Boston College

Lynch School of Education and Human Development

Department of Teaching, Curriculum, and Society

Curriculum and Instruction

PROMOTING SELF-EFFICACIOUS COMPUTER SCIENCE EDUCATION: FINDINGS FROM A SMART GREENHOUSE PROJECT, A REVIEW OF AN AI CURRICULUM, AND AN ANALYSIS OF AN AI CONCEPT INVENTORY

Dissertation

by

YIHONG CHENG

submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

2022

Abstract

Computer science literacy is the key to surviving and thriving in the digital era. Unfortunately, given the negative stereotypes about who does computer science related work and what such work entails, many individuals are dissuaded from learning more about computer science and lack belief in their competence in computer science. As such, this dissertation aims to identify ways to make computer science education more self-efficacious using three connected studies, including (1) a mixed methods study on an intervention project for non-STEM major college students, (2) a practitioner study on a novel curriculum for middle school students, and (3) a study on the internal structure of a novel concept inventory for AI concepts. Findings from the first study confirm the importance of providing learners with mastery experiences in terms of helping them developing self-efficacy in coding. Findings from the second study provide teachers with teaching tips they could use while teaching the AI curriculum in their classrooms. Findings from the third study reveal the strengths and weaknesses of the AI concept inventory in accurately measuring respondents' knowledge about AI.

1. Introduction

1.1 What is computer science literacy and why does it matter?

With the goal of creating computational systems with the capability of possessing human-like intelligence and interacting with the real world (Rosenbloom, 2013), computer scientists have made tremendous progress in the past few decades in areas like Internet of Things (IoT) and Artificial Intelligence (AI), and advancement in these fields is reshaping our society in unprecedented ways and carries the potential to bring more fundamental transformations to the world. As such, computational literacy, the "study of computers and algorithmic processes, including their principles, their hardware and software designs, their implementation, and their impact on society" (Tucker et al., 2003, p. 6), is an essential skill that today's students of all backgrounds need to acquire in order to survive and thrive in an increasingly digitalized world (Braun & Huwer, 2022). In response to this trend, strategic plans are being made around the world to facilitate the training of personnel skilled in automation and AI, in all education levels and extracurricular programs, to raise the competitiveness and vitality of digital economies (Topi et al., 2017).

These trainings are worth attending for many reasons for many important reasons. First, for individuals seeking to survive and thrive in the future job market, getting trained about these skills that will be playing increasingly critical roles in their careers gives them an edge that will very likely be favored by their employers, if not already required (Manyika et al., 2017). This is not only true for people who plan to be working in STEM (Science, Technology, Engineering, and Mathematics) fields due to the interdisciplinary nature of informatics competencies and their widespread impact on virtually all industries

(Marques, Von Wangenheim, and Hauck 2020). Future artists, for example, will likely be relying on Generative Adversarial Networks (GANs) as trustworthy assistants to facilitate with drawing (Wu, Seokin, & Zhang, 2021).

More importantly, beyond the need for the economic benefits and occupation al advantages, the prevalence of arising technologies calls for the development and dissemination of protocols for ethical and equitable uses of them. Unfortunately, compared to the rapid growth of technological advancement, insufficient ethical and philosophical preparations are in place, leaving the well-being of society at risk. In the words of Smith and Neupane (2018), this is a dangerous sign, because:

> If we continue blindly forward, we should expect to see increased inequality alongside economic disruption, social unrest, and in some cases, political instability, with the technologically disadvantaged and underrepresented faring the worst (p.12).

As such, the public needs to be knowledgeable about what does and does not constitute proper application of new technologies. This would enable them to both use new technologies to challenge the status quo (Calzada, 2021) and critically weigh the risks and benefits and take proper actions when their rights are harmed, which constitute movement toward ensuring that no one is put in a particularly advantaged or disadvantaged position to enjoy the transformative changes (Samoili et al., 2020).

1.2 What are some challenges and their solutions in computer science education?

The most salient challenge faced by educators and researchers dedicated to computer literacy education is the lack of research on and practices that promote the interdisciplinarity of computer science. While it is widely acknowledged that mathematics, albeit a discipline of its own, carries concepts and competencies that are applicable to various fields, the interdisciplinary applicability of computer science has been insufficiently studied. The majority of previous studies on computer science literacy tends to narrow their scopes to the nature of informatics competencies themselves (Wang et al., 2021) or exclusively focuses on the usefulness of these competencies in STEM subjects (Lyon & Magana, 2020). Neither of these approaches suffices to address what computer science literate individuals can do when they integrate their informatics competencies into more than just STEM fields and how best to cultivate them. That said, it's always challenging to envision the immense number of possibilities of game-changing innovations and make plans for fostering the competencies involved. As acknowledged by leading computer science education organizations, namely Institute of Electrical and Electronics Engineers (IEEE) and Association for Computing Machinery (ACM), in their jointly initiated Computing Curricula 2020 (CC2020) project, the development of education programs for computer science literacy remains largely at the proposal stage and has yet to result in a curriculum with common recognition.

Nevertheless, CC2020 pointed out that two directions of work are markedly promising: (i) creating competence based models of curricular designs and (ii) complementing the teaching of computer science concepts with ethical and philosophical considerations (CC2020 Task Force, 2020). These approaches transcend the specific domains of knowledge in the vast field of computer science and accurately reflect the core missions of computer science education that computer science educator around the world devote themselves to: to nurture individuals of all backgrounds who are not only competent within the domain of computer science itself, but also ready to creatively apply informatics competencies in whatever they do for life and work and critically evaluate and make use of new technologies and take actions when inequity is, intentionally or unintentionally, reified by these powerful tools.

1.3 What are some feasible, interim goals if not a full curriculum?

It is beyond the scope of this 3-paper format dissertation to either develop a full curriculum that is inclusive of (a) a representative set of competencies needed to form a meaningful framework that can be used to guide computer literacy education, or (b) all of the ethical considerations that should be taken in possible scenarios of applications of computer science. After all, even after searching through most of the major formal, professionally approved curricula available across regions and times, the CC2020 Task Force (2020) ended up endorsing none and predicted that it would take several more years for a curriculum worthy of recommendation to emerge. There are, however, several interim steps that can be taken to move closer to promote the quality of AI education, which I will focus on in this dissertation, which I summarize as raising self-efficacy, raising career interests, and dissecting AI concepts.

Raising Self-Efficacy. One known common beginning point of the various trajectories of development of all informatics competencies points to the development of self-efficacy (Maltese & Tai, 2011). In short, self-efficacy means one's beliefs in their capabilities to "organize and execute the courses of action required to manage prospective situations", as defined in the Social Cognitive Theory (SCT; Bandura, 1995, p. 2). Self-efficacy matters in almost all human activities because it largely lays the foundation of our inclination to succeed in a given situation (Bandura, 1997). Without sufficient self-efficacy, we are unlikely to succeed while facing obstacles of any kind even if we possess the knowledge and skills to do so, due to perceiving them as impossible to overcome and experiencing low incentives and motivation. In contrast, persons with high self-efficacy are likely to persist to work on challenging tasks, even when they encounter setbacks that they temporarily do not have the inventory of strengths to resolve, until they manage to succeed at last after continuously and actively searching for solutions (Kurbanoğlu, 2003).

Specifically in the field of computer science literacy education, self-efficacy is an important factor that deeply impacts the educational experience of learners, ranging from classroom interactions to career planning and beyond (National Science Foundation, 2017). Unfortunately, as evidenced by a cluster of empirical studies, starting from as early as middle school years, students develop increasingly low self-efficacy when it comes to computer science due to social persuasions that make them lose beliefs in or even despise developing a strong relationship with computer science (Maltese & Tai, 2011). For example, stereotypical beliefs that computer science is exclusively associated with White and Asian males constantly dissuade students of historically underrepresented in computer science from openly exploring their interests (Archer et al., 2017).

Raising Career Interests. Similarly, stigmatization of people working in computer science fields also greatly limit students' interest and participation in STEM. With little knowledge about what computer scientists really do and how informatics competencies are applied (Blotnicky et al., 2018), students often hold negative impressions about computer science related jobs and believe that they involve doing boring, uninteresting work in unpleasant surroundings and are cut off from other people (OECD, 2008). These stereotypes start to form during as early as elementary school (Luo et al., 2018) and culminate, during high school and college, into the perception that working in computer science related careers is all about memorizing facts that have little connection to the real world (Momsen et al., 2010) and that they are not as creative and relevant to their lives as other jobs (Masnick et al., 2010). Beliefs like these about computer science has widely led to low career interests and expectations in computer science (Luo, So, Wan, & Li, 2021), which has limited both the magnitude of and diversity in the workforce (Marginson et al., 2013), as predicted by Social Cognitive Career Theory (SCCT; Lent, Brown, and Hackett, 1994).

Needless to say, the aforementioned stereotypical beliefs are far from true. While some STEM concepts may be seen as independent of humanistic contexts, the studies of STEM are inherently human endeavors defined by their participants' unique lived experiences and perspectives (Franklin, 1995). Many "facts" taken for granted to be factual are but generalizations of educated guesses resulted from the representations of observations of natural phenomena, the execution of which requires creative sense-making, selecting of useful information, and problem-solving (Aikenhead, 1996). The misguided beliefs, however, are not easy to change because they are deeply rooted in learning activities embedded classroom interactions, testing, and beyond that dictate how students perceive what computer science means and how fit they are for doing related work. When the only images about computer scientists students have in mind are "geeks" and "boffins" rather than role models they can emulate (Archer et al., 2013), they do not see themselves as represented in computer science and experience difficulties even imagining what they would be doing in computer science related fields, not mentioning actually pursuing such careers (Cole and Espinoza, 2008).

Dissecting AI Concepts. An ultimate challenge that awaits individuals who manage to retain sufficient self-efficacy and career interest in AI when they finally enter AI education programs is the complexity of AI concepts and the connections among them. AI concepts are known for involving a variety of domains that are rather different from each other (Cui, Shang, and Chen, 2019). This characteristic brings a lot of difficulty to the implementation of AI education. From a curricular design perspective, for example, it is hard to determine a proper set of AI concepts and reasonably sequence them in ways that are appropriate for students entering the classroom with various cognitive levels of understanding of AI concepts (Mo, 2020). Meanwhile, due to the influence of popular media, beginning learners tend to hold many misunderstandings about AI, such as equating AI with robotics and automation, which requires much effort by AI educators to identify and correct (Hu, 2016).

In cases when misconceptions accompany their holders without being noticed even after they have engaged in deeper levels of learning, concept inventories have proven to be an effective tool for detecting (Hestenes, Wells, and Swackhamer, 1992). A concept inventory is a measurement tool that challenges respondents to answer a concise, representative set of questions about fundamental concepts in a discipline (Crouch & Mazur, 2001). In order that a concept inventory be successful, it would have to be capable of measuring a wide range of cognitive levels for two reasons. First, misconceptions can be especially hard to detect in tests where most questions can be answered through rote learning strategies, such as memorizing definitions and facts, that require low cognitive levels of understanding (Hestenes, Wells, and Swackhamer, 1992). Second, the mastery of knowledge takes a propagating sequence of learning activities involving the use of increasingly deep cognitive levels of understanding, as noted by Bloom's Taxonomy (Bloom, 1956). Without a complete set of items corresponding to each step in this trajectory, it would not be possible to locate which stage respondents are at and provide appropriate learning support.

1.4 Overall Research Questions

Given the aforementioned challenges in computer science education and related theories addressing them, I am proposing to answer the following questions in this 3 paper dissertation:

(i) How and in what ways did a novel, collaborative project based learning experience centered on the coding and building of a miniature tabletop smart greenhouse impact how a group of undergraduate students' computer science related self-efficacy in a hybrid learning space?

(ii) What are some computer science teaching standards that can be addressed by DAILy, a novel AI curriculum for middle school students, and what are some teaching tips based on observations of previous implementations?

(iii) What is the internal structure of a novel AI concept inventory in terms of (a) the AI concepts intended to be measured, (b) building blocks of AI literacy, and (c) cognitive levels defined by Bloom's Taxonomy, and what misunderstanding about AI can be detected?

2. Theoretical Backgrounds

Given the varied contexts addressed in this dissertation, I will be drawing on four theoretical frameworks, each of which is suitable for answering one research question for one of the three studies. For readability considerations and given the fact that the purpose of this dissertation is not to blend any theoretical frameworks together to form something new but to answer the three research questions, in the following subsections I will proceed to discuss each of the frameworks I have drawn upon one by one in the context of the corresponding study.

2.1 Social Cognitive Theory (Paper 1)

In this paper I will be drawing on Social Cognitive Theory (SCT; Bandura, 1986) as the primary framework that guides the research around Research Question 1. SCT is an appropriate choice as it specifically addresses the fundamental role self-efficacy plays in determining one's academic attainments and what factors may bring changes to an individual's self-efficacy, which I will explain in detail in this subsection.

Self-efficacy is defined as a person's belief in their capability to perform well in certain anticipated situations and an important determinant of the likelihood that they actually succeed (Bandura, 1997) and a central piece of SCT, a theory that delineates that learning occurs as interactions between individuals and feedbacks from social contextual factors (Bandura, 1986). It is the central piece that lays the shared foundation of the theme shared by the three studies in this dissertation: to investigate and evaluate possible ways to nurture students self-efficacious in STEM, particularly in the automation and AI branches of computer science, so that they could see through fabricated, negative stereotypes that otherwise would probably dissuade them from freely and persistently pursuing further studies and careers. As described by Bandura (1997), individuals' self-efficacy is developed primarily through making sense of information gathered from four major venues: (i) mastery experiences, (ii) vicarious experiences, (iii) social persuasion, and (iv) emotional and physiological states, as shown in Figure 1.

Figure 1

Four sources of self-efficacy

Mastery Experiences. In short, mastery experiences are experiences with success. Among the four sources of self-efficacy, mastery experiences is the most influential one because they provide the most direct evidence that one can master what it takes to succeed (Bandura, 1997). Mastery experiences are gained when one succeeds in accomplishing a task and attains stronger beliefs in their competence and undermined by failures, especially when they take place way before strong enough mastery experiences are formed.

Vicarious Experiences. Vicarious experiences are experiences of observing the success of role models. It is an important source of self-efficacy because seeing people whom they can emulate succeed strengthens the observers' beliefs that they possess the capabilities as well to successfully accomplish similar tasks (Bandura, 1977).

Social Persuasion. Social persuasion describes the reception of feedback from one's social connections, while attempting to accomplish certain tasks, that encourages them to believe that they can succeed. Receiving positive feedback from others enhances one's belief in their competence while negative feedback undermines it (Redmond, 2010).

Emotional and Physiological States. Emotional and physiological states of a person are the conditions of a person's emotional, physical, and psychological well-being. Such conditions, as well as how they are interpreted by an individual, are an important source of their self-efficacy (Bandura, 1982). It is a lot easier for one to boost their self-eefficacy when they are feeling well than when they are overwhelmed by anxiety, in which case even the strongest emotional and physiological pleasure can turn into fuels for self-doubts (Bandura, 1977). Practically speaking, this means that people who can better manage anxiety stand a higher chance of enduring challenging times and persevere.

Imaginal Experiences. More recently, another source of self-efficacy in addition to these four sources originally proposed by Bandura is suggested by Maddux and Kleiman (2016), namely imaginal experiences, which describes individuals' experiences of visualizing themselves performing well in given situations. The key to making use of this route to self-efficacy is to help individuals focus on picturing what success looks like to the extent that they see themselves as being nowhere else but the finish line (Maddux & Meier, 1995).

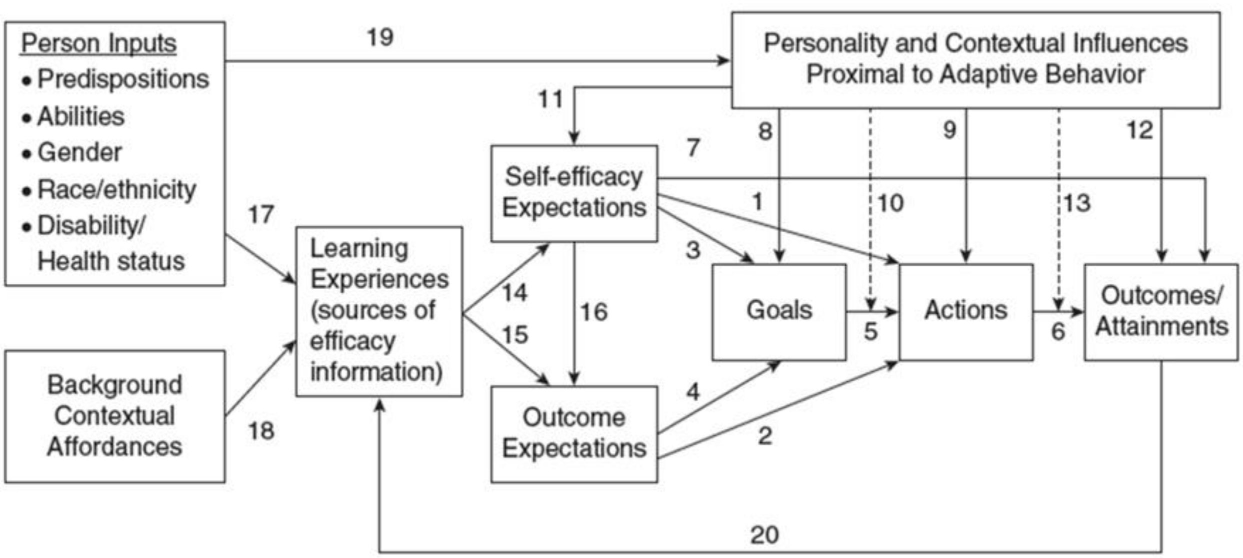
2.2 Social Cognitive Career Theory (Paper 2)

What SCT does not address very well, meanwhile, is how self-efficacy and other related factors in particular impact an individual's career choices. In order to answer Research Question 2, there is a more useful framework: Social Cognitive Career Theory (SCCT), a theory built upon SCT and self-efficacy that especially explains how self-efficacy and various factors interplay to shape the outcome of individuals' trajectories of career development, the focus of the second study.

