COLLABORATIVE PLATFORM FOR COMPUTATIONAL THINKING ASSESSMENT

by

Arjun Shakdher

A Thesis

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

Department of Computer and Information Technology

West Lafayette, Indiana

May 2019

THE PURDUE UNIVERSITY GRADUATE SCHOOL

STATEMENT OF COMMITTEE APPROVAL

Prof. Alka Harriger, Chair

Department of Computer and Information Technology

Dr. John Springer

Department of Computer and Information Technology

Dr. Baijan Yang

Department of Computer and Information Technology

Approved by:

Dr. Eric T. Matson

Head of the Graduate Program

ACKNOWLEDGMENTS

I wish to gratefully acknowledge Prof. Alka Harriger, Dr. John Springer and Dr. Baijan Yang for their insightful comments, guidance, and encouragement. I would also like to thank my parents for their continuous support throughout my education.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CS	Computer Science
CSTA	Computer Science Teachers Association
CT	Computational Thinking
CTt	Computational Thinking Test
CTSiM	Computational Thinking using Simulation and Modelling
IRB	Institutional Review Board
IS	Information System
ISTE	International Society for Technology in Education
ECN	Engineering Computer Network
NSF	National Science Foundation
NRC	National Research Council
PECT	Progression of Early Computational Thinking
PUEU	Perceived Usefulness and Ease of Use
RA	Response Average
SUS	System Usability Scale
STEM	Science, Technology, Engineering and Mathematics
TAM	Technology Acceptance Model
TESWES	The Educational Software/Website Effectiveness Survey
TECHFIT	Teaching Engineering Concepts to Harness Future Innovators and Technologists
WEBLEI	Web based Learning Environment Instrument

GLOSSARY

Computational Thinking: ”CT is a problem solving process that includes (but is not limited to) the following characteristics: formulating problems in a way that enables us to use a computer and other tools to help solve them; logically organizing and analyzing data; representing data through abstractions such as models and simulations; automating solutions through algorithmic thinking (a series of ordered steps); identifying, analyzing, and implementing possible solutions with the goal of achieving the most efficient and effective combination of steps and resources; generalizing and transferring this problem solving process to a wide variety of problems.” (CSTA and ISTE, 2011, para 2.)

ABSTRACT

Author: Shakdher, Arjun. M.S.
Institution: Purdue University
Degree Received: May 2019
Title: Collaborative Platform for Computational Thinking Assessment
Major Professor: Alka Harriger

Computational Thinking (CT) is an integral process of thinking in humans that allows them to solve complex problems efficiently and effectively by breaking down a problem in smaller parts and using abstraction to create generalizable solutions. While the term CT has gained a lot of popularity in current education and research, there is still considerable ambiguity when it comes to defining exactly what CT encompasses. Since the definition and characteristics that make up CT vary so much, it is extremely difficult to measure CT in people. This thesis explains how different industry experts and organizations view CT and describes the importance of developing and integrating such a method of thinking in everyone, not just computer science professionals. The literature review also includes a comprehensive analysis of different tests and tools created to measure CT in people. This study proposes a web-based CT assessment collaborative tool that can be an effective instrument for teachers in assessing CT skills in students who are a part of the Teaching Engineering Concepts to Harness Future Innovators and Technologists (TECHFIT) program funded through NSF DRL-1312215 and NSF DRL-1640178. The vision of this tool is to become a go-to platform for CT assessment where questions collaborated by experts can be used to reliably assess the CT skills of anyone interested in measuring them.

CHAPTER 1. INTRODUCTION

The idea of Computational Thinking (CT) is not recent and has been in practice since 1980 when Papert (1980) used it for his work on Logo, an educational programming language. Wing in her seminal article on CT stated that CT includes ”solving problems, designing systems and understanding human behavior, by drawing on the concepts fundamental to computer science” (Wing, 2006, p. 33). The gain in popularity and demand for incorporating CT in schools all over the globe after Jeanette Wing’s article in 2006 has led to the problem of finding an effective way of measuring these skills once taught (Wing, 2006). However, many different computing experts and organizations define CT in a variety of ways.

There is a need to demystify the term and find a solution that can help effectively and consistently measure CT skills. The first part of this two-step study focused on creating an online web-based platform where computing educators can collaborate, create, rate and review questions to assess CT skills. The development of the platform built on a rudimentary CT tool which was a web-based tool where contributors could add, edit and delete questions that were being stored in a database. It also allowed CT experts to rate the questions based on their alignment to a certain CT concept, habit or practice adding to their reliability. This tool as it stood was insufficient to assess CT skills because of the lack of a quizzing system, absence of a platform where teachers could administer quizzes and students would take them, an intelligent feedback mechanism and the absence of pre-validated questions, The second part of this study sought to determine if the new solution was effective in CT measurement by statistically analyzing and comparing the survey results captured through a sample of middle-school teachers involved in the CT curriculum that is an essential part of a national Science, Technology, Engineering and Mathematics (STEM) program called Teaching Engineering Concepts to Harness Future Innovators and Technologists (TECHFIT).

This section introduces the focus of this thesis; explains why the chosen research problem is significant; describes the various definitions, assumptions, limitations and delimitations; and marks the scope of the study.

1.1 Research Question

What are TECHFIT teachers' perceptions of the effectiveness of the proposed CT system in identifying their students' potential strengths and weaknesses in the programming subset of CT?

1.2 Significance

The purpose of this thesis is to create an online platform that can effectively assess CT skills in people. CT has become a buzzword in the technology education industry and the broader computing community has started acknowledging and acting on it (Barba, 2016). The National Research Council (NRC) stated that a workforce that possesses CT skills makes the United States more competent in the work economy (NRC et al., 2010). CT is being considered as the backbone and core of all STEM disciplines (Henderson, Cortina, & Wing, 2007; Weintrop & Wilensky, 2015).

Hunt and Riley (2014) characterize CT as the process of thinking and reasoning that computer scientists use. Aho (2012) states that CT is a way of thinking and formulating problems so "their solutions can be represented as computational steps and algorithms" (Aho, 2012, p. 832). CT is also defined as "the ability to think with the computer-as-tool" (Berland & Wilensky, 2015, p. 630). Organizational institutions like the Computer Science Teachers Association (CSTA) and the International Society for Technology in Education (ISTE) look at CT as a way of solving complex problems by organizing and analyzing data, automating solutions and generalizing the problem-solving process (CSTA and ISTE, 2011, para. 2). Educational definitions focus on the key elements that comprise CT like abstraction, logic, algorithms, decomposition ad evaluation (Barefoot, 2014).

CT skills are being taught to students at an early age using different approaches including programming challenges, robotics, and game design (Atmatzidou & Demetriadis, 2016; Basawapatna, Koh, Repenning, Webb, & Marshall, 2011). Researchers have also developed and leveraged in different ways to assess CT skills. Many automatic tools that measure CT while developing or programming have been developed but traditional methods like quizzes and manual tests are still widely used (Werner, Denner, Campe, & Kawamoto, 2012).

