematical Habits of Mind in Lattice Land. Mathematical Thinking and Learning, 20(1), 75–89.

Ray, M., Israel, M., Lee, C., & Do, V. (2018). A cross-case analysis of instructional strategies to support participation of K-8 students with disabilities in CS for All. Proceedings of the 49th Association for Computing Machinery (ACM) Technical Symposium on Computer Science Education (900–905). ACM.

Rose, T. Rouhani, P., & Fischer, K. (2013) The Science of the Individual. Mind, Brain and Education, 7(3), 152–158.

Rowe, E., Asbell-Clarke, J. & Baker, R. (2015). Serious game analytics to measure implicit science learning. In C.S. Loh, Y. Sheng, & D. Ifenthaler *Serious Game Analytics: Methodologies for Performance Measurement, Assessment, and Improvement* (343–360). Springer Science+Business.

Rowe, E., Asbell-Clarke, J., Baker, R., Eagle, M., Hicks, A., Barnes, T., Brown, R., & Edwards, T., (2017). Assessing implicit science learning in digital games. Computers in Human Behavior, 76, 617-630. DOI: 10.1016/j.chb.2017.03.043

Rowe, E., Asbell-Clarke, J., Baker, R., Gasca, S., Bardar, E., & Scruggs, R. (2018, April). Labeling Implicit Computational Thinking in Pizza Pass Gameplay. Late-breaking work presented at the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 18), Montreal.

Rowe, E., Asbell-Clarke, J., Shute, V. Ventura, M., Almeda, V., & Bardar, E. (2019, under review). A Performance-based Measure of Computational Thinking in Grades 3-8. Manuscript submitted for publication.

Schneps, M. H., Brockmole, J. R., Sonnert, G., & Pomplun, M. (2012). History of reading struggles linked to enhanced learning in low spatial frequency scenes. *PLoS One*, *7*(4), e35724.

Semenov, A.D. & Zelazo, P.D. (2018). The Development of Hot and Cool Executive Function: A Foundation for Learning in the Preschool Years. In Meltzer, L. (Ed.). *Executive function in education: From theory to practice*. Guilford Publications.

Shute, V. J., & Kim, Y. J. (2014). Formative and stealth assessment. In *Handbook of research on educational communications and technology* (pp. 311–321). Springer, New York, NY.

Thiede, K. W., Brendefur, J. L., Osguthorpe, R. D., Carney, M. B., Bremner, A., Strother, S., ... & Jesse, D. (2015). Can teachers accurately predict student performance?. Teaching and Teacher Education, 49, 36–44.

Tomlinson, C. A., & Moon, T. R. (2013). *Assessment and student success in a differentiated classroom.* Alexandria, Va.: ASCD.

Tissenbaum, M., Sheldon, J., Sherman, M. A., Abelson, H., Weintrop, D., Jona, K., Horn, M., Wilensky, U., Basu, S., Rutstein, D., Snow, E., Shear, L., Grover, S., Lee, I., Klopfer, E., Jayathirtha, G., Shaw, M., Kafai, Y., Mustafaraj, E., Temple, W., Shapiro, R. B., Lui, D., & Sorensen, C. (2018). The State of the Field in Computational Thinking Assessment. In

Kay, J. and Luckin, R. (Eds.). Rethinking Learning in the Digital Age: Making the Learning Sciences Count, 13th International Conference of the Learning Sciences (ICLS) 2018. London, UK.

U.S. Department of Education (WWC, 2017), Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, What Works Clearinghouse.

U.S. Department of Education, Office of Special Education Programs, Individuals with Disabilities Education Act (IDEA) database, retrieved July 10, 2017, from https://nces.ed.gov/programs/coe/indicator_cgg.asp#info
See *Digest of Education Statistics 2017*, table 204.30.

van Lier, L. (1997). Observation from an ecological perspective. *TESOL Quarterly, 31*, 783–787.

von Karolyi, C. (2001). Visual-spatial strength in dyslexia: Rapid discrimination of impossible figures. *Journal of learning disabilities*, *34*(4), 380–391.

Wang, S. How Autism Can Help You Land a Job. *The Wall Street Journal*, March 27, 2014.

Weintrop, D., Beheshti, E., Horn, M., Orton, K., Jona, K., Trouille, L., & Wilensky, U. (2016). Defining Computational Thinking for Mathematics and Science Classrooms. *Journal of Science Education and Technology*, 25(1), 127–147.

Wei, X., Yu, J. W., Shattuck, P., & Blackorby, J. (2017). High School Math and Science Preparation and Postsecondary STEM Participation for Students with an Autism Spectrum Disorder. *Focus on Autism and Other Developmental Disabilities, 32*(2), 83–92.

Wiebe, E., London, J., Aksit, O., Mott, B. W., Boyer, K. E., & Lester, J. C. (2019, February). Development of a Lean Computational Thinking Abilities Assessment for Middle Grades Students. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education* (pp. 456–461). ACM.

Willcutt, E. G. (2014). Behavioral Genetic Approaches to Understand the Etiology of Comorbidity. In S. H. Rhee & A. Ronald (Eds.), *Behavior Genetics of Psychopathology* (pp. 231–252). Springer New York.