Specifically, SCCT concerns with (i) how individuals' interests in specific academic and career pathways are formed, (ii) how individuals make choices and decisions about their academic and career pathways, and (iii) how individuals obtain success in their academic and career pathways (Lent, Brown, and Hackett, 1994). In short, SCCT holds that there are three interconnected variables that serve as the building blocks of a complex mechanism that addresses the three questions above, namely self-efficacy expectations, outcome expectations, and goals, as shown in Figure 2.

Figure 2

Mechanism of SCCT, from Lent, Brown, and Hackett (1994)



Self-Efficacy Expectations. In SCCT, self-efficacy expectations are individuals' beliefs about their abilities in performing certain tasks required in given occupations. These beliefs can dynamically shift depending on the specific domains of concern. For example, a person who has low self-efficacy expectations for their competence in sales can have very strong beliefs about their competence as an artist.

Outcome Expectations. Outcome expectations are beliefs about the outcomes of performing certain tasks required in given occupations. Such beliefs result from the picturing of the outcomes of engaging in the courses of actions involved. Careers that lead to positive outcomes like social approval, for example, correspond to higher outcome expectations and are more likely to be pursued.

Goals. Goals refer to individuals' intentions to pursue certain academic and career paths or reach certain achievements. Identifying goals enables individuals to make solid plans for the future and direct their behaviors purposefully. Goals are inevitably connected to the other two building blocks of SCCT. On the one hand, people tend to establish goals that align with their beliefs about their capabilities and what can come out of their endeavors of choice. On the other hand, successes and failures in reaching the established goals in turn can strengthen or weaken their self-efficacy beliefs and outcome expectations.

Based on these building blocks, there are three major models in SCCT that each focuses on one possible pattern of interaction among the central building blocks themselves and other input factors, including (i) the interest model, (ii) the choice model, and (iii) the performance model.

Interest Model. The interest model is the most basic model that posits that individuals are bound to develop their interests in certain endeavors when they feel self-efficacious and expect that positive outcomes will result from their courses of action. A variety of input variables contribute to this loop of growth in interest, the most direct one being sustained, authentic learning experiences with activities of interest: the more individuals are involved in doing something and reach achievements (Figure 2, arrow 14) and knowledgeable about the outcomes that await (Figure 2, arrow 15), the more they strengthen beliefs in themselves that they are capable of succeeding. Meanwhile, through actually engaging in the practices of the academic and career path of choice, they also develop a stronger sense of what to expect from their professions eventually (Figure 2, arrow 16).

Choice Model. The choice model builds upon the interest model and describes what happens next after individuals have become more self-efficacious and knowledgeable because of their direct learning experiences. Specifically, it posits that interests in careers are fostered when there are clear, achievable goals. When these the pursuit of these goals are in alignment with their self-efficacy and outcome beliefs, socially valued, and supported by significant others such as family members and friends, these goals guide individuals to draw on their self-efficacy and outcome expectations directly (Figure 2, arrows 1 and 2) and indirectly via assessing the specific goals devised based on self-efficacy and outcome expectations (Figure 2, arrows 3 and 4) to dive further into taking actions in the chosen path (Figure 2, arrow 5).

Performance Model. The performance model takes a step further from the previous two models, as it concerns with how individuals respond to the outcomes and attainments gained from the previous steps described. That is, the outcomes and attainments of individuals' actions taken in their chosen paths will largely determine their decisions to further pursue the path (more likely when the outcomes are positive) or not (more likely when the outcomes are negative). This culminating point of individuals' internalization processes in turn becomes the starting point of a new round of their learning experiences (Figure 2, arrow 20). In short, according to the performance model, in order that interests be fostered continuously, evidence of positive outcomes and attainments from career choices individuals is needed as a junction of iterations of self-sustaining interest development.

Contextual Factors. While the descriptions above is my best attempt at describing the SCCT model of development of interest, I have missed to discuss several important jigsaws that play extremely important roles behind the scene. One of them is person inputs, including one's gender, race, and health condition. The other is background contextual affordances, including barriers and supports one anticipates when they try to enter a profession of interest. As described by Lent and Brown (2002), these factors can be distal and take place way before than one makes their academic and career choices, which indirectly shapes their self-efficacy expectations through impacting their learning experiences (Figure 2, arrows 17 and 18), or proximal and take place soon before one makes their academic and career choices, which directly impacts their self-efficacy expectations (Figure 2, arrow 11).

2.3 Stereotype Threat (Paper 2)

What then are some examples of contextual factors that should be paid attention to and manipulated by computer science education researchers and practitioners? To better explore the second research question under the framework of SCCT in the second study, one important example of proximal contextual factors that has been widely discovered and discussed by previous studies is stereotype threat.

Stereotype threat is triggered when one senses a risk of confirming a negative stereotype about one's groups (Steele & Aronson, 1995), including but not limited to gender, race, ethnicity, and ability (Steele, 1997). The trigger of stereotype threats takes many highly contextual forms depending on individuals' perception of and interaction with survey items (Steele & Aronson, 1995) and instructional practices (Kreutzer & Boudreaux, 2012). Test-taking scenarios, especially when high stakes are involved, are where the most devastating effects of stereotype threats occur, as the victims are prone to falling into vicious cycles of low performance and self-doubt (Gordon, 2019).

When activated, stereotype threats seriously undermine individuals' academic and career attainments and outcomes in several ways. In the short term, stereotype threats cause individuals to be more self-conscious of their performance and the resulting outcomes, which disturbs their emotional states by making them feel more stressful and anxious to the extent that they would have to summon extra working memory to suppress such negative emotions, which inevitably leads to lower academic and career performance than usual (Schmader et al., 2008). In the long term, if one consistently encounters stereotype threats, they are likely to stop seeing themselves as associated with the group under threat and evade situations where they anticipate stereotype threats to be present (Aronson, Fried, & Good, 2002). For example, stereotype threats have been found to be a major reason for women's attrition from STEM disciplines, furthering achievement gaps between men and women (Stoet & Geary, 2012).

A number of ways have been experimented with to help reduce stereotype threats and their effects. Among them, a strategy that has been found to be useful in reducing the detrimental effect of stereotype threats is individuation, which entails distinguishing each individual from the at-risk groups that they may associate themselves with (Ambady, Paik, Steele, Owen-Smith, & Mitchell, 2004). A popular example of this strategy is values-affirmation tasks, a simple but effective intervention that proves to contribute to reducing and eliminating stereotype threats (Miyake et al., 2010). A precursor of these interventions was used by Cohen and colleagues (2006) in the form of a quick writing task that asked African American participants to write about their personal values they held that were characteristic of themselves and unrelated to the tested subject, which turned out to successfully eliminate stereotype threats among their research participants. Based on SCCT, this strategy works because it helps individuals affirm their person inputs, which directly (Figure 2, arrow 11) and indirectly (Figure 2, arrows 17 and 14) strengths their self-efficacy beliefs.

Another approach to reducing stereotype threat is through creating supportive learning and testing environments. For example, Chase and colleague (2009) found that when students' learning and testing experiences were highly interest-oriented and free of boundaries defined by learning materials, stereotype threats became virtually invisible. Similarly, Johns and colleagues (2005) successfully reduced stereotype threats by downplaying the stakes of testing and informing at-risk populations about the mechanism of stereotype threats. They observed that when they described a math test as a problem-solving test and talked about how women in STEM disciplines tend to suffer from stereotype threats before administering a test with difficult math problems, the performance between men and women was the same, whereas women performed worse than men when they described the test as a math test and did not instruct students about stereotype threat in advance. Based on SCT, these measures work because they effectively maintain or boost individuals' emotional and physiological states, thereby raising their self-efficacy.

2.4 Concept Inventory and Bloom's Taxonomy (Paper 3)

One particularly effective way to avoid causing respondents to suffer stereotype threat is to use concept inventories, a concise set of items designated to reveal misunderstandings about fundamental concepts of a discipline instead of directly measuring students' academic competence (Madsen, McKagan, & Sayre, 2017). These measurement tools, however, need to be carefully designed with comprehensive internal structures that are inclusive of various conceptual dimensions and cognitive levels. The third study is hence particularly focused on examining the layers of a novel concept inventory. To this end, it is necessary to review the origin of concept inventories and Bloom's Taxonomy, a popular way of structuring measurement tools in general.

The first well-known concept inventory was used by Hestenes, Wells, and Swackhamer (1992) and helped reveal that many students of theirs possessed fundamental misunderstandings about Newtonian force regardless of their academic attainments. Since then, concept inventories have been widely utilized in many STEM disciplines to help instructors better modify curricular and instructional designs (Taylor et al., 2014). Ironically, while the development and refinement of many of such concept inventories are enabled by AI, very few concept inventories for AI education has been created (Taylor et al., 2020).

A primary challenge faced by developers of concept inventories comes from the difficulty in determining a theoretically meaningful structure of various AI concepts. Traditionally, Bloom's Taxonomy has been a prominent choice by many to lay out the cognitive and affective levels of different items and map them. The original taxonomy (Bloom, 1956) incorporated six main categories of objectives of education, including:

(i) **knowledge**, which means the "recall of specifics and universals, the recall of methods and processes, or the recall of a pattern, structure, or setting";

(ii) **comprehension**, which means "a type of understanding or apprehension such that the individual knows what is being communicated and can make use of the material or idea";

(iii) **application**, which means the "use of abstractions in particular and concrete situations";

(iv) **analysis**, which means the "breakdown of a communication into its constituent elements";

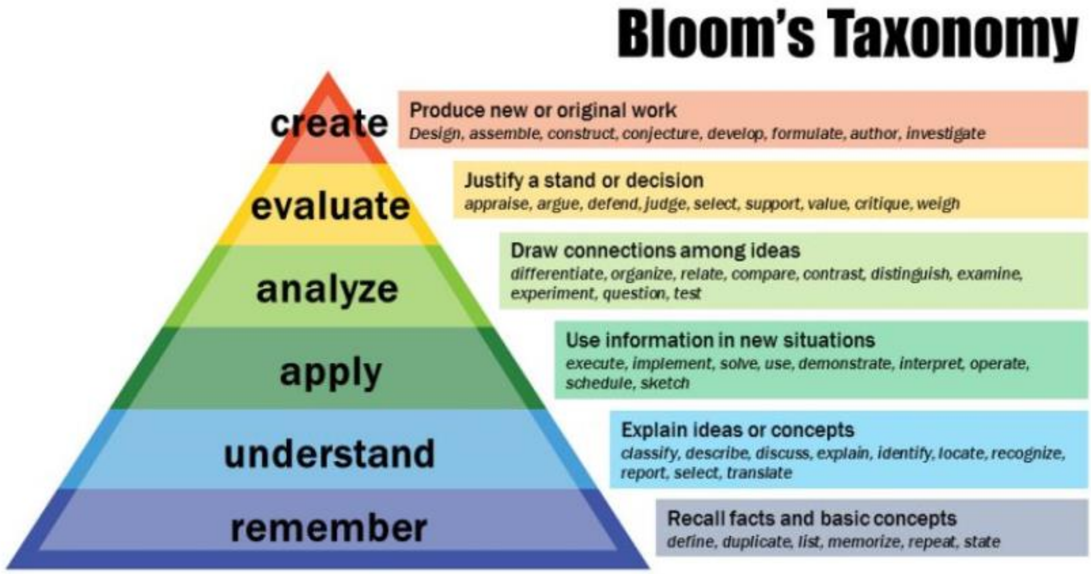
(v) **synthesis**, which means the "putting together of elements so as to form a whole"; and

(vi) **evaluation**, which means "judgments about the value of material and methods for given purposes" (Bloom, 1956, pp. 201-207).

Recognizing the static, limited definition of knowledge in the original taxonomy, the more recent version was modified, with the help of measurement and evaluation specialists, shifts into a more dynamic vision of cognitive processes of active engagement with knowledge with four major categories: remember, understand, apply, analyze, evaluate and create (Anderson, Krathwohl, & Bloom, 2001).

Figure 3

Bloom's Taxonomy hierarchies



As shown in Figure 3, the updated Bloom's Taxonomy uses verbs rather than nouns to describe a desirable sequence of propagating cognitive processes that educators and measurers could correspond their instruction and instrument designs to. Specifically, these processes include:

(i) **remember**, which entails actions such as memorizing, recalling learned concepts, and duplicating them;

(ii) **understand**, which entails actions such as summarizing and explaining ideas and concepts;

(iii) **apply**, which entails actions such as executing knowledge and implementing it in different scenarios;

(iv) **analyze**, which entails actions such as differentiating between and organizing complex ideas;

(v) **evaluate**, which entails actions such as evaluating, critiquing, and arguing for or against certain points of view;

(vi) **create**, which entails actions such as designing and generating original products.

The updated Bloom's Taxonomy (which will be referred to simply as Bloom's Taxonomy hereafter) has been widely popular among designers of a variety of education programs and measurement tools in many disciplines for many reasons (Wilson, 2004). First, since the categories are, by themselves, an ideal instructional sequence of cognitive processes and not tied to a particular construct, Bloom's Taxonomy is widely applicable to different contexts of research. Second, thanks to the clear behavioral objectives associated with cognitive and affective constructs alike, designers of assessments can structure their instruments in infinitely possible ways. One such possibility in the context of AI concept inventory is the core of Study 3.

3. Study 1: Smart Greenhouse for Future Presidents

Computer science, the "study of computers and algorithmic processes, including their principles, their hardware and software designs, their implementation, and their impact on society" (Tucker et al., 2003), has become increasingly important in our society. Educational researchers have worked extensively on researching the teaching and learning of computer science in K-12 settings and STEM major post-secondary classrooms (Lyon & J. Magana, 2020). However, relatively few studies have targeted college level courses prepared for non-STEM major students. Furthermore, given the challenges brought by the COVID-19 pandemic, additional research is much needed to explore new modes of and strategies for computer science education that take public health precautions into consideration.

3.1 Introduction

As such, to contribute to making computer science education more accessible to and safe for college students, this study focused on a "Science for Future Presidents" course taught to grades 13-16 students at a major research university in Northeast U.S. The focus of the course when the study was conducted was assembling and programming a miniature tabletop smart greenhouse. The course had previously been taking place in person before the pandemic featuring a range of STEM learning activities, but this was the first time for the smart greenhouse project to be a major component of the course, let alone being taught virtually, which made the teaching and learning experience a highly exploratory one. Given that none of the students enrolled in this course majored in STEM, the primary goal of the smart greenhouse project was, like many similar programs, helping them get interested in and feel less anxious about coding (Kaya et al., 2019; Umutlu, 2021).

Specifically, I focused on the following research questions in this research study:

1. How and in what ways did the smart greenhouse project influence students enrolled in the "Science for Future Presidents" course in terms of their interest, competence belief, and anxiety toward coding?

2. What reactions did students enrolled in the "Science for Future Presidents" course have regarding pedagogical designs and practices used in the smart greenhouse project?

3.2 Theoretical Framework

The theoretical framework I used that led us to asking the first research question was Bandura's self-efficacy theory, as the subconstructs I used, namely interest, competence belief, and anxiety, were adapted from the motivational, cognitive, and affective aspects of development in one's self efficacy (Bandura, 1993). Concrete development in these aspect of self-efficacy would serve as the foundational first steps that learners must take before they could further expand their knowledge and skills in computer science.

Meanwhile, the Social Infrastructure Framework (SIF; Bielaczyc, 2013) guided us to ask the second research question. Specifically, I was concerned with the dynamic interactions between students, the professor, teaching assistants, physical computing devices, and online learning materials in the "Science For Future Presidents" course, as informed by the 18 SIF design considerations. Specifically, I was particularly interested in the consideration of "student-teacher-cyberspace configurations", given that the course took place in a virtual space that consisted of fundamentally new elements of inter-personal and person-material interactions.

In addition, the design and implementation of the smart greenhouse project was inspired by previous educational studies on physical computing, which showed that abstract ideas in computer science can be more easily understood if they were manifested through observable physical phenomena (Przybylla & Romeike, 2018). This finding encouraged us to show students that the smart greenhouse could be programmed to conveniently turn on and off devices like fans and lights, depending on readings from various sensors of environmental variables, such as temperature, humidity, and light level.

3.3 Backgrounds

Cambria University (CU; pseudonym) is a private, selective, and pre-dominantly white research university in Northeast U.S. The "Science for Future Presidents" course is particularly designed for undergraduate students at CU who do not major in STEM in an effort to help them explore STEM topics that may potentially be of interest to them and learn STEM through practice-oriented project based learning experiences. During the time the study was conducted, 87 students were enrolled in the course. Eventually, 21 of them were able to complete both the pre-course and the post-course surveys.

Out of these 21 students, 16 were female and 5 were male. Among the 16 female students, 11 self-identified as non-Hispanic whites, 4 self-identified as Asian Americans, and 1 self-identified as Hispanic or Latinx. Among the 11 non-Hispanic white female students, 6 majored in Education, including 5 Sophomores and 1 Senior; 1 majored in Business and Finance and was a Junior; 1 majored in International Studies and was a First-year; 1 double majored in Political Science and Communication and was a Sophomore; 1 majored in Accounting and was a Junior; and 1 majored in Communication and was a Junior.