Literature has shown a number of experts characterizing CT in a variety of ways, depending on the industry they come from. Different projects, classes and programs assess CT differently as well (Atmatzidou & Demetriadis, 2016; Cetin, 2016; Denner, Werner, Campe, & Ortiz, 2014; Israel, Pearson, Tapia, Wherfel, & Reese, 2015; Werner et al., 2012). There is a lack of a generic platform where users interested in assessing their CT skills can take a test and measure it.

CT has been called the 'new basic requirement' (Snapp, n.d.) that has resulted in a surge in the number of jobs that require CT skills. The requirement of manual or routine cognitive jobs has gone down in the past three decades (I. Lee, Martin, & Apone, 2014), and it is suggested that our current and future workforce become skilled in CT as CT will become the core for future job opportunities (*Computational Thinking Will Be Vital For The Future Job Market*, 2018). CT has also been predicted to become one of the skills required for the top jobs in 2025 (Moran & Moran, 2016). It was predicted that jobs that benefit from computing skills would increase to 4.4 million by 2017 (Yadav, Hong, & Stephenson, 2016). Emphasis has been put to apply CT in K-12 because of future job opportunities (*Computational Thinking for Every Student*, n.d.; *Engaging Students Through Computational Thinking*, 2018; Jennings, 2018). Many universities and corporations as described in the literature review are focusing on CT. However, a common, convenient and readily available platform for the assessment of these skills is absent.

More specifically, national programs like TECHFIT have faced difficulty assessing CT skills of participating students, with the TECHFIT leaders attributing it to the lack of a common, widely accepted definition and the absence of a viable tool to assess it. Hence, there is a need for a platform for assessing CT. This mixed-methods study will show how an online collaborative, cross-platform, web-based system can be used to consistently and effectively assess CT skills in geographically dispersed people following the same curriculum for TECHFIT, a national STEM program.

1.3 Scope

The scope of this study is to create an online, web-based, collaborative platform where middle-school teachers participating in the TECHFIT curriculum can assess their students' CT skills through quizzes. Teachers would be able to track their students' progress as to how their CT skills have improved over time. The standardization of the tests would be done by creating a collaborative environment where computing educators can create tests and peer review their legitimacy by rating quality questions with respect to different dimensions of CT they address including different concepts, habits, and perspectives. The system would provide insights on areas for improvement to hone their CT skills. The study then evaluates if the new system is effective in assessing CT skills using a survey analysis. This study is limited to geographically dispersed middle-school teachers participating in the TECHFIT curriculum from 2014-2019. The scope is also limited to the programming subset of CT.

1.4 Assumptions

The assumptions taken for the research conducted are as follows:

- All participants will answer the survey questions in an honest, candid and unbiased way to the best of their knowledge.

- All participants of the study will use the proposed system via a stable Internet connection reducing the variation in user experience.
- The subjects participating in the survey would be comfortable contacting the researcher via email, phone, text or video conferencing if they encounter any problems.
- The number of participants in the study would be adequate to draw conclusions from.

1.5 Limitations

The limitations of this study are as follows:

- Because the researcher will not be working with the subjects in person, the subjects may not be able to access the system sufficiently and understand its usefulness.
- Participants may drop out at any time, which directly impacts sample size.
- Participants may be unable to accurately share and describe their experience in the surveys.
- The current system is limited to only using multiple choice questions (MCQ) to assess CT skills for quizzing.
- CT habits cannot be measured using MCQs, so they are not measured by the system.
- The results may be limited by the small sample size of participants.

1.6 Delimitations

This research is performed acknowledging the following delimitations:

- The study participants are middle-school teachers in STEM disciplines from the TECHFIT project.

- The proposed system is a web-based solution that requires an active Internet connection.

1.7 Summary

This chapter provided the scope, significance, research question, assumptions, limitations, delimitations, definitions, and other background information for the research study.

CHAPTER 2. REVIEW OF LITERATURE

The idea of Computational Thinking (CT) is not recent and has been in practice since 1980 when Papert (1980) used it for his work on Logo, an educational programming language. These early references of the principal of CT by Papert and Harel (1991) emphasized programming and procedural thinking. The term CT began to gain broad appeal after it was used in 2006 in an article by Wing (2006) presenting her viewpoint on CT. She stated that CT includes ”solving problems, designing systems and understanding human behavior, by drawing on the concepts fundamental to computer science” (Wing, 2006, p. 33). CT involves the use of separation of concerns, being able to choose the best approach and model to solve the problem, using invariants to describe a system concisely, and making our solution modular so that it can be decoupled and re-used when needed (Wing, 2006). She further explained that the process of thinking computationally involves heuristic reasoning to break down complex problems into smaller ones, analyzing their feasibility and coming up with possible optimized solutions (Wing, 2006). Solving a problem efficiently may sometimes include modifying or transforming the problem to a form that is understandable and feasible, possibly by reduction, embedding, transformation, or simulation (Wing, 2006).

2.1 Characteristics of CT

This first definition has evolved and has been redefined numerous times by a number of people but still has not reached a consensus. A few years later, Wing (2010) added that CT ”is the thought processes involved in formulating problems and their solutions so that solutions are represented in a form that can be effectively carried out by an information processing agent” (Wing, 2010, para. 2). Wing (2010) states the solutions that we come up with the need to be able to be carried out by a combination of humans and machines. Other researchers argue that CT also encompasses data collection, organization, analysis, automation, efficiency and generalization (Barr, Harrison, & Conery, 2011, p. 21). Bers, Flannery, Kazakoff, and Sullivan (2014) include abstraction, generalization, and trial debugging as part of the description of CT.

CT has been closely associated with the academic discipline of Computer Science though it is not the same. It is linked to programming in the following definition, ”students using computers to model their ideas and develop programs” (Israel et al., 2015, p. 264). It has been characterized as the process of thinking and reasoning that computer scientists use (Hunt & Riley, 2014). Aho (2012) characterizes CT by simplifying the definition stating that CT is a way of thinking and formulating problems so ”their solutions can be represented as computational steps and algorithms” (Aho, 2012, p. 832). Additionally, CT has also been defined as ”the ability to think with the computer-as-tool” (Berland & Wilensky, 2015, p. 630).