Among the 5 male students, 4 self-identified as non-Hispanic white, of which 2 majored in Computer Science (both Sophomores) and 2 double majored in Business and Finance (1 Sophomore and 1 Senior); and 1 self-identified as mixed (white/Asian) and double majored in Business and Finance.

STEM teaching and learning can be a tedious and unpleasant experience when learners do not engage in hands-on activities that allow them to apply what they learn and interact with the world (Conde et al., 2020; Jayathirtha et al., 2020; Miles, Huberman, & Saldaña, 2014). Given the fact that the course had to take place virtually and in an effort to maximize the quality of students' learning experience, kits of raw materials for the smart greenhouses were made and delivered to their residential halls. This allowed students to each watch instructions on assembling and coding the smart greenhouses and follow along at their own pace. In addition, two teaching assistants were available for students to reach out to in case they needed individualized in-person support.

The coding of the smart greenhouse is achieved via micro:bit, a block-based coding system that has gained popularity in recent years due to its beginner-friendly design and its applicability to various physical computing devices (Przybylla & Romeike, 2018). Eventually, students learned to code a variety of sensors and actuators to make the greenhouse automatically measure environmental variables and take actions to maintain these variables within an ideal range for plants' growth. Adopting the strategy of "decomposition of complex skills and tasks into minimal constituent components" (Reiser & Tabak, 2014, p. 47), learning activities (see Table 3.3.1) in the course were highly modularized in ways that focused on one sensor/actuator each time over the course of three weeks. Figure 4 shows a sample set of block codes students can possibly produce for a functioning smart greenhouse.

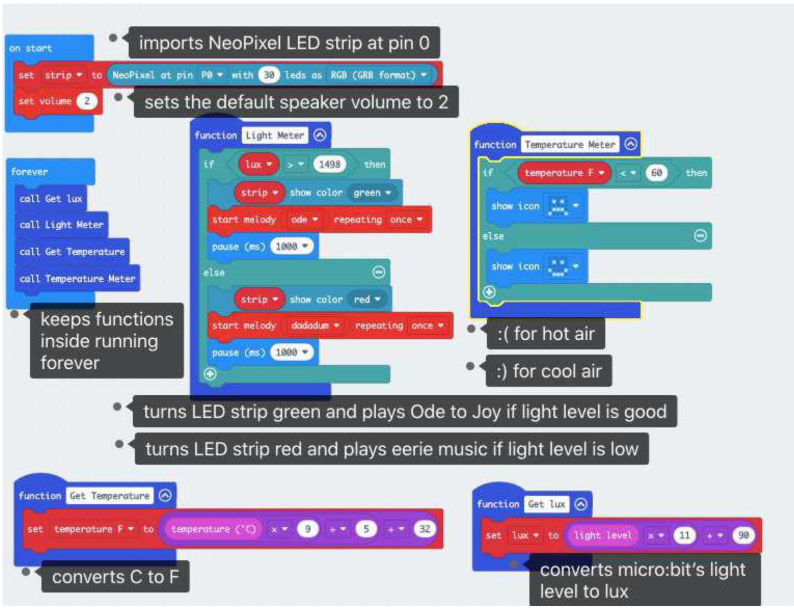
Table 3.3.1

Learning activities summary

Module	Topic(s)
1	Introduction to BBC Micro:bit & Grove shield hardware; introduction to MakeCode software; how to transfer files; use of micro:bit LEDs, Grove OLED screen, and Grove LED strip
2	Introduction to (programming) functions; using a Grove temperature-and-humidity sensor
3	Introduction to if-then(-else) loops; using a relay (switch) to activate or deactivate circulation or exhaust fans
4	Using the micro:bit's built-in light sensor to turn on or off an LED lamp; performing arithmetic in MakeCode
5	Calibrating and controlling a servo motor to open or close the greenhouse's windows
6	Communication between two or more micro:bits; using a Grove gesture sensor and/or passive infrared sensor to control various outputs
7	Demonstration of integration with Google Sheets; introduction to micro:bit Classroom (learning management software); planning future sessions; time for focus-group and survey

Figure 4

Final product micro:bit code for smart greenhouses



3.4 Methods

Using a convergent mixed-methods design, I collected and analyzed qualitative data generated from students' reflections written throughout the project and quantitative data collected from pre- and post-surveys, in hopes that comparisons of the two types of data could help us achieve deeper understandings of students' learning experience and better answer our research questions.

Specifically, the qualitative data I collected came from students' response to a series of weekly reflection questions that targeted the key constructs of this study, namely interest, competence belief, anxiety, and pedagogical designs and practices. Given the exploratory nature of this study, I chose to perform multiple iterations of Grounded Theory (Strauss & Corbin, 1990) procedures to obtain an inductive understanding of patterns in the qualitative data, including: (1) open coding, in which I thoroughly read through students' responses and labeled key moments that represented students' feedback with open codes, (2) axial coding, in which I looked through all open codes generated in the previous step and grouped similar open codes together to form more abstract axial codes, and (3) selective coding, in which I re-visited the raw data, examined the applicability of the newly generated axial codes, and further modified the axial codes as well as open codes accordingly.

Meanwhile, the quantitative analysis in this study was achieved via performing paired sample t-tests on the pre- and post-surveys containing Likert scale questions that targeted the aforementioned key constructs. More specifically, items in all aforementioned subscales with the exception of “suggestions for pedagogical designs and practices” were adapted from the “modified Attitudes Toward Science Inventory” instrument (mATSI; Weinburgh & Steele, 2020). This modified version of mATSI has been validated in previous studies of similar contexts (Jackson, Cheng, Meng, & Xu, 2022). The major difference between it and the original version was that it replaced the word “science” with the word “coding” in multiple items to better fit the physical computing centered nature of the smart greenhouse project.

3.5 Results: Statistics and Primary Codes

Eventually, quantitative analysis shows that participants displayed statistically significant positive shifts in the subscales of Interest in Coding Jobs ($p=.04$, Cohen’s $d=.34$), Interest in Learning about Coding ($p=.02$, Cohen’s $d=.34$), and Competence Beliefs in Coding ($p=.0002$, Cohen’s $d=.59$) and statistically insignificant shifts in Anxiety ($p=.7$, Cohen’s $d=.07$) and Connection between Coding and Science ($p=.5$, Cohen’s $d=.7$). These are surprisingly promising signs that the Smart Greenhouse project was contributing to boosting participants’ self-efficacy in coding.

There are but two major limitations that have to be noted: (1) the sample size (n=21) was quite small after all, and (2) the 21 participants self-selected themselves to participate in this study. As such, it is hard to say to what extent stories of the 21 participants were representative of the experience of the whole class, or whether it was those who enjoyed coding more than others in the class that decided to complete the weekly reflections and respond to both the pre- and post-project surveys.

Bearing these limitations in mind, I attempted to uncover, to the greatest extent possible given the available data, different stories of participants exhibiting different shifts in their perceptions toward coding. To this end, I ordered the 21 participants based on their overall shifts in self-efficacy scores (all dimensions combined) from low to high, read through their weekly reflections case by case, and coded their reflections. In this process, I first used the dimensions of self-efficacy as primary codes to highlight segments in participants' reflections that reflected changes in their self-efficacy. After that, I generated secondary codes based on themes and patterns that repeatedly appeared across participants' reflections and went back to code each reflection with the new generated secondary codes. In the next few subsections, I will present in detail the cases of 7 participants selected based on their rank in net shift in self-efficacy score (3 top highest, 3 top lowest, 1 zero net shift) and discuss the secondary codes generated in the Discussion section.

3.51 CR1

CR1 displayed the highest increase in overall self-efficacy score (from 3.15 to 4.31, d=1.16) among the 21 students. She entered the project with anxiety and little prior knowledge about coding but eventually gained visibly stronger competence belief and interest in coding.

In week 2, CR1 already possessed more competence belief in coding after attending one session. She wrote in her first reflection that “I had no coding knowledge whatsoever when I first came into class, and even though I’ve only done one lab that involves coding, I still think my knowledge of coding and its advantages has grown tremendously”, indicating a clear boost in her competence beliefs. In addition, she wrote that “I feel a greater confidence in my abilities to learn new skills like coding and not shy away from challenges, and I no longer see it as something unrelated to my future career field”, which, in addition to providing evidence of growth in competence belief, shows that she developed greater interests in coding related jobs. In addition, CR1 also talked about the alleviation of her anxiety. She acknowledged that “I was quite apprehensive about the coding coming into this lab” but noted that at the moment “I think I have a better grasp”.

In the joint reflection for week 3 and 4, CR1 focused on writing about how visible impacts of several features of the smart greenhouse made her become more interested in coding. For example, she wrote that “manipulation of variables and monitoring their effect on my microgreens taught me a lot of patience and changed my perspective on the level of interaction you can have with plants”. In addition, she mentioned that coding the LED strips was “by far my favorite concept that I learned how to code”. Also, similar to before, she expressed her competence belief in coding (“I am not scared to learn about coding”) and her interest in coding related jobs (“I no longer see it as unrelated to my future career field”). This increase in competence belief was also evidenced by her willingness to share her coding experience with others. She mentioned that she showed the LED strips she coded to her roommate and that both of them were impressed when they saw the right code was “making the LED strip turn rainbow”

3.52 NM2

NM2 displayed the second highest positive shift in overall self-efficacy score (from 2.15 to 2.84, d=.69). She started the project with very little confidence in her ability to code and finished with the belief that she could definitely succeed in beginner level coding.

In her week 2 reflection, she demonstrated considerable growth in competence belief in coding. She clearly knew that “I am building a mini greenhouse and using coding in order to control different factors and collect data”. When asked about her learning experience in the first two weeks in general, she wrote that “I definitely believe that my skills to program a micro:bit has gotten significantly better since the lab has started”.

Later in her weeks 3 and 4 reflection, NM2 described again the growth in her competence belief. When asked about the impact of the project on her competence belief, she responded “I would have considered myself at 0 before and at 5 now”. This time, however, she also expressed her awareness that the project was but a beginning step to coding. She commented that “I think this lab was a good introduction to see what it is like to code” but “there is still so much about coding that I do not know about” and that “I believe this lab was just a quick preview into the coding world”.

3.53 CP3

CP3 displayed the third highest increase in self-efficacy score (from 3.62 to 4.28, d=.66). Similar to many, at the beginning CP3 did not know much about coding and had a lot of anxiety, but she managed to pick up a lot of competence belief as she managed to successfully follow along the instructions and create her fully functioning smart greenhouse.

While reviewing her experience with the lab so far in her week 2 reflection, CP3 noted that “my confidence in my capacity/skills to program a micro:bit has definitely skyrocketed after starting this lab”. She remembered being “a bit scared to code because I thought it would have been a difficult process to do”.

Reading through CP3’s weeks 3 and 4 reflection, it becomes clearer how the smart greenhouse project helped CP3 develop stronger self-efficacy in coding. Specifically, she talked about how micro:bit as a block based coding language was a convenient tool to utilize. She noted that “using the drop down arrows to set certain factors of the code was much easier than doing it on other platforms” because “for instance, Java is very particular about the syntax of your code and will fail if just one comma is off, so the drop down options helped streamline this process”.

In addition, she particularly mentioned that she benefited from watching the step-by-step instructional video recorded by the instructor of the course. She recalled being “unsure of what the actual coding meant and how it functioned”, but fortunately “the instructions provided by the YouTube videos for class and instructions manual provided by Professor B instilled confidence in me” and “led to the increase in my comfort levels”.

Also, similar to the case of CR1, this rise in her self-efficacy was reflected by her willingness to share her coding experience with others: “when talking with suitemates, it became the main topic of conversation that I would bring up”.

3.54 SL21

SL21 displayed the largest drop in overall self-efficacy score among the 21 participants (from 3.28 to 2.44, d=-.84). Her reflections, however, reflected some quite positive changes. Similar to CR1, SL21 did not have much great experience with coding prior to the project (“I was hesitant to learn about coding before because I had very little exposure to coding and the exposure I did have was not very successful”) but ended up believing that coding could be easier than she thought to be.

SL21 had a less than ideal experience with the smart greenhouse project from the beginning. In week 2, after being introduced to micro:bit for one week, she was conservative but clear in describing her engagement in the smart greenhouse project. She wrote in her reflection plainly that “I am programming a micro:bit to collect observable data in order to solve a research question based on our plants” after recognizing that “I am still not completely comfortable with it”.

In the reflection for weeks 3 to 4, however, there were several positive changes. First, SL21 wrote that “I realized that coding is really not as complicated as it seems” because of the smart greenhouse project. Even though her coding experience was not a smooth one and that she did not get to program a completely functioning smart greenhouse (“I was able to do only one experiment successfully”), she was confident that she would be able to do better “if we were able to have classes in-person and I could actively ask questions”.

3.55 MM20

MM20 displayed the second highest drop in overall self-efficacy score (from 2.82 to 2.44, d=-.38) and tied with SL21 for the lowest eventual overall self-efficacy score (2.44). Interestingly, MM20's reflection also indicated that she did benefit from the smart greenhouse project to some extent and became more self-efficacious in coding than before.

It was clear that MM20 has a very rough, stressful start unseen in most of her classmates in week 1, but it was already getting better when she finished week 2. In her week 2 reflection, MM20 wrote that in the first week "I did not feel at all comfortable with programming the micro:bit, I was very nervous", but in week 2 "I feel more comfortable about the actual coding part, but am still a bit unsure as to correct coding mistakes or how to transfer all of this information to the micro:bit and allowing it to run without errors".

In her weeks 3 and 4 reflection, coding clearly became a more encouraging process for MM20 as her concerns about making and correcting errors were clearly getting addressed. She noted that "the videos in this lab that walked me through how to create a code really helped". Her experience became a successful one to the extent that "I could even see myself incorporating this (the project) into STEM lessons as a teacher in the future" and that "I would now consider myself at 5 (comfort level, full score=10)".

That said, similar to NM2, MM20 described her success as but a beginning point. She was certainly proud that "even with the roadblocks I experienced in this lab, I developed some skills in how to fix problems myself" but also expressed that "I think I still have a very long way to go".

3.56 AN19

AN19 displayed the third largest drop in self-efficacy score (from 3.02 to 2.84, d=-.28). Similar to SL21 and MM20, however, AN19's reflection reflected a relatively successful story of how she gradually became more self-efficacious in coding.

In week 2, for example, AN19 did reveal that she was "a little nervous about what it (the project) entailed". She also believed that "if I didn't have tutorial videos step by step and detailed explanations on how to program the micro:bit", she would not "have sufficient skills or background knowledge to program it on my own". However, being able to follow along instructional videos itself was a sign of learning.

In her weeks 3 and 4 reflection, it became clearer that AN19 had become more self-efficacious. She felt that "I am better prepared to learn more about coding compared to before" and that she would "rate myself at about 7" and indicated that because of her increased confidence, she "would be willing to take on a new software or program to play with and figure out". AN19 also spoke highly of the YouTube instructional videos and the personal support she received, as she found that "YouTube playlist in order of what to do, and the TAs and professor were extremely attentive and helpful and answered me very quickly whenever I had a question".

3.57 HL0

HL0 displayed an exact net zero shift in her overall self-efficacy score (from 3.59 to 3.59), a unique result unseen in others. Reading through her reflections, it appears that the smart greenhouse project did help her become more comfortable with coding to some extent, despite challenges she encountered in her virtual learning experience.

This positive impact is most salient in her week 3 and 4 reflection, in which HL0 revealed that while she “did have some background knowledge about how coding works because I did learn python in a computer science class”, it was “a whole new experience to see coding in action physically rather than just some results on a computer”. She further explained that that seeing the visible impact of her micro:bit code on the physical world made her “better prepared to learn more about coding now” because physical computing with micro:bit “made it much easier to visualize what was happening; it was color coded and I could see what I coded happening in real life”. She concluded that “prior to the lab, I would rate my comfort level at a 3, but after the lab, I would rate my comfort level at a 5”.

3.6 Discussion: Secondary Codes

Reading through the reflections of the participants, there are several salient patterns of learning trajectories that could be found across different individuals.

3.61 **Temporary Rough Start**

I use the secondary code Temporary Rough Start to describe a common situation that many participants indicated that they had encountered. Specifically, many reported that they were feeling nervous about what to do when they entered the project largely because of their lack of exposure to and/or successful experiences with coding prior to the project. In as early as week 2 of the project, however, most of them turned out to be much more self-efficacious in coding.

This pattern could most clearly be seen in the case of CR1, who indicated in her week 2 reflection that she “was quite apprehensive about the coding coming into this lab” in week 1 and that at the moment of writing the week 2 reflection she thought “I have a better grasp". Similarly, in the case of NM2, the change she experienced between the start of the project and week 2 was so significant that she "would have considered myself at 0 before and at 5 now". CP3's experience was also almost the same, as she went from being "a bit scared to code because I thought it would have been a difficult process to do" in week 1 to "my confidence in my capacity/skills to program a micro:bit has definitely skyrocketed after starting this lab". This was also the case for MM20, who described herself as "I did not feel at all comfortable with programming the micro:bit" at first and "I feel comfortable about the coding part" except for feeling "still a bit unsure as to correct coding mistakes" in week 2.