Besides these general definitions, operational organizations around the world have also coined definitions for CT. The aim of such a definition is to help CT educators teach by providing them with a framework. The Computer Science Teachers Association (CSTA) and the International Society for Technology in Education (ISTE) state that:

> CT is a problem-solving process that includes (but is not limited to) the following characteristics: formulating problems in a way that enables us to use a computer and other tools to help solve them; logically organizing and analyzing data; representing data through abstractions such as models and simulations; automating solutions through algorithmic thinking (a series of ordered steps); identifying, analyzing, and implementing possible solutions with the goal of achieving the most efficient and effective combination of steps and resources; generalizing and transferring this problem-solving process to a wide variety of problems. (CSTA and ISTE, 2011, para. 2)

More educational definitions of CT focus on key concepts or dimensions that are part of an entire CT framework. Computing at School (CAS) characterizes CT through six different concepts (abstraction, logic, algorithms, decomposition, and evaluation) and five distinct approaches (tinkering, creating, debugging, persevering, and collaborating) for teaching in the classroom (Barefoot, 2014). Brennan and Resnick (2012) define a similar framework comprising three dimensions namely Computational Concepts, Computational Practices, and Computational Perspectives (Brennan & Resnick, 2012, p. 1). The National Research Council (NRC) characterizes CT by five crucial elements:

- hypothesis testing - testing systematically if the hypothesized statement is valid or not.
- data management - collect and manage the data needed for solving the problem.
- parallelism - being able to come up with a computationally feasible solution that can be sped up by executing tasks in parallel.
- abstraction - solving a problem in layers.
- debugging - being able to find out why something is not working the way it is expected to and being able to diagnose the problem. (NRC et al., 2010, p. 40)

Similarly, Anderson (2016) described CT using the following components: Problem decomposition, pattern recognition, abstraction, algorithmic design for solutions, evaluation of the solution. Several courses that are being taught in high schools use their own definition of CT. One such course, that is created by the College Board and the National Science Foundation (NSF) focuses on the practices element of CT and comprises the seven pillars of computing, namely (Grover, Pea, & Cooper, 2015, p. 209):

- Computing is a creative human activity
- Abstraction reduces information and detail to focus on concepts relevant to understanding and solving problems
- Data and information facilitate the creation of knowledge

- Algorithms are tools for developing and expressing solutions to computational problems
- Programming is a creative process that produces computational artifacts
- Digital devices, systems, and the networks that interconnect them enable and foster computational approaches to solving problems
- Computing enables innovation in other fields, including science, social science, humanities, arts, medicine, engineering, and business. (Grover et al., 2015, p. 209)

2.1.1 Relationship of CT with other disciplines

There is also mass confusion among people who mistake CT with other thinking methodologies such as mathematical, engineering, design, systems and also with the field of Computer Science (CS).

2.1.1.1 CT and mathematical thinking

Figure 2.1 describes the key elements common between CT and mathematical thinking. The main common feature in both is problem-solving (Wing, 2008). Other common elements from both fields including modeling, data analysis, statistics, and probability show that there is a strong overlap of the two methodologies even though they are different as a whole.

2.1.1.2 CT and engineering thinking

As stated by Pawley (2009), engineering thinking involves ”applied science and math, solving problems, and making things” (Pawley, 2009, p. 310). CT is similar to engineering thinking in the sense that both focus on problem-solving and understanding how complex systems work as a whole (Wing, 2010).

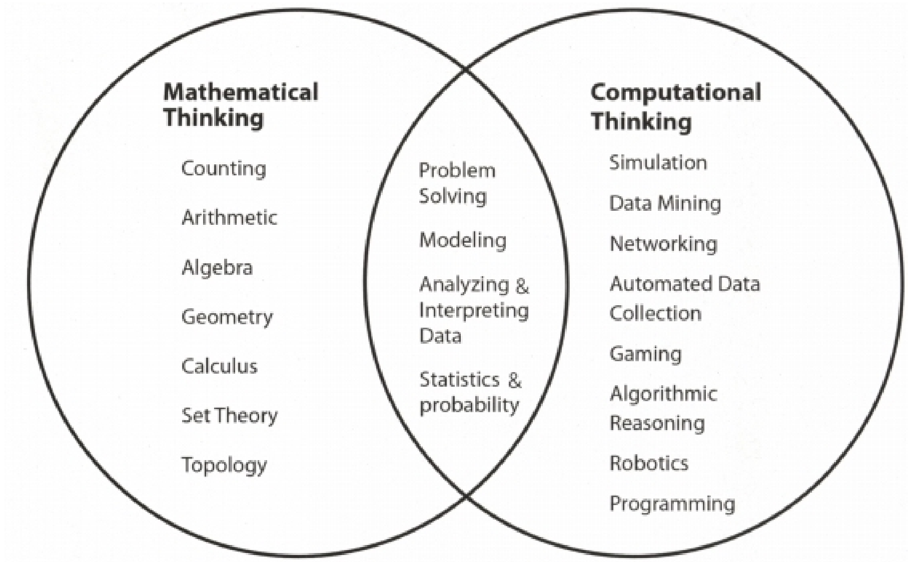


Figure 2.1. Union of Mathematical and Computational Thinking. Reprinted from ”Computational thinking in high school science classrooms”, by Sneider, Stephenson, Schafer, and Flick (2014, p. 53)

2.1.1.3 CT and design thinking

Design thinking, on the other hand, requires one to concentrate on solving problems by focusing on the design and by thinking like a designer (Razzouk & Shute, 2012). Such a way might be limited by physical constraints but unlike design thinking, CT is not constrained and can be used to solve both theoretical as well as practical questions.

2.1.1.4 CT and systems thinking

Systems thinking is a way of treating a system as a whole where the different components are working in tandem. Shute, Sun, and Asbell-Clarke (2017) mention that even though both these fields involve careful system modeling, CT is a much broader term than systems thinking.

2.1.1.5 CT and computer science (CS)

The difference gets unclear when it comes to CS. As described by Wing (2006), ”computer science is the study of computation - what can be computed and how to compute it” (Wing, 2006, p. 34). Just like CS, CT involves thinking recursively, analyzing the benefits and dangers of aliasing, repercussions of not using unique identifiers and weighing in the cost of different ways of doing the same thing. Above all, a focus on simplicity, elegance, and aesthetics is crucial in creating a good solution (Wing, 2006). Thinking the way computer scientists think is a method of thinking in various levels of abstraction. Ioannidou, Bennett, Repenning, Koh, and Basawapatna (2011) warn that there is a stark difference between CT and CS skills, but programming knowledge is a plus point when it comes to thinking computationally. As described by NRC et al. (2010); Wing (2006), programming is a tiny subset of CT, which is much broader as it is a way of thinking that is extended to even daily activities and not just computer science problems. This is represented in figure 2.2.

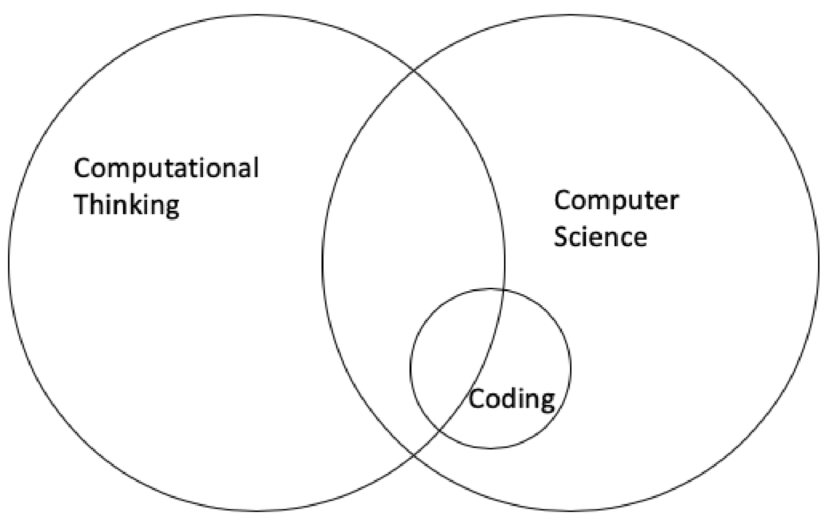


Figure 2.2. Computational Thinking and Computer Science. Adapted from ”Advancing Computational Thinking Across K-12 Education”, by *Advancing Computational Thinking Across K-12 Education* (2019)

CT has also been characterized by favoring:

- conceptualizing, instead of programming - visualizing at various levels of abstraction on how to solve the problem.
- fundamental, instead of rote skill - thinking elementary and not in a mechanical way is a skill all humans should possess to function in modern society.
- the way humans think, instead of the way computers do - thinking in a creative/imaginative way and not being dull/boring with the solutions.
- combining mathematical and engineering thinking - taking the best of our mathematical foundation and applying our engineering skills to create highly capable systems.
- ideas instead of artifacts - not concentrating on just producing the software/hardware artifacts but focusing on the computational concepts using to approach and solve the problem.
- its applicability to everyone - it needs to be so integral that it disappears as an explicit philosophy. (Wing, 2006, p. 35)

CT uses similar fundamentals as CS like ”abstraction, debugging, remixing and iteration” to tackle problems (Brennan & Resnick, 2012; Ioannidou et al., 2011; Wing, 2008). Wing (2010) stresses the importance of abstraction by stating that ”Designing efficient algorithms inherently involves designing abstract data types. Abstraction gives us the power to scale and deal with complexity.” (Wing, 2010, para. 5). Wing (2010) warns that our computational systems are often limited by the computing power of the physical hardware, so our solutions must take into consideration the worst-cases with respect to the potential failures and the unpredictability of the real world (Wing, 2010, p. 1). On the other hand, since the software is not constrained directly by physical reality, we can build virtual worlds that have no limitations (Wing, 2010, p. 2).

The common ground where most authors agree on is the fact that CT is a transferable skill and not just for computing professionals. Wing, in her seminal article on CT, stated that CT ”represents a universally acceptable attitude and skill set everyone, not just computer scientists, would be eager to learn and use.” (Wing, 2006, p. 33). Breaking down a large complex problem and solving it piece by piece using concepts like abstraction, reduction, pattern-finding, and transformation is something that is beneficial to everyone in their day to day life (Wing, 2006). The elements of CT that include prevention, protection, and recovery from worst-case scenarios are principles that are applicable to every professional job one could perform (Wing, 2006). CT is making a large impact on industries outside of Science and Engineering (Wing, 2010). Fields like ”algorithmic medicine, computational archaeology, computational economics, computational finance, computational law, computational social science, and digital humanities” are some examples that are benefiting from CT (Wing, 2010, p. 2).

2.2 Need for CT Skills

There is a huge, industry-wide need for the development of CT skills, so the new workforce needs to acquire the active problem-solving methodology it entails. The National Research Council (NRC) stated that a workforce that possesses CT skills makes the United States more competitive in the world economy (NRC et al., 2010). The need for mining and analyzing volumetrically and dimensionally large amounts of data we have presently using traditional modeling techniques in combination with modern machine learning has led computation to be recognized as the third pillar of science, besides theory and experimentation (Reed et al., 2005). As a result of the awareness and the importance of CT, it is increasingly gaining popularity in the education industry and is being launched as a classroom program or be integrated into existing classes. In the United Kingdom, the introduction of computer science fundamentals as early as primary school shows the recognition of the value such a method of thinking generates (Brown et al., 2013). CT is being considered as the backbone and core of all STEM disciplines (Henderson et al., 2007; Weintrop & Wilensky, 2015).

2.3 Teaching CT Skills

The goal of teaching CT to young students is to develop their thinking where they can create better mental models that help solve bigger problems with or without the use of technology across a broad range of disciplines. The CS Unplugged project is one such instance where CS and general computing principles were taught to elementary and middle school children without using a computer (Bell, Witten, Fellows, Adams, & McKenzie, 2005). Kim, Kim, and Kim (2013) also created a strategy called Paper and Pencil programming (PPS) which allowed students of a non-CS background to understand CT visually and in a more effective way.

There has been an increased interest in the implementation of CT not just in college research, but, has also expanded to K-12 schools (Wing, 2010). Carnegie Mellon University has held multiple summer workshops of CS4HS which gives the teachers understanding about CT (Wing, 2010). By 2010, Microsoft and Google had helped expand CS4HS to 20 schools in the USA and 14 in Europe, the Middle East, and Africa (Wing, 2010, p. 5). In 2007, The Carnegie Mellon Center for Computational Thinking, that supports research and CT outreach programs, was funded by Microsoft (Wing, 2010). Yu (2014) has proposed that CT be taught to children at an early stage and CT be incorporated into the school's computer related courses. Educators across the globe have tried to leverage different ways of teaching students CT skills through programming, robotics, game design, etc. Since programming and CT are so tightly coupled, coding inherently teaches the elements of CT and is therefore used widely for CT instruction. Scratch, a free programming language created at The Lifelong Kindergarten group in Massachusetts Institute of Technology, is an easy way for young students to pick up programming skills through Scratch's drag and drop block programming. Grover et al. (2015) constructed a seven-week course to teach algorithmic thinking to fifty-four 7th and 8th graders using Scratch. Similarly, students were also given tasks using Alice, a tool that uses building blocks to create 3-dimensional programming objects (Denner et al., 2014). Other traditional languages that have been used to teach programming and computing include Logo and Python (Ahamed et al., 2010; Aiken et al., 2013; Lin & Liu, 2012).

Code.org, a non-profit that aims to expand CS education to minorities has achieved incredible success with their hour of code program. The program aimed at children gives the opportunity to introduce anyone to computing in just one hour via an engaging activity. Backed by large corporations including Amazon, Google, Facebook, Infosys, and Microsoft, they have engaged 15% of students in the world according to their 2018 annual report. Code.org has also been instrumental in enacting policies in many states to require computing coverage in K-12. (*Code.org 2018 Annual Report*, n.d.; Meyers & Huang, 2019)

TECHFIT facilitates a 6-day professional development summer workshop where the teachers from STEM disciples are trained in CT so that they can teach the TECHFIT curriculum to their middle school students through a 10-week, after school class or program. The program involves the use of programming in both Scratch and NanoNavigator. Students are also expected to use physical computing and work together as a group to build an entire exer-game. The skills they use for the program are directly related to what the various definitions of CT entail including abstraction, decomposition, algorithmic design, and pattern matching. The original project ended in Summer 2017, however, another NSF project that built on TECHFIT was funded in Fall 2016 and continues through 2019. The new project continues to teach the TECHFIT curriculum to middle school teachers and seeks to examine differences in student motivation depending on the delivery mode (in-school class or after-school program). The TECHFIT project has grown in size and has involved 165 teachers and 2850 middle-school students since 2013. (Harriger & Harriger, 2017)

The field of robotics is another area where tools are being used to develop CT skills. CT skills were enhanced by Lego Mindstorms by learning decomposition, abstraction, generalization, and automation (Atmatzidou & Demetriadis, 2016). Programming in robotics forces the mind to think of decomposing larger problems and developing algorithms while trying to automate it. The essential skill of debugging is also something that one learns from programming and robotics.