Responses like these suggest that the smart greenhouse project was effectively and quickly helping participants become more self-efficacious in coding. This is a promising sign that implies that the smart greenhouse project was, using the terms of SCT, bringing them a successful mastery experience of coding, something that most of them indicated that they lacked and truly needed for the sake of developing their competence belief in coding.

***3.62* Visible Impact**

I use the secondary code Visible Impact to describe cases in which participants were obviously benefiting from seeing visible impacts of their codes on the physical world in terms of obtaining, again in the terms of SCT, successful mastery experiences of coding and hence developing stronger competence belief in coding.

CR1 was a perfect example of this pattern, as she described coding the LED strips as "by far my favorite concept that I learned how to code" and mentioned excitedly sharing the process of "making the LED strip turn rainbow" with her roommate and thinking she should no longer consider coding as unrelated to her future career, indicating clear a rise in self-efficacy in coding. This was also the case for HL0, who felt she was "better prepared to learn more about coding now" because physical computing with micro:bit "made it much easier to visualize what was happening; it was color coded and I could see what I coded happening in real life", a benefit of micro:bit coding that, as she rightfully pointed out, could not be enjoyed while engaging in line based coding languages such as Python.

3.63 Bane of Error

Bane of error is the secondary code I generated to another commonly seen pattern in participants' reflection: sensitivity to errors in coding. This pattern manifested differently for different participants. For participants who were able to correct errors they made with support from their instructor and Teaching Assistants, the successful experience of overcoming mistakes consolidated their confidence belief in coding. Otherwise, for participants who did not get a chance to correct all errors in their codes, they tended to be reserved in estimating their competence in coding and/or seeking more support.

For example, CP3 felt that it was a huge advantage of micro:bit that she could use "drop down arrows to set certain factors of the code" and criticized Java for being "very particular about the syntax of your code and will fail if just one comma is off". She also highly appreciated that "the instructions provided by the YouTube videos for class and instructions manual provided by Professor B instilled confidence in me" because she was definitely "unsure of what the actual coding meant and how it functioned" prior to accessing the step-by-step guidance in the videos and the manual. Similarly, MM20 indicated that she was still “a bit unsure as to correct coding mistakes or how to transfer all of this information to the micro:bit and allowing it to run without errors” and was thankful that “the videos in this lab that walked me through how to create a code really helped”.

In contrast, SL21 lamented that she was “able to do only one experiment successfully” and that she would have benefited from the project more “if we were able to have classes in-person and I could actively ask questions”, revealing her demand for more instructional support that was not available in the virtual learning environment that the course had to take place in.

3.64 Beyond the Classroom

Beyond the Classroom is the secondary code I generated to describe another common pattern seen in participants who had gained increased competence belief in coding: thinking and talking about possibilities of using coding in settings other than their virtual learning environments. In the words of SCCT, this indicated that these participants were actively adjusting their outcome expectations for coding to include more possibilities, which was a contributing factor to continued growth in self-efficacy.

For example, CR1 knew that the fact that she “no longer see it as something unrelated to my future career field” marked her “greater confidence in my abilities to learn new skills like coding and not shy away from challenges”. Similarly, MM20 chose to express her increase in competence belief in coding by saying that “I could even see myself incorporating this (the project) into STEM lessons as a teacher in the future”.

3.7 Concluding Thoughts

In this study I examined the impact of an undergraduate course designated for non-STEM major students that centered on the programming and assembling of a novel miniature smart greenhouse. Findings from analyzing qualitative and quantitative data suggested that the smart greenhouse project did contribute to boosting students' self-efficacy in coding through supporting them in a step-by-step way to successfully code with a beginner friendly language that produced visible impacts on the physical world. This study is hopefully going to serve as a useful reference point for researchers and practitioners as they work toward making STEM education more encouraging for students who did not have much prior knowledge about coding and inspire them to see themselves as competent in coding and explore possibilities of connecting coding with their life.

4. Study 2: Review and Evaluation of the Teaching of an AI Career Curriculum

4.1 Introduction

Artificial Intelligence (AI) systems are intelligent machines constructed by humans to perform intelligent tasks (AAAI, 2020). Once an abstract concept in science fictions, AI is rapidly advancing and ubiquitously reshaping society in many ways. Nevertheless, the general public is largely underprepared for understanding the mechanism of AI algorithms and learning about the ethical and social justice related implications of AI, which greatly limits their ability to maximally enjoy the benefits brought by AI and safeguard their well-being from harmful, biased designs and uses of AI (Long & Magerko, 2020). As such, youth in this eve of the era of AI, middle school aged youth in particular (Manyika et al., 2017), urgently need to be equipped with sufficient knowledge and skills to creatively and critically make use of AI artifacts in order to survive and thrive in the AI-enabled future (DiPaola et al., 2020).

Among a plethora of AI literacy centered lessons and learning platforms developed in recent years, the DAILy (Developing Artificial Intelligence Literacy) curriculum is a promising example for educators to experiment with in their classrooms for many reasons. First, empirical studies have shown the effectiveness of DAILy in raising youth's AI literacy within a short period of time, as indicated by positive shifts in participating students' ability to both describe the mechanism of AI systems and the socio-ethical implications of AI after experiencing an online summer workshop style implementation (Zhang et al., 2022). Second, DAILy provides educators with a lot of flexibility in terms of the platform (i.e., in-person or online) and the length (i.e., a 2 to 3 week workshop or a full course over a semester). Third, the cost of implementing DAILy is low, given that non-commercial use of it is free and the majority of basic instruction materials, including lesson plans, slides, programs for plugged and unplugged hands-on activities, and more, can be found on its official website (DAILy Workshop, 2022).

4.2 Purposes

There are two things that the website does not have yet. First, the website has no listed widely accepted K-12 standards, such as CSTA (Computer Science Teacher Association) standards, that the lessons address, the addition of which would greatly facilitate teachers in more effectively navigating through the DAILy curriculum and decide about where each lesson fits their teaching plan. Second, the website has not documented up-to-date teaching tips that inform teachers about teaching and learning related challenges that have occurred (and likely will occur again in future implementations), and potential viable solutions.

As a graduate research assistant who has observed multiple implementations of DAILy, I am therefore writing this practitioner study to complement the DAILy website and help teachers interested in teaching DAILy better utilize the resources that the DAILy website has to offer, especially for those whom the development team of DAILy has not had the chance to reached out to and provide DAILy related professional development training (Lee et al., 2021). To this end, in this practitioner study I will try to answer the following questions that may probably be of interest to AI educators:

(1) What CSTA standards do lessons in DAILy address, and in what ways?

(2) For each lesson in DAILy, what are some obstacles that could take place during the process of teaching and learning, and what are some actions that could be taken accordingly, based on observations of previous implementations of DAILy in general?

4.3 Modules and Lessons in DAILy

DAILy curriculum contains a series of lessons on AI concepts and AI ethics that can be taught both in-person and online that takes 30 hours or more to complete. Designed by a team of experienced AI educators and researchers and guided by the goal of making AI education more accessible for youth without much prior knowledge about AI, lessons in DAILy situate complex AI concepts in everyday life contexts and reveal the relevance of AI to social-ethical issues that are directly relevant to students and their communities (Zhang et al., 2022). After careful revision of the outcomes of pilot testing of the prototype (Ali et al., 2019), the development team included in DAILy a hierarchy of five modules, including:

(1) Introduction to AI, which covers topics such as what is and is not AI;

(2) Logic Systems, which covers mainly the mechanism of decision trees that classify complex data into desirable groups;

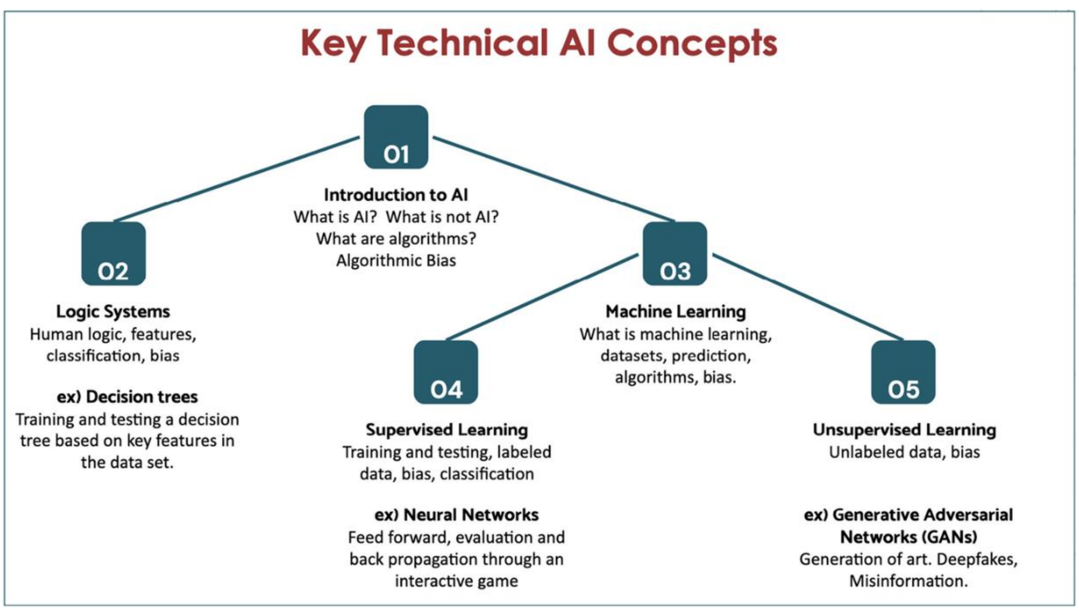
(3) Machine Learning, which covers what machine learning is in general and the difference between supervised and unsupervised machine learning;

(4) Supervised Learning, which is a type of Machine Learning and covers topics such as the training and testing of neural networks; and

(5) Unsupervised Learning, which is another type of Machine Learning and covers topics such as how Generative Adversarial Networks (GANs) generate art from unlabeled data.

Figure 4.3.1

Hierarchy of modules of AI concepts in DAILy



DAILy lessons built on AI lessons created by researchers in the Personal Robot Group at the MIT Media Lab and MIT STEP (Scheller Teacher Education Program) Lab that aimed at providing youth with accessible, hands-on learning opportunities that would inspire their consideration of social-ethical issues related to AI. For example, The module Unsupervised Learning (especially the lessons What are Deepfakes and Spread of Misinformation), draws from the "Creative AI" curriculum (Ali et al., 2021; DiPaola et al., 2021) that guided youth to explore how GANs can be used to generate videos, pictures, and texts, as well as social-ethical consequences of AI-generated and AI facilitated transfer of media. The unit Introduction to AI carries over the essence of the "How To Train Your Robot" curriculum (Williams et al., 2021), namely interactive exploration of algorithmic bias in AI and ways to mitigate such bias (especially the lessons Ethical Matrix, Investigating Bias, and Redesign YouTube), through exposing students to how supervised machine learning models are inevitably affected by their designers' interests, encouraging them to consider ways in which these models may be subject to bias and violate ethical principles, and challenging them to reimagine and redesign AI to more fairly serve the interests of different stakeholders.

4.4 Analysis of Sample Lessons

In this section, I will focus on a selection of lessons that are both representative of the five modules and the focus of DAILy on AI concepts and AI ethics. Specifically, as mentioned in Section 4.2, I will analyze for each lesson (1) what CSTA standards are addressed and in what ways and (2) potential challenges in teaching and learning and possible solutions. Before I start, here is a table that summarizes the outcome of this process for a quick preview.

Name of Lesson	CSTA Standard	Module
What is AI	3B-AP-08	Module 1 Introduction to AI
Algorithm as Opinions	3A-IC-24	Module 1 Introduction to AI
Decision Trees	2-DA-07	Module 2 Logic Systems
Investigating Bias	2-IC-21	Module 3 Machine Learning
Introduction to Supervised Machine Learning	2-AP-16	Module 4 Supervised Learning
Neural Networks	3B-AP-09	Module 4 Supervised Learning

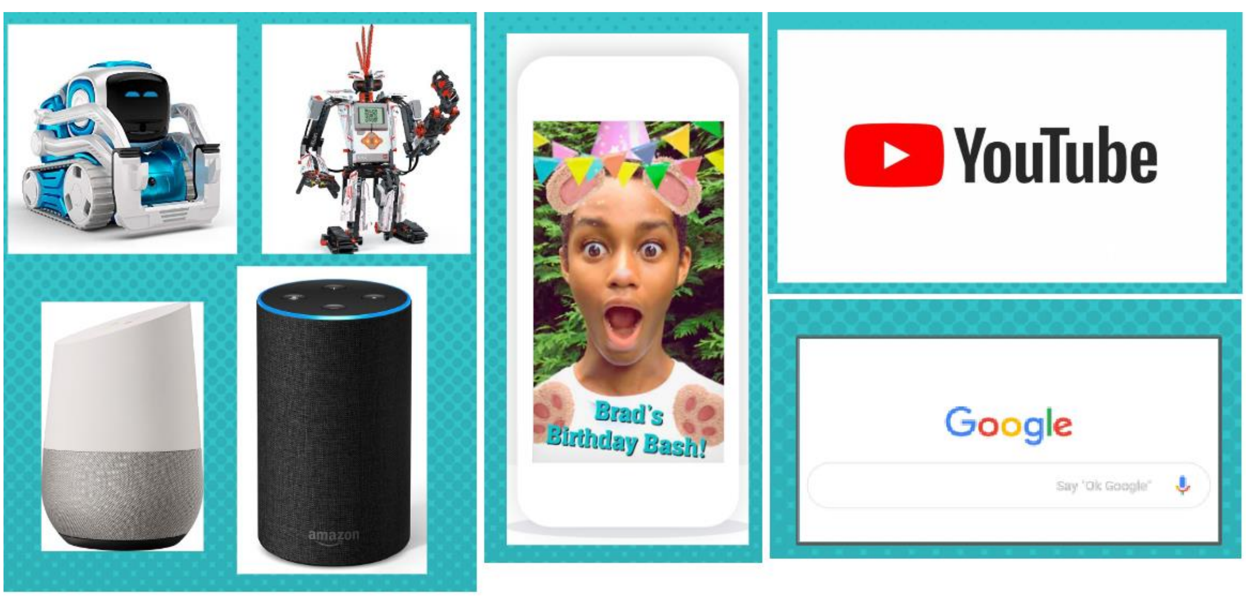
What are Deepfakes	3B-IC-25	Module 5 Unsupervised Learning

4.41 What is AI (3B-AP-08)

The first lesson of the DAILy curriculum, "What is AI", engages students in thinking about what technologies around them are and are not examples of AI and for what reasons. This lesson belongs to Module 1 (Introduction to AI) and aligns well with CSTA standard 3B-AP-08 of "describe how artificial intelligence drives many software and physical systems". The way this standard is addressed in this lesson is through asking students what "Artificial Intelligence" make them think of, giving them examples of software and physical systems they see and use in everyday life, and discussing with them why or why not these examples are or are not AI-driven.

Figure 4.41.1

Examples of AI and non-AI



One thing that is worth keeping in mind is that at this stage of the curriculum, students do not need to know how to distinguish between AI and non-AI properly yet. The essence of this lesson is to encourage students to draw on their prior knowledge, be it correct or incorrect about AI, through justifying their responses. It would be tempting to tell students, for example, that a robot is not quite an example of AI after they respond that robot is the first thing they think about when they hear the word AI. However, a better practice would be to ask students to first further explain why they thought of robot. Doing so would not only help them more clearly review their preconceptions about AI, but also prepares them for learning about "learning from data" as a key feature of AI systems later.

4.42 Algorithms as Opinions (3A-IC-24)

The second lesson of the DAILy curriculum, "Algorithms as Opinions", introduce students to the fact that algorithms as a key component of AI are rarely neutral and objective and many of them would think and deeply impacted by the opinions of their designers, knowingly or unknowingly. This lesson belongs to Module 1 (Introduction to AI) and addresses directly the CSTA standard 3A-IC-24 of "evaluate the ways computing impacts personal, ethical, social, economic, and cultural practices", as youth will learn that AI can have positive and negative impacts on society and that algorithmic bias can result in benefiting (or harming) certain demographic groups more than others.

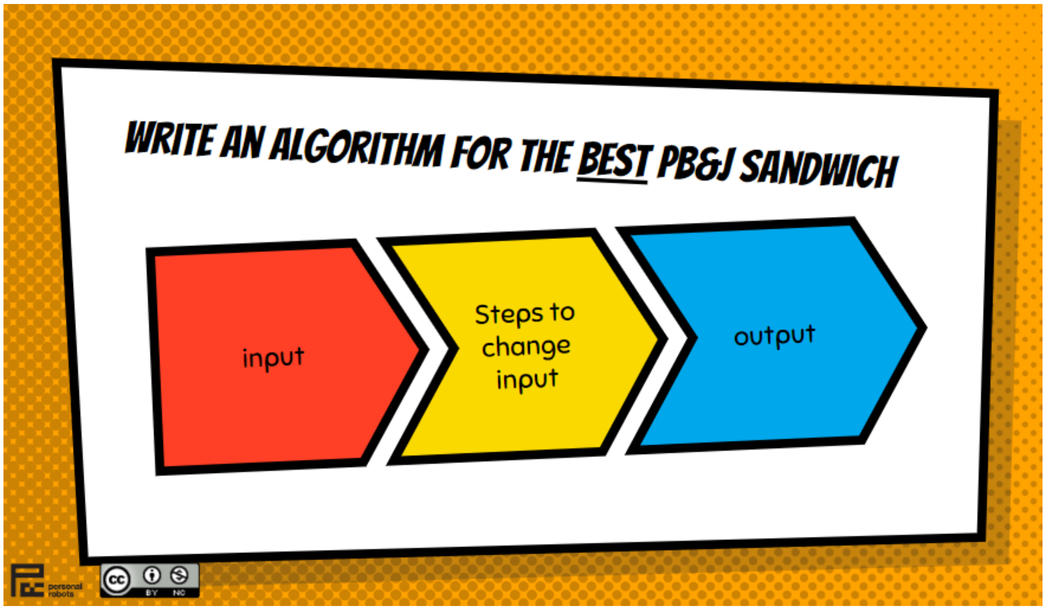
In order to embed the concept of algorithmic bias into a daily context that students can understand, the lesson uses food recipes, which also involves the set-up of rigorous procedures to convert certain input into favorable output, as an approximation. Specifically, students are challenged to describe their recipes for the "best" PB&J (peanut butter and jelly) sandwich and explain what makes their PB&J sandwich the best. Eventually, after small group and whole class discussions, students will learn that the meaning of "best" can be entirely different stakeholders because of differences in what they prioritize (i.e., taste from the perspective of students and nutrition from the perspective of their parents).