AgentSheets, a tool that teaches CT through game design was used to learn how to animate interactions via programming (Basawapatna et al., 2011). Game design teaches an individual, how to strategize and plan different solutions for solving the problem. Spreading awareness and teaching the concept and principles of CT is just one perspective. Like with most domains that have numerous ways to develop essential skills, having a widely-accepted or peer-reviewed way to assess the domain skills is important.

2.4 Available CT Assessments

There are a number of ways CT has been measured. Different individuals and organizations have come up with various methods of assessing it by either creating tools that capture and analyze computational concepts, practices, and perspectives, or by creating manual tests, questionnaires and quizzes containing questions that measure CT.

One of the main contributions to the CT assessment community is the Fairy Assessment that scores students on computational concepts used while solving problems (Werner et al., 2012). The downside of this approach is that it requires knowledge of Alice, a programming learning environment. Also, this assessment lacks the reliability analysis or validity study (Ketenci, Calandra, Margulieux, & Cohen, n.d.). Scratch, a block-based programming tool is a popular method of assessing CT skills due to its interactive, user-friendly approach to programming (Brennan & Resnick, 2012). Snippets of code from Scratch have been included in tests created by Grover et al. (2015) in the form of multiple choice questions. These questions, in conjunction with other open-ended questions, form the test by Grover et al. (2015). CT has also been measured using automated tools that capture the different computational thinking patterns used while developing games (Koh, , Bennett, & Repenning, 2010). Dr. Scratch is a similar tool that analyzes uploaded Scratch projects and produces feedback about the scope of improving computational skills (Moreno-León & Robles, 2015). It calculates the overall score based on seven CT dimensions: Abstraction and problem decomposition; Parallelism; Logical thinking; Synchronization; Flow control; User interactivity; and Data representation (Moreno-León & Robles, 2015). This is an online web application that is offered to everyone for free and is an excellent tool to collect feedback and learn programming (Moreno-León & Robles, 2015). Seiter and Foreman (2013) developed another Scratch-based assessment called PECT (Progression of Early Computational Thinking). This tool assessed CT skills by using information from Scratch programs and allocated points based on how proficient the test taker was in terms of the concepts, patterns, and perspective facets.

Game development and design have also been used as a platform for assessing CT using pattern analysis when students are developing their games (Basawapatna et al., 2011; Ioannidou et al., 2011). This tool generates a visual report of the student's abilities, also prompting them about their areas of improvement. More traditional and systematic methods include Computational Thinking using Simulation and Modelling (CTSiM), an open-ended learning environment that assesses these skills in people (Basu, Kinnebrew, & Biswas, 2014). Object-oriented programming languages like Java and C++ are very similar to the concepts addressed by CT (Smith, Cypher, & Tesler, 2000). However, they do not prove to be effective for people who are from a non-computer background because they involve an overhead of learning the syntax and semantics. All of the described assessments measure CT skills in a different aspect: some target the computational concepts and practices while others simply assign a score based on programming skills.

Cognitive approaches have also been used correlated with CT to measure the overall intelligence. As suggested by Ambrósio, Xavier, and Georges (2014), CT is related to the following elements of the Cattell-Horn-Carroll (CHC) model of intelligence (McGrew, 2009):

- Fluid reasoning (G_f), defined as: "the use of deliberate and controlled mental operations to solve novel problems that cannot be performed automatically. Mental operations often include drawing inferences, concept formation, classification, generating and testing hypothesis, identifying relations, comprehending implications, problem-solving, extrapolating, and transforming information. Inductive and deductive reasoning are generally considered the hallmark indicators of G_f" (McGrew, 2009, p. 5)

- Visual processing (G_v), defined as ”the ability to generate, store, retrieve, and transform visual images and sensations. G_v abilities are typically measured by tasks (figural or geometric stimuli) that require the perception and transformation of visual shapes, forms, or images and/or tasks that require maintaining spatial orientation with regard to objects that may change or move through space” (McGrew, 2009, p. 5)

- Short-term memory (G_{sm}), defined as ”the ability to apprehend and maintain awareness of a limited number of elements of information in the immediate situation (events that occurred in the last minute or so). A limited-capacity system that loses information quickly through the decay of memory traces, unless an individual activates other cognitive resources to maintain the information in immediate awareness” (McGrew, 2009, p. 5)

Questions and surveys are seen as the most popular method for measuring CT skills, mostly because they can be evaluated quickly. (Atmatzidou & Demetriadis, 2016; Denner et al., 2014). Surveys that capture the understanding of the CT concepts, practices, and perspectives have been designed to assess CT (Cetin, 2016). These surveys, distributed to elementary and middle school students, besides measuring the proficiency in CT concepts, also measure interest in the subject (Jun, Han, Kim, & Lee, 2014; Kim et al., 2013). Interviews have been constructed to understand how developments of CT skills takes place (Cetin, 2016; Israel et al., 2015). However, all of the quizzes are localized and differ in their definition of what constitutes CT.

Many quizzes exist on popular quizzing websites like Chegg and Quizziz that test people on their knowledge of CT. They include knowledge-based questions based on decomposition, pattern recognition, algorithms and evaluating solutions along with definition-based questions (*Computational Thinking*, n.d.; *Computational thinking Quiz*, n.d.; *Digital Literacy Quiz*, 2016; *K3 Computational Thinking Starter Quiz*, 2018). Other quizzes include puzzle-based questions that are more generic (*Computational* *thinking Quiz*, n.d.). Besides these, there are some tools like Quizmaker and Quizly that are capable of tracking the progress that utilize App Inventor for grading. However, all of these sources do not have any validation as they are created by unverified contributors on websites.

The CT quiz, based on Brennan and Resnick's CT framework contains 12 multiple-choice questions (MCQs) based on different CT concepts and practices (Ketenci, Calandra, Margulieux, & Cohen, 2019). The authors have proven this tool to be reliable (alpha=0.82) in measuring students' CT skills. The assessment developed uses CT concepts, practices and perspectives as the CT dimensions. CT concepts include sequences, loops (Figure 2.3), parallelism (Figure 2.4), events, conditionals, operators, and data while programming. CT practices include abstracting and modularizing, reusing and remixing or debugging and testing, and being incremental and iterative. The questions were adapted from a tool called App Inventor, which uses flowchart programming to develop apps for Android OS. The questions were aligned with California K-12 CS framework (Committee et al., 2016). This assessment is a proven, reliable and useful quiz that can be used to measure students' CT skills with the limitation being it is a standalone quiz that measures CT skills using flowchart programming. (Ketenci et al., 2019)

Figure 2.3. Sample question with Loops as the CT concept and Abstraction as the CT practice. Adapted from "Computational Thinking Assessment", by Ketenci et al. (2019)

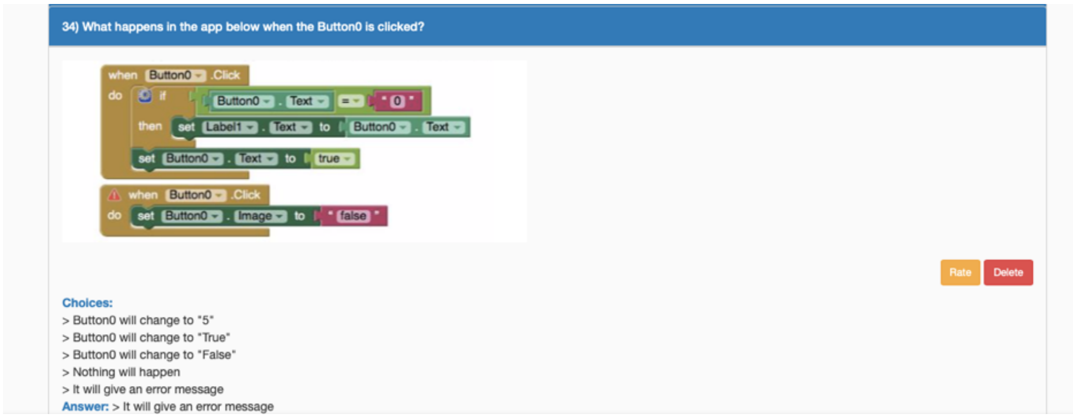


Figure 2.4. Sample question with Parallelism as the CT concept and Abstraction as the CT practice. Adapted from ”Computational Thinking Assessment”, by Ketenci et al. (2019)

The Computational Thinking Test (CTT) is a valid and reliable test created for 8th grade middle-school children. The reliability analysis yielded a Cronbach’s alpha of 0.772 and McDonalds omega of 0.779 showing that it can successfully measure CT skills in the students. The problems created in the test are not programming based but lie in the math and science domain. However, the items in the test are open-ended and hence making it difficult to be graded by a computer making it the biggest limitation of this paper. (Bati, 2018)

Another test, known as the Computational Thinking Test (CTt) is a test created specifically to measure CT skills in middle school students from the 5th grade to 9th grade. The CTt uses the dimensions as outlined by Weintrop. It was reviewed by five experts and after feedback, was revised and tested with 110 8th-graders. This tool was found to be a valid and reliable test that can successfully measure the computational thinking skills in middle school children. This test is generic, based on math and science principles and not completely on programming (even though it contains a few questions with Scratch snippets). (Román-González, Pérez-González, & Jiménez-Fernández, 2017)

The CTt is built on the following CT dimensions (Román-González et al., 2017, p. 681):

- ”Computational concept addressed: each item addresses one or more of the following seven computational concepts, ordered in increasing difficulty: Basic directions and sequences (4 items); Loopsrepeat times (4 items); Loopsrepeat until (4 items); Ifsimple conditional (4 items); If/elsecomplex conditional (4 items); While conditional (4 items); Simple functions (4 items). These computational concepts are aligned with some of the CT framework (Brennan & Resnick, 2012) (Table 2.1) and with the CSTA Computer Science Standards for 7th and 8th grade (CSTA and ISTE, 2011).” (Román-González et al., 2017, p. 681)

- ”Environment-Interface of the item: CTt items are presented in any of the following two environments-interfaces: The Maze (23 items) or The Canvas (5 items). Both interfaces are common in popular sites for learning programming such as Code.org (Kalelioğlu, 2015).” (Román-González et al., 2017, p. 681)

- ”Answer alternatives style: in each item, the response alternatives may be presented in any of these two styles: Visual arrows (8 items) or Visual blocks (20 items). Both styles are also common in popular sites for learning programming such as Code.org (Kalelioğlu, 2015).” (Román-González et al., 2017, p. 681)

- ”Existence or non-existence of nesting: depending on whether the item solution involves a script with (19 items) or without (9 items) nesting computational concepts (a concept embedded in another to a higher hierarchy level) (Mühling, Ruf, & Hubwieser, 2015). ”(Román-González et al., 2017, p. 681)

- ”Required task: depending on which of the following cognitive tasks is required for solving the item: Sequencing: the student must sequence, stating in an orderly manner, a set of commands (14 items); Completion: the student must complete an incomplete given set of commands (9 items); Debugging: the student must debug an incorrect given set of commands (5 items). This dimension is partially aligned with the aforementioned computational practices from the CT framework (Brennan & Resnick, 2012) (Table 2.1).” (Román-González et al., 2017, p. 681)

Table 2.1. Crosstab intersecting CT framework with the sampling domain of the Computational Thinking Test (CTt)

Dimension	Component	CTt Sampling Domain
Concepts	Sequences	Yes
Concepts	Loops	Yes
Concepts	Events	No
Concepts	Parallelism	No
Concepts	Conditionals	Yes
Concepts	Operators	Yes
Concepts	Data	No
Practices	Experimenting and Iterating	No
Practices	Testing and Debugging	Partly
Practices	Reusing and Remixing	Partly
Practices	Abstratcing and Modularizing	Partly

Example items from the CTt including which CT concept and practice they address are shown in Figure 2.5 and 2.6.

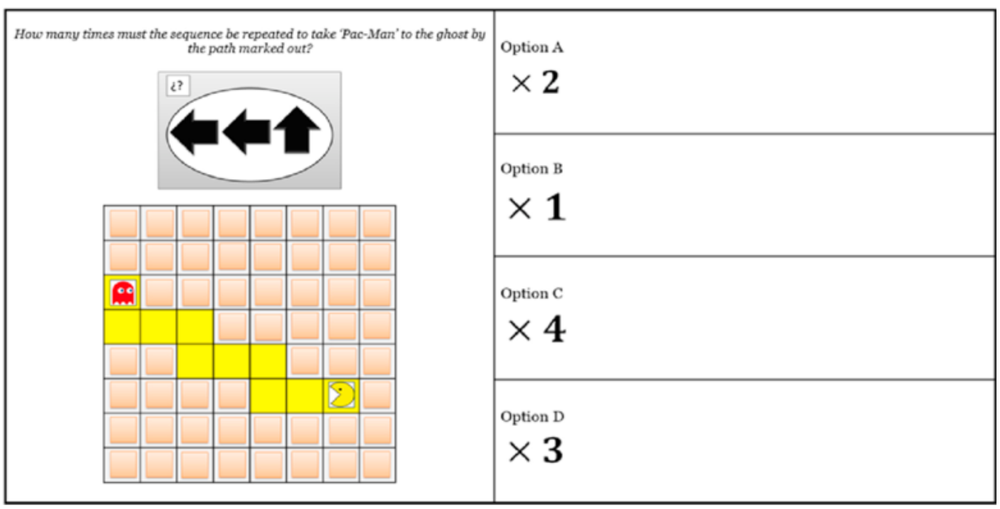


Figure 2.5. Computational Thinking Test (CTt), item 6: loops; completion. Reprinted from ”Which cognitive abilities underlie computational thinking? Criterion validity of the Computational Thinking Test”, by Román-González et al. (2017, p. 681)