In teaching this lesson, it is important to not drift away from the purpose of helping students think of algorithms as opinions that can be biased. To this end, it's important to help them notice how "best" could mean different things to different people and that there is not a single correct answer. Teachers should therefore encourage students to (a) be clear about the steps in their recipe and (b) explain well why they believe the output of their recipe is going to be the best, (c) consider how others may think of "best" in different ways.

In addition, observations of previous implementations show that younger learners (around Grades 6 to 7) may find it difficult to think from the perspectives of others. When teaching these students, teachers will need to make the best use of whole class discussions to direct students' attention to how different individuals think differently of what it means to the best PB&J sandwich.

Figure 4.42.1

Best PB&J sandwich



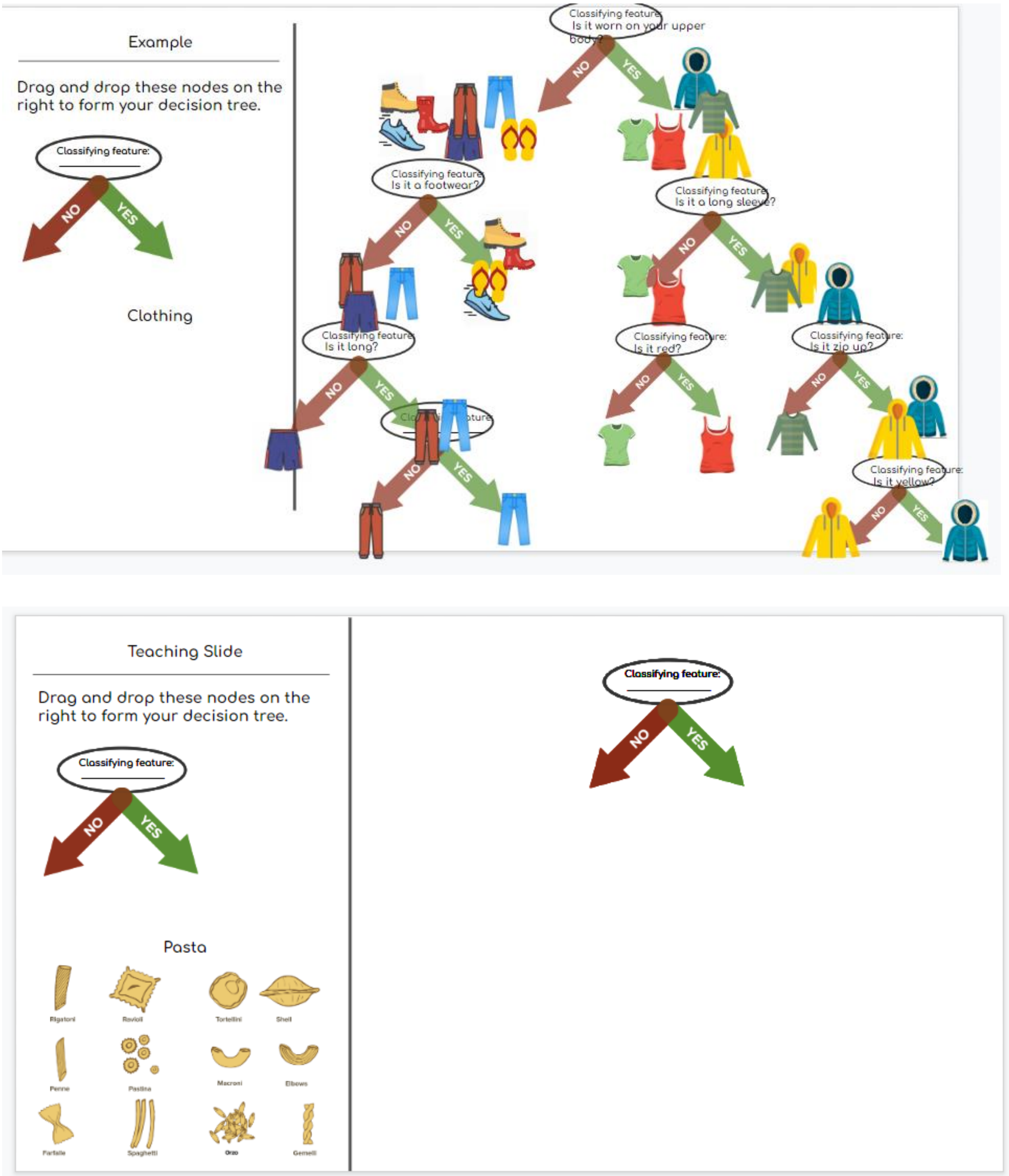
4.43 Decision Trees (2-DA-07)

The third lesson of the DAILy curriculum, “Decision Trees”, introduces students to decision trees, a basic algorithm that can be used in AI systems to classify complex data into meaningful groups. This lesson belongs to Module 2 (Logic Systems) and addresses the CSTA standard 2-DA-07 of “represent data using multiple encoding schemes”, as students will learn how to use decision trees to represent and categorize complex data into meaningful groups in an organized format.

Specifically, in the learning activity “Is it winter”, students learn to build decision trees that can classify a variety of clothes into groups based on the specific weather conditions the clothes are suitable for. In the learning activity “Queen of Pastaland”, students are challenged to make decision trees that would help the Queen of Pastaland, a cat that can only say yes and no, effectively tell her cook which kind of pasta she would like for dinner. In this process, abstract concepts in decision trees like nodes and branches are concretely context-based, whether it’s picking the right clothes for the right weather or finding the right pasta for the Queen of Pastaland.

Figure 4.43.1

Decision trees for Is it winter and Queen of Pastaland



The major challenge in teaching this activity is to help students realize that AIs think in binary ways, which limits the design of classifying questions. For example, it has been common to see students design classifying questions like "is the pasta long" for the Queen of Pastaland activity. In cases like this, it is important for teachers to use guiding questions to help students reconsider whether their classifying question is something that machines can understand and work with them to form more AI-friendly questions.

Another challenge that students are likely to encounter is lack of understanding of the purpose of decision trees. For example, in the Queen of Pastaland activity, while the final decision tree should only have one kind of pasta left at each node, it's very likely that students will stop half-way without realizing that their decision tree is incomplete. In such moments, students would benefit from reminders of what their end goal is: to design a decision tree in which each pasta can be uniquely identified after a chain of simple yes or no questions.

Furthermore, there are occasions in which these activities are not accessible to students. For example, in regions where winter is as warm as the rest of the year, "Is it winter" should be modified to reflect what students typically wear in a particular time of the year. For students who are completely unfamiliar with pastas, it would be a good practice to replace pastas with other types of food that they see on their tables, in order to make the activity more accessible for them.

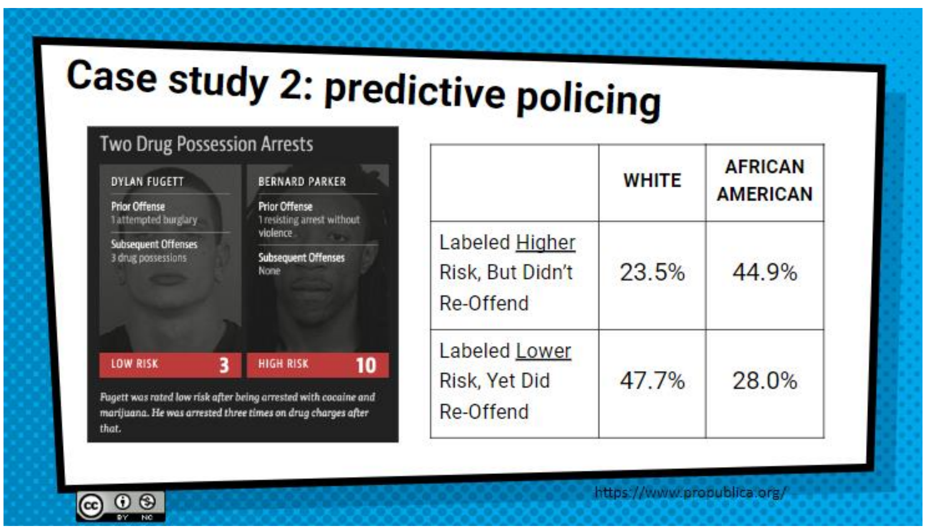
4.44 Investigating Bias (2-IC-21)

The fourth lesson of the DAILy curriculum, "Investigating Bias", students will be introduced to more real-life examples of biases in AI and consider the consequences. This lesson belongs to Module 3 (Machine Learning) and addresses the CSTA standard 2-IC-

21 of “discuss issues of bias and accessibility in the design of existing technologies”, as students will engaged in in-depth discussion of the meaning of fairness and bias in AI and how AI systems designed with positive intentions can lead to negative consequences due to biases in their design or training procedure. For instance, students will see that a predictive policing system that was actually developed and used in B County in F state to assess the risk of re-offending and determine qualification for a parole, because of using zip code as a key factor (a stand-in variable for race in that region) in its algorithm, tended to make decision unfairly in favor of Whites.

Figure 4.44.1

Biased predictive policing system due to unfair training



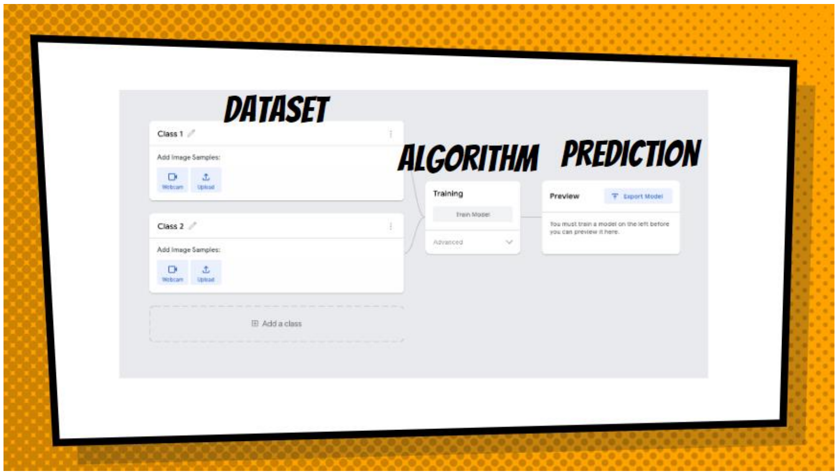
When viewing this example, a typical reaction students may have, largely due to lack of full understanding of the biased, is questioning with confusion how AI can be racist, because in their mind, being racist is something that only humans do and that AI has to be neutral and objective. In situations like this, it is very important that teachers clarify that while it may look like the predictive policing AI in this example is being racist in terms of its effect, the AI is still simply following an algorithm, albeit a biased one. In addition, previous implementations have also shown that students may be saddened by the policing system's unfair treatment of people of color. In such occasions, is important that teachers encourage students to look at the bright side and discuss with them what they can do to retrain the system and make it more fair.

4.45 Introduction to Supervised Machine Learning (2-AP-16)

The next lesson "Introduction to Supervised Machine Learning" belongs to Module 4 (Supervised Learning) and addresses the CSTA standard 2-AP-16 of "incorporate existing code, media, and libraries into original programs, and give attribution", as students will be creating original programs that interact with their body movements using two existing tools. The first tool is Teachable Machine, an accessible web-based platform developed by Google that enables convenient training of supervised machine learning models that learn to classify the body movements of the trainer. The second tool is Scratch, a block-based coding interface developed by the MIT Media Lab, for which members of the DAILy development team has developed a specific block that makes it possible for students to load their Teachable Machine models and use them to control their original Scratch codes (Ali et al., 2021).

Figure 4.45.1

Accessible training of supervised machine learning models with Teachable Machine



One thing for teachers to keep in mind while teaching this lesson is that the training of supervised machine learning models via Teachable Machine is another great opportunity for students to peek into the black box of algorithmic bias and the importance of diversity in data. For example, it would be a good practice to let students consider if their models would perform better if they provided Teachable Machine with more variations of the body movements they picked or with simply a greater number of pictures of their body movements. Previous implementations have shown that at this stage of the curriculum, many students would still believe that the latter would suffice, even though the former truly matters.

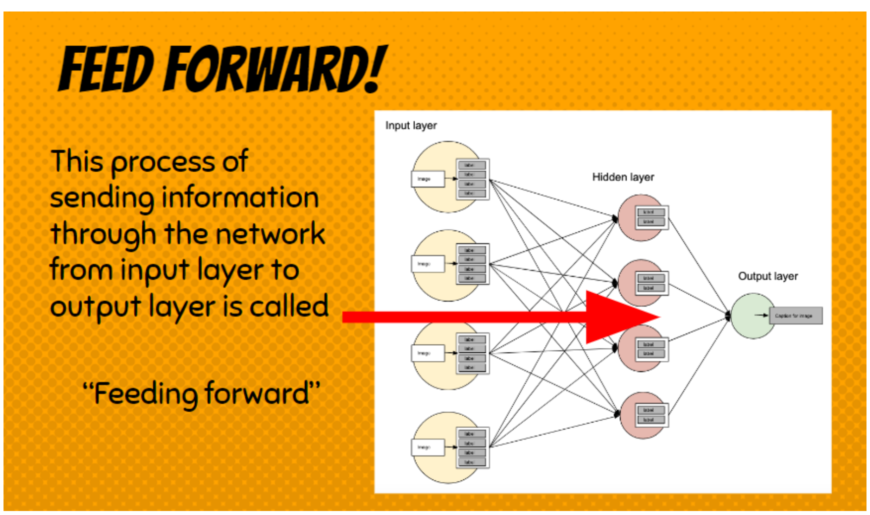
4.46 Neural Networks (3B-AP-09)

The next lesson "Neural Networks" belongs to Module 4 (Supervised Learning) and addresses the CSTA standard 3B-AP-09 of "implement an artificial intelligence to play a game against a human opponent or solve a problem", as they get to play a game in which each of them acts as if they were a component of a neural network to guess the captions of a series of pictures.

Specifically, after being shown a picture, the "input layer" each student picks four words that best describe the picture and pass the words to the "hidden layer" students, each of whom then picks two words they receive and pass to the "output layer" student, who picks four words they receive to create a sentence as the output of the first round (models the procedure of "feeding forward" in the training of neural networks). The true caption of the picture is then revealed, and the "hidden layer" students who are able to feed the output layer student more correct words are assigned higher "weights", and so are the "input layer" students who are able to feed the "hidden layer" students more correct words (models the "backward propagation" in the training of neural networks). In sum, this game models how neural networks learn step by step through rounds and rounds of trials and errors, which allow them to produce better and better output. The next round of the game then continues with a new picture like before, except that words provided by students with higher "weights" are going to be prioritized.

Figure 4.46.1

Neural network



One major challenge in the teaching of this lesson is that steps in the nueral network game can be quite complex to explain to students at the beginning. Fortunately, previous implementations have shown that after playing a round or two, the rules all start to make sense to students. It's therefore recommended that teachers explain the game while playing it with students. Meanwhile, one thing to keep in mind while playing the game is that the ultimate purpose of the game is not to guess the captions correctly, but to guide students to see that the neural network they form are able to get better and better results because of continuously guessing and assigning higher weights to the providers of more correct guesses, which is largely similar to how actual neural networks learn.

4.47 What are Deepfakes (3B-IC-25)

The lesson "What are Deepfakes" belongs to Module 5 (Unsupervised Learning) and addresses the CSTA standard 3B-IC-25 of "evaluate computational artifacts to maximize their beneficial effects and minimize harmful effects on society", as students will learn from this lesson ways in which deepfakes should and should not be used for social-ethical considerations. Specifically, they will be introduced to how deepfakes are AI-enabled products that blends real audio, photo, and video source data into fake outputs, strategies they can use to identify deepfakes, and consider the consequences of deepfakes.

Figure 4.47.1

Examples of deepfakes with real life impacts

While teaching this lesson, it is important to avoid spending too much time on introducing strategies to identify deepfakes. As interesting as this process can be fore students, students should stay aware that there are no accurate ways to call out deepfakes every time, and they should not be too concerned with catching deepfakes to start with. Instead, they simply need to be aware of the positive and negative impacts deepfakes can have on society so that they can make decisions on how best to make use of deepfakes.