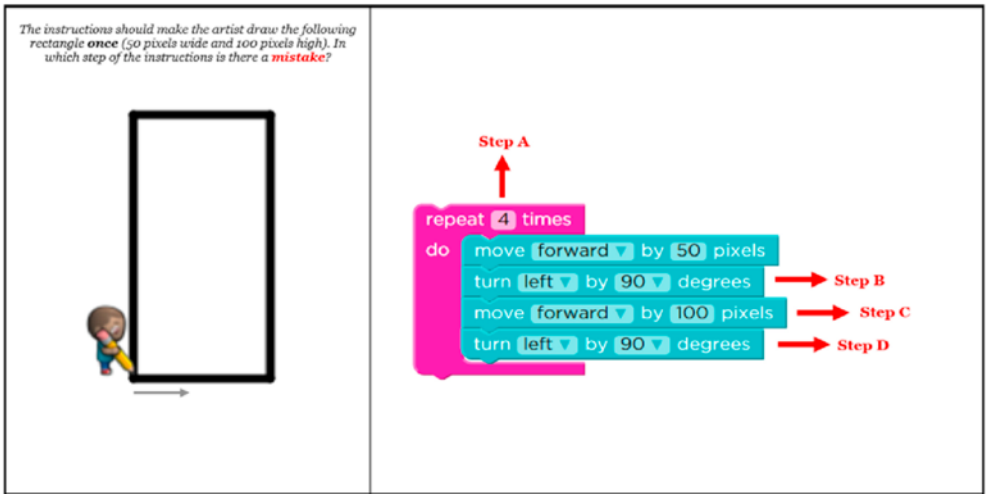


Figure 2.6. Computational Thinking Test (CTt), item 7: loops; debugging. Reprinted from ”Which cognitive abilities underlie computational thinking? Criterion validity of the Computational Thinking Test”, by Román-González et al. (2017, p. 681)

The CTt has proven to be effective as ”it can be administered in pretest conditions to measure the initial development level of CT in students without prior programming experience from 5th to 10th grade; it can be collectively administered so it could be used in massive screenings and early detection of students with high abilities (or special needs) for programming tasks; it can be utilized for collecting quantitative data in pre-post evaluations of the efficacy of curricula or programs aimed at fostering CT, which would be a desirable practice versus the qualitative approach that has been mostly used in the literature so far (Lye & Koh, 2014); and it could be used along academic and professional guidance processes towards STEM disciplines”. (Román-González et al., 2017, p. 687)

The reliability and predictive validity tests of the CTt have yielded promising results. The CTt has predictive validity with respect to academic performance and grade-point average (GPA) (Román-González, Pérez-González, Moreno-León, & Robles, 2018). CTt also has predictive validity with respect to coding achievement in middle school and the predictive validity to distinguish between computational top and regular thinkers (Román-González et al., 2018).

The main limitation of the CTt is that it is a static test and the authors recommend using it with other assessments as a whole. Another limitation of this study is that CTt is focused on CT concepts and partly on practices. Combining this assessment with other assessments could alleviate the limitation. (Román-González et al., 2017)

The CT quiz (Ketenci et al., 2019) and the CTt (Román-González et al., 2018) are two good choices to populate the CT assessment system with. The only downside is that independently, they are quizzes and not a complete assessment system. The TECHFIT program has had students drop-out and schools not completing the curriculum within the time frame. TECHFIT researchers suspect the lack of a generic platform for CT assessment to be a possible factor contributing to the problem. Such a platform could help alleviate the problem of localized CT assessments by the collaboration of different validated quizzes. This could also help the TECHFIT leaders gauge the effectiveness of the curriculum coverage by participating schools and help each teacher identify areas of strength and weakness in their team, so they can provide additional feedback to their students. (Harriger & Harriger, 2017)

2.5 Survey Instrument

Perceived effectiveness of software in the context of an information system has been defined as ”a multidimensional concept that includes user perceptions of the usefulness and ease of use” (Atkinson & Yeoh, 2008, p. 226). In 1989, Davis created the Technology Acceptance Model (TAM) that measures how well the software is accepted based on perceived usefulness and perceived ease of use (Davis, 1989). The TAM was voted as the ”most widely applied theoretical model in the IS field” (Y. Lee, Kozar, & Larsen, 2003). Perceived usefulness has been defined as ”the degree to which a person believes that using a particular system would enhance his or her job performance” (Davis, 1989, p. 320). Davis characterized perceived ease of use as ”the degree to which a person believes that using a particular technology would be free from effort” (Davis, 1989, p. 320). User satisfaction has been noted as a key factor in an Information System (IS) being successful.

The System Usability Scale (SUS) is proven to be a reliable usability scale that can quickly measure a system's usability. It is robust, reliable and correlates well with other measures of usability (Brooke et al., 1996).

The Educational Software/Website Effectiveness Survey (TESWES) is a useful survey instrument can be used by educational instructors in evaluating if the software under review is educationally appropriate and sound for instruction and behavior management. (Furner & Daigle, 2004)

The Perceived Usefulness and Ease of Use (PUEU) is a well-tested survey instrument created by Davis for reliably evaluating user acceptance of a system by evaluating its perceived usefulness and perceived ease of use. (Davis, 1989)

The Web based Learning Environment Instrument (WEBLEI) is a survey instrument created to capture the perceptions of a web-based learning environment (Chang, 1999). It has been proven valid and reliable in capturing perceptions in a study done by Ozkok (ÖZKÖK, 2013).

2.6 Summary

Although a substantial amount of work has been done since Wing (2006) popularized the term CT, the definitions and methods of assessment still vary. Most of the assessments created to measure CT are specific to the group and type of work they were created for. Demystification of CT will happen when a consensus regarding its concept is reached. Further, there is a need for researching and developing a platform where computing educators can collaborate and create peer-reviewed tests for measuring CT skills. To the best of the author's knowledge and literature review, such a comprehensive web-based platform that gives the teachers the ability to build quizzes from multiple valid and reliable sources and gain meaningful feedback on their student's reports is not available. The proposed tool is a quick, reliable and consistent metric that could be used to assess different people on the basis of their generic CT ability, no matter which domain they want to study or work in. The insights produced from the tool would give meaningful feedback to the users which could help hone their CT skills.

CHAPTER 3. METHODOLOGY AND IMPLEMENTATION

This section provides the methodology and the framework used for building the collaborative platform for CT assessment. It also details the sample, population set, instrumentation, and data analyses that will be used to explore the effectiveness of the proposed platform.

This mixed-methods study is performed in two parts. First, the proposed tool was developed. After the development of the tool, its effectiveness as a CT assessment tool was evaluated using surveys sent out to a chosen sample containing both rating and open-ended questions. The information gathered from these surveys was used to evaluate the research question described earlier.

3.1 Population

The population for the study consists of geographically distributed teachers following the curriculum for TECHFIT, a national STEM program.