4.48 Environmental Impact of AI (3A-IC-24)

The lesson "Environmental Impact of AI" addresses the CSTA standard 3A-IC-24 of "evaluate the ways computing impacts personal, ethical, social, economic, and cultural practices", as students will explore the usually unseen environmental cost of training AI. Specifically, they will look a series of statistics, including the amounts of carbon dioxide emitted by (1) a car in a year, (2) by the average household in a year, and (3) by the power plants because of the training of a large neural network that makes Google Translate possible. After seeing that the third number is way higher than the first two, they engage in discussion of ways to mitigate the environmental impact of training AI.

Previous implementations have shown that students may not be aware of the connection between carbon dioxide emission and the generation of electricity and why excessive emission of carbon dioxide is harmful, which necessitates a quick review of some background knowledge about fossil fuels and global warming. Meanwhile, some students may feel powerless about what they can do and sad about how their households are contributing to the global warming, albeit a small contribution compared to the training of large neural networks. In situations like this, it would be a good practice to engage them in discussion of useful strategies to reduce carbon emission, including but not limited to planting more trees and avoiding training large AI systems unless there are no other alternatives or when the benefits outweigh the costs.

4.5 Conclusion

Overall, DAILy aligns well with CSTA standards in many aspects, especially when it comes to evaluating the impact of technology on society and issues of bias and accessibility, and offers abundant learning opportunities for students to develop knowledge about and skills in AI concepts and AI ethics. It is my hope that this practitioner study will serve as a valuable guide for teachers interested in using DAILy in their virtual and physical classrooms to prepare for potential challenges in teaching and learning and make AI education more engaging and accessible.

5. Study 3: Exploration of the Structure of an AI Concept Inventory

5.1 Abstract

The purpose of this study is to examine a novel AI concept inventory to reveal its internal structure and provide a reference point for further validation work, such as dimensional analysis and item difficulty item analysis. This is achieved through answering the following three questions: (1) what constructs are intended to be measured by the concept inventory; (2) are there an appropriate number of items designed for each

cognitive level; and (3) what misunderstandings about AI can be detected by this concept inventory?

5.2 Introduction

Artificial Intelligence (AI) has become an inseparable part of our life. From the manufacturing of automobiles to the targeted advertising of mobile apps, every corner of society is feeling the increasingly ubiquitous influence of AI in some ways. However, the development of relevant instruments remains largely in its infancy (Taylor et al., 2014), which necessitates education researchers and practitioners to explore ways of assessing everyone's knowledge about AI, especially considering the projection that 14% of the global workforce may have to either upgrade their skills to work with intelligent machines or end up being displaced by them by 2030 regardless of their occupations (Manyika et al., 2017).

As a beginning step toward addressing this pressing challenge, 5 tenets of AI education have been established by the AI4K12 initiative (Touretzky et al., 2019) as an outline of basic ideas that youth should know about AI, including 1) computers perceive the world using sensors; 2) agents maintain models/representations of the world and use them for reasoning; 3) computers can learn from data; 4) making agents interact with humans is a substantial challenge for AI developers; and 5) AI applications can impact society in both positive and negative ways. Based on this framework, Long and Magerko (2020) further summarized that AI literacy should be measured using 5 questions as building blocks, namely 1) "what is AI"; 2) "what can AI do"; 3) "how does AI work"; 4) "how should AI be used"; and 5) "how do people perceive AI".

5.21 What are some challenges developing instruments in AI education?

While these questions have received wide approval, given the numerous strands of AI concepts that need to be covered, the majority of existing instruments for AI literacy are designed to be singular tools, each of which assesses no more than one specific dimension of AI literacy (Samarakou, Fylladitakis, Prentakis, & Athineos, 2014) due to the complexity of mapping out an extensive net of interwoven AI concepts. While these singular instruments suffice for the specific domains covered, they are not as effective when AI educators intend to learn about their students' overall progress in developing AI literacy.

Meanwhile, these instruments can be viewed as high-stake tests of their abilities, which may cause their respondents to suffer stereotype threats (Gordon, 2019), a stressful emotional state that occurs when one senses a risk of confirming a negative stereotype about racial, gender, religious, and other groups they belong to (Steele & Aronson, 1995). When triggered, stereotype threats create an extra emotional and physiological burden and consume the working memory of their victims as they try to suppress the negative reactions, which, more often than not, makes it unnecessarily hard for stereotype-threatened respondents to demonstrate their true academic and career performance (Schmader et al., 2008). Furthermore, low academic and career attainments reflected by these tests continue to harm these respondents' self-efficacy in computer science, forming a self-sustaining downward loop that drags them farther and farther away from academic and career success in computer science (Schmader et al., 2018).

Computer scientists, unfortunately, is one of those groups that has been assigned a stigmatizing label that carries such negative connotations that elementary, middle school, high school, and college students alike, including those who are interested in and competent at computer science, refrain from associating themselves with (Luo et al., 2018; Momsen et al., 2010) due to powerful social persuasions that portray computer scientists as doing non-creative and life-irrelevant work (Masnick et al., 2010). As a common coping mechanism students resort to in cases when they are repeatedly stereotype-threatened, they tend to disassociate themselves with the group by no longer viewing themselves as future computer scientists and evading situations in which they are going to be tested for their competence at computer science (Aronson, Fried, & Good, 2002). Specifically, numerous studies have shown that stereotype threat is the major reason that women and racially and ethnically minoritized groups quit pursuing STEM disciplines, thereby widening the existing achievement gaps between them and Whites and Asian males (Stoet & Geary, 2012).

5.22 Why is concept inventory a viable but immature alternative?

One approach to reducing the risk of stereotype threat while maintaining a precise measurement of students' learning attainments is to use alternative, low stake tests (Gordon, 2019). Concept inventory, a concise set of items aimed at revealing one's conceptual understanding and misunderstanding of fundamental disciplinary concepts, is one of such choices (Madsen, McKagan, & Sayre, 2017). The first concept inventory was developed for Newtonian concepts about force by a group of physics educators (Hestenes, Wells, & Swackhamer, 1992) and has enabled educators in many other STEM disciplines to initiate pedagogical transformations that better support students' learning experience accordingly (Crouch & Mazur, 2001).

Unfortunately, in the field of AI education, the development of a valid measurement tool remains largely in its infancy (Taylor et al., 2014). Unidimensional instruments that each exclusively measure students' understanding of a particular dimension of AI concepts, which makes it difficult to obtain full-scope, systemic findings about students' understanding about AI concepts in general or compare across results generated from different instruments (Almstrum et al., 2006). Scholars like Porter, Taylor, and Webb (2014) have therefore suggested that multidimensional, modular concept inventories should be developed in ways that consist of multiple sub-inventories, each of which corresponds to a subset dimension of AI concepts, so that they can be flexibly refined, expanded, and unified if needed.

Another jigsaw that is missing in the design of existing instruments for AI education is the connection between measurement tools for AI education and the cognitive levels of understanding needed to answer them. Cognitive levels of understanding, as defined by Bloom's Taxonomy, (Anderson, Krathwohl, & Bloom, 2001), are reflected by a sequence of behavioral objectives, including Remember, Understand, Apply, Analyze, Evaluate, and Create (see Section 2.4 for more details about the definition of these terms). Students' cognitive levels of understanding are important to measure because they are representative of different degrees of mastery of knowledge and can be obscured by instruments that do not differentiate between different cognitive levels. As discovered by Hestenes, Wells, and Swackhamer (1992), respondents could perform quite well in tests with low demands for cognitive levels of understanding even if they possessed fundamental misunderstandings about a discipline through doing rote learning, such as memorizing facts and equations. It's therefore important to make sure that AI literacy instruments contain an appropriate set of items that reflect a wide range of levels of cognitive understanding about AI in order to measure authentic and comprehensive development of respondents.

5.23 What does this study aim to achieve?

Following Long and Magerko's (2020) call for modularity, a novel concept inventory of AI concepts (AICI) was recently developed by a group of researchers at Massachusetts Institute of Technology at Boston College. Designed as the culmination of a selection of basic dimensions of AI concepts and their subsets, AICI proved to measure positive shifts in the understanding of several domains of AI concepts among students who have attended the corresponding summer online intervention programs (Zhang, Lee, Ali, DiPaola, Cheng, & Breazeal, 2022). However, the internal structure of AICI, a prerequisite for effectively performing validity tests such as dimension analysis item difficulty analysis (Akaike, 1974), has not been completely sorted out. First, while each item in AICI was designed to measure one definite dimension of AI concepts, it is currently unclear which components of AI literacy outlined by Long and Magerko (2020) the items can be mapped to. Second, even after the first shortcoming is fixed, more effort is needed to identify how the items map to different cognitive levels of understanding, as defined in Bloom's Taxonomy (Anderson, Krathwohl, & Bloom, 2001).

Recognizing these two aspects of weakness, in this aspect I aim to thoroughly examine and clarify the internal structure of AICI through exploring how the items in AICI can be mapped to Touretzky's (2019) AI4K12 initiative as well as the cognitive levels outlined by Bloom's Taxonomy. Due to the lack of student data, this process will be fully conceptual, but nevertheless paves the road for further statistical analysis after data from a large enough sample have been collected by locking down possible hypotheses to be made. Specifically, this study is driven by and aimed at answering three guiding questions as follows.

5.24 Research Question 1: How are aspects of AI literacy measured by AICI

The first research question focuses on the match between items in AICI and the five areas of knowledge about AI that youth should master as outlined by the AI4K12 initiative (Touretzky, 2019), a comprehensive and age appropriate . Successfully answering this question is important for addressing the theoretical significance of AICI by further clarifying the internal structure of AICI, as it would help take a crucial preparatory step toward creating the construct map for AICI, a key feature of theoretically meaningful instruments (Wilson, 2004).

5.25 Research Question 2: How are cognitive levels of understanding measured by AICI

The second research focuses on the match between items in AICI and Bloom's Taxonomy of cognitive levels of conceptual understanding. This question is important to answer, as it offers insights into the cognitive depth of AICI. Doing so further enriches the preparation work needed to build the construct maps for AICI as it helps reveal the powerfulness of AICI in differentiating between students possessing different levels of cognitive understanding of AI concepts.

5.26 Research Question 3: How useful is AICI in revealing misunderstandings about AI

The third research question focuses on the usefulness of AICI as a tool to detect misunderstandings about AI for AI educators to modify their curricular designs accordingly. While answering this question does not necessarily complement the other two, given that it's unlikely to add anything new to the theoretical structure of AICI, it reveals situations in which partial credits should be given for incompletely true responses provided by students. The outcome I expect to achieve is a series of misconceptions that AICI may uncover, which can then be used by AI educators as the ground for targeted instruction. In addition, for items that do have the potential reveal certain misunderstandings about AI, I will suggest ways to refine and enable them to better help AI educators discover misconception their students have about AI.

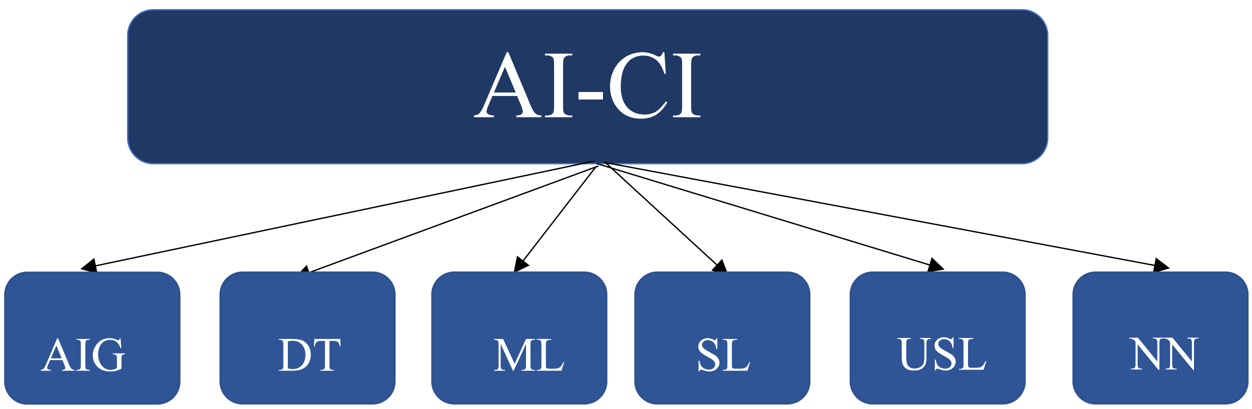
5.3 Dimensions of AI Concepts

AICI contains 31 multiple choice questions designed to measure youth's understanding of the following dimensions of basic concepts of AI that young adolescents should be knowledgeable about, given how AI can impact their lived experiences and communities in unprecedented ways (Williams et al., 2019). Before mapping these dimensions to the building blocks of AI literacy suggested by the AI4K12 initiative (Touretzky, 2019), here is a brief introduction of the conceptual dimensions themselves.

The first dimension is "AI General (AIG)", which measures students' understanding of what AI is and what it is not. The second dimension is "Decision Trees (DT)", which measures students' understanding of how decision trees, a fundamental logic system, function to classify data into categories. The third dimension is "Machine Learning General (ML)", which measures students' understanding of the difference between supervised learning and unsupervised learning. The fourth dimension "Supervised Learning (SL)", a secondary dimension of the third dimension, specifically measures students' understanding of how supervised learning systems make predictions based on the datasets they are given. The fifth dimension "Unsupervised Learning (USL)", another secondary dimension of the third dimension, specifically measures students’ understanding of how unsupervised learning systems work, with a particular focus on what they know about Generative Adversarial Networks (GANs). Finally, the sixth dimension “Neural Networks (NN)” measures students’ understanding of the terminology and the sequence of neural networks. In sum, these six dimensions are highly correlated and distinguishable between each other and can be aggregated to measure students’ conceptual understanding of basic AI concepts.

Figure 5

Theoretical map of AICI with dimensions of AI concepts



5.4 AI4K12 Initiative

What does youth nowadays truly need to know about AI? A joint effort by the Association for the Advancement of Artificial Intelligence (AAAI) and the Computer Science Teachers Association (CSTA) resulted in five areas of knowledge about AI known as Five Big Ideas that provide a working guideline for K-12 AI education (Touretzky, 2019). Specifically, Five Big Ideas can be summarized as follows.

Perception. Youth who master the area of knowledge of Perception are aware that AI systems “perceive” the world through extracting and processing meaning from sensory signals, which allows them to take actions accordingly.

Representation and Reasoning. Youth who master the area of knowledge of Representation and Reasoning know that AI systems build representations of raw data and that algorithms allow them to reason through such representations.

Learning. Youth who master the area of knowledge of Learning are knowledgeable about how AI systems learn from training data provided by humans or machines through finding patterns and making statistical inferences.

Natural Interaction. Youth who have knowledge about the area of Natural Interaction know that there are many complex social and cultural factors at play when it comes to making AI systems perform in ways that are close to humans.

Social Impact. Youth who are knowledgeable about the area of Social Impact know that AI systems can impact society in both positive and negative ways, which means that attention must be paid to

5.5 Bloom's Taxonomy in the Context of AICI

Bloom's Taxonomy describes a sequence of cognitive levels of mastery of knowledge as reflected by clearly defined behavioral objectives (Anderson, Krathwohl, & Bloom, 2001). It has been widely used in the design of measurement tools in STEM disciplines (Wilson, 2004), but not so much in the context of AI education (Taylor et al., 2014) yet. Given the lack of established connections between Bloom's Taxonomy and items for AI concepts, in this section I will build on the original definitions of cognitive levels in Bloom's Taxonomy (see Section 2.4 for detail) and lay out how the levels could refer to in relation to items in AICI.

Remember. In the original Bloom's Taxonomy, the first cognitive level, Remember, refers to memorizing facts and recalling them when needed. In the context of

AICI, an item should be classified as this level if one can answer the item through remembering and repeating what they have read about definitions of AI concepts.

Understand. The second cognitive level, Understand, originally refers to describing learned knowledge and translating it. In the context of AICI, an item should be classified as this level if one can answer the item through explaining what they have learned about AI concepts in their own words.

Apply. The third cognitive level, Apply, is defined as implementing what one has learned in situations that are different from the original context of learning. In the context of AICI, an item should be classified as this level if the item involves challenging respondents to think about a specific scenario they have not encountered before and make use of what they have learned about AI to solve problems.

Analyze. The fourth cognitive level, Analyze, means organizing learned ideas and building connections between them through comparing and contrasting. In the context of AICI, an item should be classified as this level if it requires respondents to compare and contrast between similar AI concepts and make sense of the relationship between them.

Evaluate. The fifth cognitive level, Evaluate, refers to arguing for or against stands and decisions after considering and weighing multiple views and sources of information. In the context of AI-CI, an item should be classified as this level if it involves asking respondents to make arguments for or against a view about AI after carefully taking various factors into considerations.

Create. The sixth cognitive level, Create, centers generating new products through engaging in original, creative work. In the context of AICI, an item should be classified as this level if it involves the creation something new based on respondents' original ideas using AI.

5.6 Labeling the Items

Having described the theoretical backgrounds for labeling items in AICI, I will dedicate this section specifically to demonstrating how exactly the labels are eventually labeled using a series of acronyms. After that, for the sake of readability, I will take the cases of five items as examples of how the labeling procedure is performed on the full 31 items in AICI in subsections 6.61 to 6.65.

To start, first, given the three theoretical frameworks mentioned in the previous section, each item in AICI is going to be assigned three types of labels, namely six dimension of AI concepts as described in Section 6.3, five Big Ideas of AI knowledge as described in Section 6.4, and six cognitive levels in Bloom's Taxonomy as described in Section 6.5. Specifically, first, the dimensions of AI concepts will be abbreviated as AIG, DT, ML, SL, USL, and NN, as shown in Table 5.6.1.

Table 5.6.1

Acronym chart for dimensions of AI concepts

Dimension of AI Concepts	Acronym	Meaning
AI General	AIG	general knowledge about what AI is and is not

Decision Trees	DT	knowledge about how decision trees work
Machine Learning	ML	general knowledge about what machine learning is
Supervised Learning	SL	knowledge about how supervised learning works
Unsupervised Learning	USL	knowledge about how unsupervised learning works
Neural Networks	NN	knowledge about how neural networks work

Second, five Big Ideas of AI knowledge will be labeled as PE, RE, LE, NI, and SI respectively, as shown in Table 5.6.2:

Table 5.6.2

Acronym chart for Five Big Ideas

Big Idea	Acronym	Meaning

Perception	PE	knowledge about how AI systems perceive the world
Representation & Reasoning	RE	knowledge about how AI systems reason through representations of data
Learning	LE	knowledge about how AI systems learn through inferring from data
Natural Interaction	NI	knowledge about social and cultural features AI systems need to possess to approach humans
Social Impac	AIL5	knowledge about how AI systems can impact society in both positive and negative ways

Third, cognitive levels in Bloom's Taxonomy will be abbreviated as REM, UND, APP, ANA, EVA, and CRE, as shown in Table 4.

Table 4

Acronym chart for Bloom's Taxonomy

Cognitive Level	Acronym	Meaning
Remember	REM	remember and repeat definitions of AI concepts
Understand	UND	summarize and explain AI concepts
Apply	APP	execute and implement knowledge about AI
Analyze	ANA	differentiate between and organize AI concepts
Evaluate	EVA	evaluate, critique, and argue for or against different points of view
Create	CRE	design and generate original products

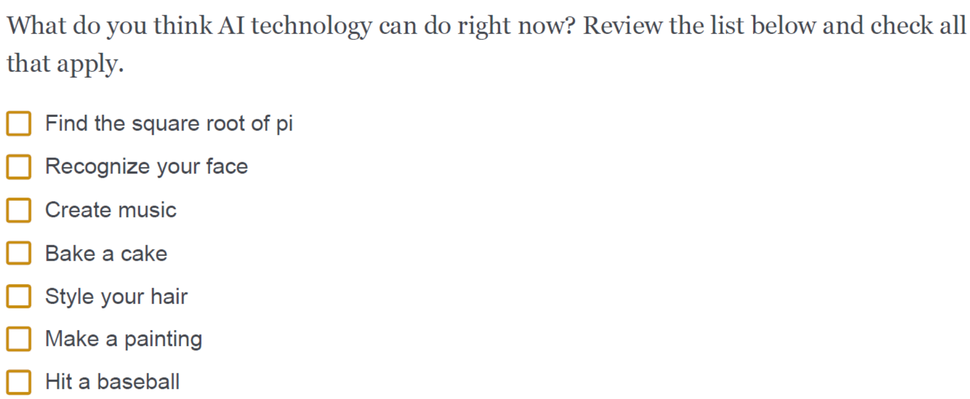
Next, in the following subsections I will provide a few sample items and demonstrate the process of how they are labeled and the reasoning behind.

5.61 Item 1 AIG-SI-APP

The first item in AICI is a multiple choice question that asks respondents to choose among a list of options what they think AI can do at the moment, as shown in Figure 6.1:

Figure 6.1

AICI Item 1



What do you think AI technology can do right now? Review the list below and check all that apply.

- Find the square root of pi
- Recognize your face
- Create music
- Bake a cake
- Style your hair
- Make a painting
- Hit a baseball

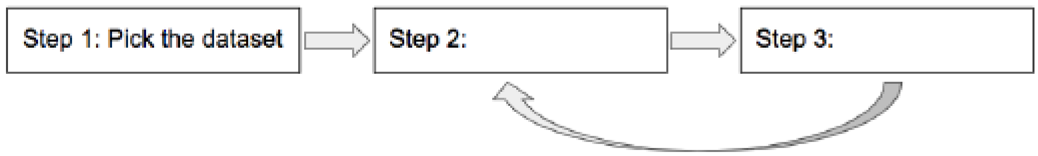
In terms of the dimension of AI concepts covered, this item should be labeled as AIG-SI-APP. First, the key step it takes to the task of picking the right choices in the provided list is differentiating between what AI does and what it does not, which exactly fits the scope of AIG. Second, in terms of the Five Big Ideas, SI is the appropriate label since this item reveals respondents' knowledge about how AI can impact our life. Third, Apply is the cognitive level required to answer this question from a Bloom's Taxonomy perspective, because here in this item respondents need to consider a variety of scenarios for which they need to apply what they know about AI to make decisions. Overall, the label for this item should be AIG-SI-APP.

5.62 Item 10 DT-RE-UND

The 10th item in AICI tests students' knowledge about how decision trees work in general by presenting them with an incomplete conceptual diagram of three steps decision trees take to work and asking them to fill out two missing steps, as shown in Figure 6.2. As such, the dimension of AI concept covered is DT, as students need knowledge about decision trees to reasonably select the steps that should be taken. RE is the Big Idea addressed, since it takes knowledge about how decision trees structurally represent data. The cognitive level needed for this item is UND, as students are not dealing with a practical situation, but a highly conceptual case here. As long as they are able to interpret the definitions of decision trees and recall them in a modified way, they will be able to select the correct steps that aren't too far off from the definitions. Overall, the label for this item is DT-RE-UND.

Figure 6.2

AICI Item 10



Please select what the computer should do in step 2

- ○ Decide what the leaves of the tree would be
- ○ Split the set based on the questions
- ○ Pick the height of the tree
- ○ Pick the features of the data to ask about

Please select what the computer should do in step 3

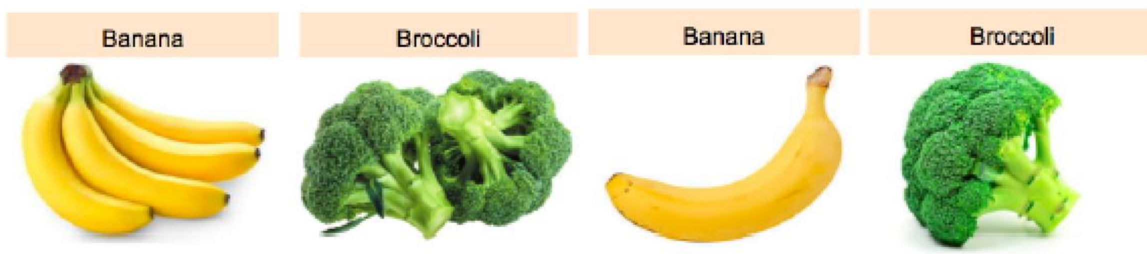
- ○ Decide what the leaves of the tree would be
- ○ Split the set based on the questions
- ○ Pick the height of the tree
- ○ Pick the features of the data to ask about

5.63 Item 11 SL-LE-APP

Item 11 asks respondents to select a prediction that an AI is most likely to make after being trained with a specific dataset of four pictures of broccolis and bananas, as shown in Figure 6.3. This item is a test of students' knowledge about how supervised learning enables AI to learn from the data it is provided with and thus should be labeled as SL. The Big Idea relevant to this item is LE, as respondents need to know about how machine learning enables AI systems to learn from training data. The cognitive level needed to answer this item is APP, as respondents will need to apply what they learn about machine learning to a new scenario. Overall, the item should be labeled SL-LE-APP.

Figure 6.3

AICI Item 11



Which of the following would you expect the technology to predict?

- ○ All things in multiples are broccoli
- ○ All crunchy things are broccoli
- ○ All things in multiples are bananas
- ○ All yellow things are bananas

5.64 Item 16 USL/NN-LE-REM

Item 16 asks respondents to determine whether a series statements about neural networks and two components of Generative Adversarial Networks (GANs), generator and discriminator, are true or not, as shown in Figure 6.4. The dimension of knowledge being tested in this item includes NN and USL, as respondents will need to know both what neural networks are and how unsupervised learning enables GANs to learn in order to correctly pick the right answers. The Big Idea involved here is LE, as answering this item requires knowledge about how supervised and unsupervised machine learning enable AI systems to learn. The cognitive level it takes to answer this item is Remember, since respondents will be able to pick the right answers as long as they remember basic definitions of discriminator and generator. Overall, the label for this item is USL/NN-LE-REM.

Figure 6.4

AICI Item 16

	True	False
A generator and discriminator are both neural networks	O	O
The discriminator gives feedback to the generator	O	O
The generator and discriminator are working in competition with one another	O	O

5.65 Item 18 ML-LE-ANA

Item 18 challenges respondents to determine whether a social media that can recognize one's face and tag it their name is an example of classifying or generating, as shown in Figure 6.5. In this item, students need knowledge about machine learning in general to differentiate between classifying and generating, thus ML is an appropriate label. The Big Idea addressed here is LE, as classification and generation are both outcomes of AI systems' learning. Analyze is the cognitive level needed for this item, as respondents are going to have to differentiate between two similar yet different concepts in machine learning. Overall, the label for this item is ML-LE-ANA.

Figure 6.5

AICI Item 18

A social media site just released a feature that allows you to upload a photo of yourself, and the system will recognize your face and tag your name.

Is this an example of an AI classifying or a generating?

- Classifying
- Generating

5.7 Conclusion

Overall, in terms of the 5 big ideas from the AI4K12 initiative, 24 items address the idea of Learning and 7 the idea of Representation and Reasoning. This indicates that the current version of AICI primarily focuses on testing respondents' knowledge about how AI systems learn from data. The complete absence of Perception, Natural Interaction, and Social Impact, however, weakens the capability of AICI in measuring respondents' knowledge about social-ethical issues relevant to AI. Meanwhile, in terms of Bloom's Taxonomy, 16 items address the level of Applying, 7 the level of Analyzing, 6 the idea of Understanding, and 2 the idea of Remembering. This reflects the strength of AICI in measuring respondents' ability to apply their knowledge in different scenarios. However, none of the items address the levels of Evaluate and Create, which limits the power of AICI to measure respondents' ability to weigh the benefits and costs of AI systems and create original content.

Given these findings, AICI could benefit from the removal of over-represented items (those that address Learning and Applying) and addition of items that address the missing big ideas and cognitive levels, if AICI were to holistically measure respondents' AI literacy.

References

AAAI. (2020). Association for the Advancement of Artificial Intelligence Homepage. Retrieved from AAAI: https:// www. aaai. org/

Aikenhead, G. S. (1996). Science education: border crossing into the subculture of science. *Studies in Science Education, 27*(1), 1–52.

Akaike, H. (1974). A new look at the statistical identification model. *IEEE Transaction Automatic Control, 19*, 716–723.

Akkermans, J., & Tims, M. (2017). Crafting your career: how career competencies relate to career success via job crafting. *Applied Psychology, 66*, 168-195.

Ali, S., DiPaola, D., Lee, I., Hong, J., & Breazeal, C. (2021). Exploring Generative Models with Middle School Students. *CHI Conference on Human Factors in Computing Systems*.

Ali, S., Payne, B. H., Williams, R., Park, H. W., & Breazeal, C. (2019). Constructionism, ethics, and creativity: Developing primary and middle school artificial intelligence education. In *International workshop on education in artificial intelligence K-12* (pp. 1–4).

Almstrum, V. L., Henderson, P. B., Harvey, B., Heeren, C., Marion, W., Riedesel, C., & Tew, A. E. (2006). Concept inventories in computer science for the topic discrete mathematics. *ACM SIGCSE Bulletin, 38*, 132-145.

Ambady, N., Paik, S. K., Steele, J., Owen-Smith, A., & Mitchell, J. P. (2004). Deflecting negative self-relevant stereotype activation: The effects of individuation. *Journal of Experimental Social Psychology, 40*(3), 401-408.

Archer, L., DeWitt, J., Osborne, J., Dillon, J., Willis, B., & Wong, B. (2013). 'Not girly, not sexy, not glamorous': primary school girls' and parents' constructions of science aspirations 1. Pedagogy, Culture and Society, 21(1), 171–194. https://doi.org/10.1080/14681366.2012.748676.

Anderson, L. W., Krathwohl, D. R., & Bloom, B. S. (2001). *A taxonomy for learning, teaching, and assessing : A revision of Bloom's taxonomy of educational objectives*. New York: Longman.

Archer, L., Dawson, E., DeWitt, J., Godec, S., King, H., Mau, A., . . . Seakins, A. (2017). Killing

curiosity? An analysis of celebrated identity performances among teachers and students in nine London secondary science classrooms. Science Education, 101(5), 741-764.

Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology, 38*(2), 113-125.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist, 37*(2), 122–147. https://doi.org/10.1037/0003-066X.37.2.122

Bandura, A., & National Inst of Mental Health. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall, Inc.

Bandura, A. (1995). Exercise of personal and collective efficacy in changing societies. In A. Bandura (Ed.), *Self-efficacy in changing societies* (pp. 1-45). Cambridge University Press.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* W H Freeman/Times Books/ Henry Holt & Co.

Biaggi, C., & Wa-Mbaleka, S. (2018). Grounded theory: A practical overview of the Glaserian school. *JPAIR Multidisciplinary Research, 32*(1): 1–29.

Bloom, B.S. (1956). *Taxonomy of Educational Objectives, Handbook: The Cognitive Domain*. David McKay, New York.

Blotnicky, K. A., Franz-Odendaal, T., French, F., & Joy, P. (2018). A study of the correlation between STEM career knowledge, mathematics self-efficacy, career interests, and career activities on the likelihood of pursuing a STEM career among middle school students. *International Journal of STEM Education, 5*(1), 1–15.

Blustein, D. (2019). *The importance of work in an age of uncertainty: The eroding work experience in America.* Oxford University Press.

Braun, D., & Huwer, J. (2022). Computational literacy in science education: A systematic review. *Frontiers in Education.*

Bull, G., & Bell, L. (2009). TPACK: A framework for the CITE Journal. *Contemporary Issues in Technology and Teacher Education, 9*(1).

Bush, S. B., Cook, K. L., Edelen, D., & Cox, R. (2020). Elementary students' STEAM perceptions: Extending frames of reference through transformative learning experiences. *The Elementary School Journal, 120*(4), 692-714.

Calzada, I. (2021). *Digitranscope: The governance of digitally transformed society.* Luxembourg: Publications Office of the European Union.

Carsten Conner, L.D., Tsurusaki, B.K., Tzou, C., Sullivan, P.T., Guthrie, M., & Pompea, S.M.

(2019). *Fostering a STEAM mindset across learning settings. Connected Science Learning, 12*, 1-11.

Cavanaugh, C. S. (2001). The effectiveness of interactive distance education technologies in K-12 learning: A meta-analysis. *International Journal of Educational Telecommunications, 7*(1), 73-88.

CC2020 Task Force. (2020). *Computing Curricula 2020: Paradigm for future computing curricula*. Association for Computing Machinery and IEEE Computer Society. New York.

Cheng, Y., Zhang, H., Jackson, D. W., Lee, I. A., Brown, N. J. S., Szendey, O., Ali, S.,DiPaola, S. (2020). *Raising minoritized middle schoolers' AI career awareness and adaptability: Findings from two online summer camps.* Presented at American Educational Research Association.

Cole, D., & Espinoza, A. (2008). Examining the academic success of Latino students in science technology engineering and mathematics (STEM) majors. *Journal of College Student Development, 49*(4), 285–300.

Creswell, J. W. (2013). Qualitative inquiry and research design: Choosing among five approaches. Los Angeles, CA: Sage Publications.

Crouch, C. H., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. American Journal of Physics, 69, 970-977.

Cui, Y., Shang, C., & Chen, S. (2019). Overview of artificial intelligence: Development of AI. *Radio Communication Technology, 45*(3), 225-331.

Curry, J. R., & Milsom, A. D. (2017). *Career and college readiness counseling in P-12 schools* (2nd ed.). New York, NY: Springer.

DAILy Workshop. (2021). DAILy Curriculum Homepage. Retrieved from: https://raise.mit.edu/daily/index.html

DeWitt, S. (2018). Why middle school? Why now? *Techniques, 93*(1).

DiPaola, D., Ali, S., & Breazeal, C. (2021). What are GANs?: Introducing Generative Adversarial Networks to Middle School Students. *Proceedings of the 11th Symposium on Education Advances in Artificial Intelligence*.

DiPaola, D., Payne, B. H., & Breazeal, C. (2020). Decoding design agendas: an ethical design activity for middle school students. Proceedings of the Interaction Design and Children Conference, (pp. 1–10).

Erickson, F. (2006). Definition and analysis of data from videotape: Some research procedures and their rationales. In J. L. Green, G. Camilli, P. B. Elmore, A. Skukauskaite, & E. Grace (Eds.), *Handbook of complementary methods in education research* (pp. 177-191). Mahwah, NJ: Lawrence Erlbaum Associates.

Franklin, S. (1995). Science as culture, cultures of science. *Annual Review of Anthropology, 24*(1), 163–184.

Garad, A., Al-Ansi, A. M., & Qamari, I. N. (2021). The role of e-learning infrastructure and cognitive competence in distance learning effectiveness during the covid-19 pandemic. *Jurnal Cakrawala Pendidikan, 40*(1), 81-91.

Godbey, S., & Gordon, H. R. (2019). Career exploration at the middle school level: Barriers and opportunities. *Middle Grades Review, 5*(2).

Gordon, A. (2019). Don't remind me: Stereotype threat in high-stakes testing. *University of Baltimore Law Review, 48*(3), 387-412.

Hadzigeorgiou, Y. (2016). *Imaginative science education: The central role of imagination in science education*. Switzerland: Springer International Publishing.

Harding, S. (2015). *Objectivity and diversity: Another logic of scientific research*. Chicago, IL: The University of Chicago Press.

Hazen, E. P., Abrams, A. N., & Muriel, A. C. (2016). Child, adolescent, and adult development. In T. A. Stern, M. Fava, T. E. Wilens, & J. F. Rosenbaum (Eds.), *Massachusetts general hospital comprehensive clinical psychiatry* (2nd ed.). Elsevier.

Hu, K. (2016). Misread intelligent machines. *Well-off, 5*, 44-45.

Jin, G. & Bierma, T. (2013). STEM for non-STEM majors: Enhancing science literacy in large classes. *Journal of College Science Teaching, 42*(6), 20-26.

Jordan, B., Devasia, N., Hong, J., Williams, R., & Breazeal, C. (2021). PoseBlocks: A Toolkit for Creating (and Dancing) with AI. *Proceedings of the 11th Symposium on Education Advances in Artificial Intelligence*.

Karacan-Ozdemir, N., & Yerin Guneri, O. (2017). The factors that contribute to career adaptability of high-school students. *Eurasian Journal of Educational Research, 67*, 183-198.

Kerka, S. (2000). *Middle school career education and development.* Center on Education and Training for Employment.

Kintu, M. J., Zhu, C., & Kagambe, E. Blended learning effectiveness: the relationship between student characteristics, design features and outcomes. *International Journal of Educational Technology in Higher Education, 14*(7). https://doi.org/10.1186/s41239-017-0043-4

Kreutzer, K., & Boudreaux, A. (2012). Preliminary investigation of instructor effects on gender gap in introductory physics. *Physical Review Special Topics-Physics Education Research, 8*(1), 010120.

Kurbanoğlu, S. (2003). Self-efficacy: A concept closely linked to information literacy and lifelong learning. *Journal of Documentation, 59*(6), 635-646.

Lee, I., Zhang, H., Moore, K., Zhou, X., Perret, B., Cheng, Y., Zheng, R., and Pu, G. (2021). AI Book Club: An innovative professional development model for AI education. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education*.

Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior, 45*(1), 79-122.

Lent, R. W., Brown, S. D., & Hackett, G. (2002). Social cognitive career theory. *Career Choice and Development, 4*(1), 255-311.

Long, D., & Magerko, B. (2020). What is AI Literacy? Competencies and Design Considerations. Proceedings of the 2020 CHI Conference on Human Factors in

Computing Systems (pp. 1–16). New York, NY, USA: Association for Computing Machinery.

Luo, T., Li, J., & So, W. W. M. (2018). Analysis of students' visual presentations of STEM professionals. Hualien: Paper presented at 2018 International Conference of East-Asian Association for Science Education.

Luo, T., So, W. W. M., Wan, Z. H., & Li, W. C. (2021). STEM stereotypes predict students' STEM career interest via self-efficacy and outcome expectations.

Lyon, J. A., and Magana, A. J. (2020). Computational thinking in higher education: A review of the literature. *Computer Applications in Engineering Education 28*(5), 1174–1189.
doi: 10.1002/cae.22295

Madsen, A., McKagan, S. B., & Sayre, E. C. (2017). Best practices for administering concept inventories. *The Physics Teacher, 55*(9), 530-536.

Maddux, J. E., & Meier, L. J. (1995). Self-efficacy and depression. In J. E. Maddux (Ed.), *Self-efficacy, adaptation, and adjustment: Theory, research, and application* (pp. 143–169). Plenum Press. https://doi.org/10.1007/978-1-4419-6868-5_5

Maddux, J. E., & Kleiman, E. M. (2016). Self-efficacy: A foundational concept for positive clinical psychology. In A. M. Wood & J. Johnson (Eds.), *The Wiley handbook of positive clinical psychology* (pp. 89–101). Wiley Blackwell. https://doi.org/10.1002/9781118468197.ch7

Maltese, A. V. & Tai, R. H. (2011). Pipeline persistence: Examining the association of educational experiences with earned degrees in STEM among U.S. students. *Science Education,* 1-31.

Manyika, J., Lund, S., Chui, M., Bughin, J., Woetzel, J., Batra, P., & Sanghvi, S. (2017). Jobs lost, jobs gained: Workforce transitions in a time of automation. San Francisco, CA: McKinsey Global Institute.

Marginson, S., Tytler, R., Freeman, B., & Roberts, K. (2013). *STEM: country comparisons: International comparisons of science, technology, engineering and mathematics (STEM) education. Final report*. Melbourne: Australian Council of Learned Academies.

Marques, L. S., von Wangenheim, C. G., & Hauck, J. C. R. (2020). Teaching machine learning in school: A systematic mapping of the state of the art. *Informatics in Education*, *2*, 283-321.

Menekse, M. (2015). Computer science teacher professional development in the United States: A review of studies published between 2004 and 2014. *Computer Science Education 25*(4): 325–350.

Miles, M. B., Huberman, M. A., & Saldaña, J. (2014). Qualitative data analysis: Amethod source book (3rd ed.). Los Angeles, CA: Sage Publications.

Minichino, M. J. (2016). Career and technical exploration in middle school. *Techniques, 91*(2), 46-49.

Miyake A, Kost-Smith LE, Finkelstein ND, Pollock SJ, Cohen GL, Ito TA (2010). Reducing the gender achievement gap in college science: a classroom study of values affirmation. Science 330, 1234-1237.

Mo, H. (2020). *Introduction to Artificial Intelligence*. Beijing, China: People's Post Press.

Momsen, J. L., Long, T. M., Wyse, S. A., & Ebert-May, D. (2010). Just the facts? Introductory undergraduate biology courses focus on low-level cognitive skills. *CBE—Life Sciences Education, 9*(4), 435–440.

National Science Foundation, National Center for Science and Engineering Statistics. (2017).

Women, minorities, and persons with disabilities in science and engineering. Retrieved from www.nsf.gov/statistics/wmpd/

Organization for Economic Co-operation and Development (OECD) (2008). *Encouraging student interest in science and technology studies*. Global Science Forum. Retrieved from the internet December 9, 2019: https://www.oecd.org/publications/encouraging-student-interest-in-science-and-technology-studies-9789264040892-en.htm

Porter, L., Taylor, C., & Webb, K. (2014). Leveraging open source principles for flexible concept inventory development. In *Proceedings of the 19th Annual Conference on Innovation and Technology in Computer Science Education*, 243-248. Uppsala.

Redmond. B.F (2010). *Self-efficacy Theory: Do I Think that I Can Succeed in My Work? Work*

Attitudes & Motivations. The Pennsylvania State University; World Campus.

Rosenbloom, P. (2013). *On Computing*. The Fourth Great Scientific Domain. MIT Press.

Rottinghaus, P. J., & Miller, A. D. (2013). Convergence of personality frameworks within vocational psychology . *Handbook of Vocational Psychology*, 105-131.

Rudolph, C. W. (2016). Lifespan developmental perspectives on working: A literature review of motivational theories. *Work, Aging, and Retirement, 2*(2), 130-158.

Samarakou, M., Fylladitakis, E., Prentakis, P., & Athineos, S. (2014). Implementation of artificial intelligence assessment in engineering laboratory education. In *Proceedings of the Multi Conference on Computer Science and Information Systems*. Lisbon, Portugal: International Association for Development of the Information Society.

Samoili S., Righi R., Cardona M., López Cobo M., Vázquez-Prada Baillet M., & De Prato G. (2020). *TES Analysis of AI Worldwide Ecosystem in 2009-2018*. Publications Office of the European Union. doi:10.2760/85212, JRC120106

Savickas, M. L. (2005). The theory and practice of career construction. In S. D. Brown, & R. W. Lent, *Career development and counselling: Putting theory and research to work* (pp. 42-70). Hoboken, NJ: Wiley.

Savickas, M. L., & Porfeli, E. J. (2012). Career adaptabilities scale: construction, reliability, and measurement equivalence across 13 countries. *Journal of Vocational Behavior, 80*(3), 661-673.

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological review, 115*(2), 336.

Schoenfield, A. H. (1985). *Mathematical Problem Solving.* New York, NY: Academic Press.

Shanahan, M.-C., & Nieswandt, M. (2009). Creative activities and their influence on identification in science: Three case studies. *Journal of Elementary Science Education, 21*(3), 63-79.

Smith, M. & Neupane, S. (2022). Artificial intelligence and human development: toward a research agenda.

Southern Regional Education Board. (2011). *A new mission for the middle grades: Preparing students for a changing world.* The SREB Middle Grades Commission.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. Journal of Personality and Social Psychology, 69(5), 797–811. https://doi.org/10.1037/0022-3514.69.5.797

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52*(6), 613–629. https://doi.org/10.1037/0003-066X.52.6.613

Stoet, G., & Geary, D. C. (2012). Can stereotype threat explain the gender gap in mathematics performance and achievement. *Review of General Psychology, 16*(1), 93-102.

Tai, R. H., Liu, C. Q., Maltese, A. V., & Fan, X. (2006). Planning early for careers in science. *Science, 312*(5777), 1143-1144.

Taylor, C., Clancy, M., Webb, K. C., Zingaro, D., Lee, C., & Porter, L. (2020). The practical details of building a CS concept inventory. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education* (pp. 372-378).

Taylor, C., Zingaro, D., Porter, L., Webb, K. C., Lee, C. B., & Clancy, M. (2014). Computer science concept inventories: Past and future. *Computer Science Education, 24*(4), 253-276.

Topi, H., Karsten, H., Brown, S. A., Carvalho, J. A., Donnellan, B., Shen, J., Tan, B. C-Y. and Thouin, M. F. (2017). MSIS 2016 global competency model for graduate degree programs in information systems. *Communications of the Association for Information Systems, 40*(18).

Touretzky, D., Gardner-McCune, C., Martin, F., & Seehorn, D. (2019). Envisioning AI for K-12: What should every child know about AI. AAAI Conference on Artificial Intelligence, 33, pp. 9795-9799.

Tucker, A. *A Model Curriculum for K-12 Computer Science: Field Report of the ACM K-12 Task Force Curriculum Committee.* New York, NY: Association for Computing Machinery.

Wang, C., Shen, J., & Chao, J. (2021). Integrating computational thinking in stem education: A literature review. *International Journal of Science and Mathematics Education*, 1-24.

Wilson, M. (2004). *Constructing measures: An item response modeling approach.* New York, NY, USA: Routledge.

Wu, C., Seokin, K., & Zhang, L. On GANs art in context of artificial intelligence art. In *Proceedings of the 5th International Conference on Machine Learning and Soft Computing.* New York, NY, USA: Association for Computing Machinery.

Yadav, A. (2017). Computer science teacher professional development: Towards a research agenda on teacher thinking and learning. In *Proceedings of the 12th Workshop on Primary and Secondary Computing Education*. New York, NY, USA: Association for Computing Machinery.

Yoon, K. S., Duncan, T., Lee, S., Scarloss, B., & Shapley, K. L. (2007). *Reviewing the evidence on how teacher professional development affects student achievement.* Issues & Answers. ERIC Clearinghouse.

Zhang, H., Lee, I., Ali, S., DiPaola, D., Cheng, Y., & Breazeal, C. (2022). Integrating Ethics and Career Futures with Technical Learning to Promote AI Literacy for

Middle School Students: An Exploratory Study. International Journal of Artificial Intelligence in Education, 1-35.

Appendix 1

AICI Items

	Dimension	Big Idea	Cognitive Level	Content
Item 1	AIG	RE	APP	What do you think AI technology can do right now? Review the list below and check all that apply. ☐ Find the square root of pi ☐ Recognize your face ☐ Create music ☐ Bake a cake ☐ Style your hair ☐ Make a painting ☐ Hit a baseball
Item 2	AIG	RE	APP	AI / Not AI A software that can classify different types of hats A drawing app that can generate new paintings by combining features from other paintings A face recognition software that can tell whether you or your friends are in the picture A smartphone app that reminds people to turn off the lights in their house if the time is between 9am to 3pm
Item 3	AIG	LE	APP	Yes / No Does this app use a dataset? Does this app make a prediction?
Item 4	AIG/ML	LE	APP	*Answer the following questions based on the technology described above.* Yes / No Does this automatic door use a dataset? Does the automatic door make a prediction?
Item 5	AIG/ML	LE	APP	Yes / No Does this app use a dataset? Does the app make a prediction?
Item 6	AIG/ML	LE	APP	*Answer the following questions based on the technology described above.* Yes / No Does this software use a dataset? Does the software make a prediction?

Item 7	DT	RE	APP	Imagine you categorize a blueberry using this decision tree.  In which category will it end up? ○ A ○ B ○ C ○ D
Item 8	DT	RE	APP	Pick which question was most likely used at the "?" oval to classify this set of fruit: ○ Is it long? ○ Is it blue? ○ Is it small? ○ Is it sour?
Item 9	DT	RE	APP	Pick which question was most likely used at the "?" oval to classify the set of fruit: ○ Is it red? ○ Is it large? ○ Is it blue? ○ Is it small?
Item 10	DT	RE	UND	Please select what the computer should do in step 2 ○ Decide what the leaves of the tree would be ○ Split the set based on the questions ○ Pick the height of the tree ○ Pick the features of the data to ask about
Item 11	SL	RE	APP	Please select what the computer should do in step 3 ○ Decide what the leaves of the tree would be ○ Split the set based on the questions ○ Pick the height of the tree ○ Pick the features of the data to ask about
Item 12	SL	LE	APP	Which of the following would you expect the technology to predict? ○ All things in multiples are broccoli ○ All crunchy things are broccoli ○ All things in multiples are bananas ○ All yellow things are bananas
Item 13	SL	LE	APP	Based on the above dataset, how would it classify the image below? 
Item 14	SL	LE	APP	Based on the above dataset, how would it classify the following image below?  ○ Broccoli ○ Banana

Item 15	SL	LE	APP	 ○ Broccoli ○ Bananas ○ It will be the same for broccoli and bananas
Item 16	USL/NN	LE	REM	During the training phase, if the technology identifies incorrectly, what does it do to improve the system? ○ Change the label of the picture so that the AI technology won't identify it wrong ○ Strengthen or weaken connections in the network ○ Ask the algorithm to remember the picture that AI technology identifies wrong ○ Delete the picture
Item 17	SL	LE	APP	The discriminator gives feedback to the generator ○ ○ The generator and discriminator are working in competition with one another ○ ○
Item 18	ML	LE/SI	ANA/APP	○ Some brand labels don't make women's shoes ○ Women's shoes are smaller than men's shoes, making them difficult to identify ○ The dataset it uses to learn to identify shoes only includes women's shoes ○ The dataset it uses to learn to identify shoes only includes men's shoes
Item 19	ML	LE/SI	ANA/APP	A social media site just released a feature that allows you to upload a photo of yourself, and the system will recognize your face and tag your name. Is this an example of an AI classifying or a generating?
Item 20	ML	LE/SI	ANA/APP	A doctor sends x-ray scans out to an AI to determine whether or not you have broken a bone. Is this an example of an AI classifying or a generating? ○ Classifying ○ Generating
Item 21	USL	LE/SI	APP	An app converts the style of your face to a Van Gogh painting. Is this an example of an AI classifying or a generating? ○ Classifying ○ Generating
Item 22	USL	LE	ANA/APP	A GAN is being trained to generate images of clouds. The generator creates an image and sends it over to the discriminator. The discriminator does not classify the image as a cloud. What happens next? ○ The discriminator generates a new image ○ The GAN stops running ○ The generator generates a new image
Item 23	USL	LE	ANA/APP	A machine is given a dataset of thousands of audio clips (containing sounds only) and generates a new audio clip similar to the dataset. ○ Unsupervised Learning ○ Supervised Learning
Item 24	USL	LE	ANA/APP	A machine is given a dataset of images of flowers with the identification of whether the image is a of a "tulip" or "not tulip" and is trained to identify when an image contains a tulip. ○ Unsupervised Learning ○ Supervised Learning
Item 25	USL	LE	ANA/APP	A machine is given a dataset of images of dogs (without naming them) and can group different dogs by their types. ○ Unsupervised Learning ○ Supervised Learning

Item 26	USL	LE	UND	Which of the following is a difference between the training and testing phases in supervised learning? ○ Where each phase takes places ○ The dataset used in each phase ○ Whether or not the data is labeled ○ How the prediction is scored
Item 27	SL	LE	UND	What is the purpose of a label on an image in Supervised Learning? ○ The label is used to check if the prediction made was correct ○ The label is used to see if the image contains words ○ The label is used to tell the supervisor to classify a model ○ The label is not useful in supervised learning
Item 28	NN	LE	REM	What are the layers in a middle of a neural network called? ○ Beginning, middle, and end layers ○ First, second, and third layers ○ Input, hidden, and output layers ○ Big, medium, and small layers
Item 29	NN	LE	UND	What happens in the second step? ○ The information gets learned ○ A prediction is made ○ The prediction gets scored ○ Information is sent back to the input nodes
Item 30	NN	LE	UND	Understanding Which of the following ways enables a neural network to learn? (check all that apply) ☐ By being told what to do by a supervisor ☐ By making a guess and finding out if the guess was right or wrong ☐ By rearranging the network ☐ By strengthening connections in the network ☐ By changing the algorithm it uses to make guesses
Item 31	NN	LE	UND	Understanding In Neural Networks, when does learning happen? ○ During the "testing phase" when a neural network is tested to see how good it is at making predictions on new data ○ During the "training phase" when a neural network is built using training data