3.2 Sample

Using Convenience Sampling, 146 teachers from different middle-schools in STEM disciplines who had participated in the TECHFIT curriculum from 2014-2019 were sent an invitation to use the tool online along with a survey link. 12 teachers responded to the questionnaire out of which 4 had incomplete responses to the MCQs. 7 respondents had complete responses to the open-ended questions.

3.3 Instrumentation

This section describes the instrumentation used to build the proposed CT platform. It contains the software and hardware required to build along with information on where the tool is deployed. It also describes the survey instrumentation used to evaluate the study's research question.

3.3.1 Development of the tool

The instrumentation for this study included creating a responsive web application that will be the collaborative tool for CT assessment. The following programming languages and frameworks were used for the development of the application due to their popularity and accessibility at Purdue University, West Lafayette, Indiana, USA:

- Back-end development - C# 7.0
- Front-end development - ASP.NET
- Frameworks - Bootstrap
- Database - Microsoft SQL Server 2014
- Editor - Microsoft Visual Studio Code Version 1.28

3.3.2 Population of question repository

After consulting with Prof. Alka Harriger, TECHFIT Program Manager, pre-validated questions that have proven to be reliable in assessing CT skills were selected to populate the question bank for this platform. Currently, the two main sources for questions are the CTt (Román-González et al., 2018) and Computational Thinking quiz (Ketenci et al., n.d.). Permission from the authors of both the sources was taken before being used in the tool (Appendix C).

3.3.3 Deployment of the tool

The web application is hosted on the Engineering Computer Network (ECN) servers at Purdue University, West Lafayette, Indiana, USA. The URL for accessing the website is: https://ecniisdev.ecn.purdue.edu/dev.techfit/CTQuestions

The hardware configuration of the servers is shown in Table-3.1.

Table 3.1. Hardware specification of the ECN server.

Specification	Measure
CPU	2x Intel Xeon E5-2623 v3 @ 3.00 GHz
Memory per node	32GB
Operating System	Windows 2008 R2

3.3.4 Evaluating the research question

The research question described in the previous sections was tested, and the created tool was evaluated using the survey which was sent to everyone in the chosen convenience sample. The survey contained a mixture of ranking-based questions using Likert scales and open-ended questions asking about the effectiveness of the tool in assessing CT skills. The survey was created and administered through Qualtrics. The data sources are described in the following section.

3.3.5 Survey instrument

To study the perceived effectiveness of the proposed tool, a questionnaire containing multiple pre-validated instruments was created after an extensive review of the literature. Questions from the following pre-validated reliable instruments were used:

- 8 questions from the SUS measuring perceived system usability (Brooke et al., 1996)
- 8 questions from TESWES measuring perceived age appropriateness and perceived motivation level (Furner & Daigle, 2004)

- 2 questions from PUEU measuring perceived usefulness (Davis, 1989)
- 4 questions from the WEBLEI measuring perceived user satisfaction and perceived content appropriateness (Chang, 1999)

A total of 25 questions with 22 MCQs and 3 open-ended questions were included in the final survey. Some of the words in the existing questions were replaced with the tool's name. For example, 'this website' in SUS was replaced with 'this web tool'. It has been proven that doing minor modifications to make the questions more specific is generally a good practice that makes the questionnaire more understandable to the respondent and does not affect the reliability or validity of the survey (Sauro, 2011). The complete survey instrument can be found in Appendix A.

Generally, questionnaires use multiple items to study a common focal item on which they wish to gain insight (Robinson, 2018). For this study, the common item was perceived effectiveness, which literature defines as the sum of many sub-scales. Reliability and validity of a scale are very important features where the former is a prerequisite for the latter (Kline, 2013). While there are several methods to gauge the reliability of an instrument, Cronbach's alpha coefficient (Cronbach, 1951) is the most popular method for calculating the reliability of a created questionnaire (Robinson, 2018). While there is considerable debate about the accepted alpha level, Cortina (1993) suggests that 0.75 is the widely-accepted convention to call the questionnaire reliable.

The reliability testing of the created questionnaire was performed using Cronbach's alpha test. A prerequisite of the test is that negatively worded questions should be re-coded for the analysis (Field, 2009). For example, "using a traditional 5-point rating scale, the reverse coding would proceed as follows: strongly disagree (1 to 5), disagree (2 to 4), neutral (3 to 3), agree (4 to 2), strongly agree (5 to 1)" (Robinson, 2018, p. 748). The questionnaire had 4 questions that were negatively worded (Q10, Q12, Q14, and Q16). These were reverse-coded using R.

The overall reliability of the questionnaire was α=0.94 (figure 3.1). Though there are a number of factors like the number of items and sample size that can drive up the alpha score, a score of 0.8 or higher is assumed to be safe (Field, 2009). This implies that the questionnaire is reliable. Another step included finding out how each of the questions contributed to the overall reliability of the questionnaire (Field, 2009). The raw alpha score for all items remains considerably consistent (0.03 to 0.95), which implies items do not need to be removed to make the questionnaire reliable.

```
 raw_alpha std.alpha G6(smc) average_r S/N   ase mean   sd median_r
      0.94      0.94       1      0.43  16 0.021  4.3 0.66     0.48

 lower alpha upper     95% confidence boundaries
0.9 0.94 0.98

 Reliability if an item is dropped:
      raw_alpha std.alpha G6(smc) average_r S/N var.r med.r
Q1         0.94      0.94       1      0.44  16 0.118  0.48
Q2         0.94      0.94       1      0.43  16 0.122  0.47
Q3         0.94      0.94       1      0.42  15 0.113  0.46
Q4         0.94      0.94       1      0.43  16 0.123  0.48
Q5         0.94      0.94       1      0.43  16 0.119  0.48
Q6         0.94      0.94       1      0.43  16 0.117  0.48
Q7         0.94      0.95       1      0.45  17 0.118  0.49
Q8         0.94      0.94       1      0.44  16 0.122  0.49
Q9         0.93      0.94       1      0.41  14 0.115  0.45
Q10-       0.94      0.94       1      0.43  16 0.121  0.47
Q11        0.95      0.94       1      0.45  17 0.114  0.49
Q12-       0.95      0.95       1      0.47  19 0.098  0.50
Q13        0.94      0.94       1      0.42  15 0.113  0.46
Q14-       0.95      0.95       1      0.47  18 0.104  0.49
Q15        0.94      0.94       1      0.42  15 0.118  0.46
Q16-       0.94      0.94       1      0.41  14 0.115  0.46
Q17        0.94      0.94       1      0.43  16 0.114  0.47
Q18        0.94      0.94       1      0.42  15 0.113  0.47
Q19        0.93      0.93       1      0.41  14 0.115  0.45
Q20        0.94      0.94       1      0.41  15 0.118  0.46
Q21        0.94      0.94       1      0.41  15 0.116  0.46
Q22        0.94      0.94       1      0.41  15 0.116  0.45
```

Figure 3.1. Overall Cronbach's α

Table 3.2. Cronbach's α for the sub-scales

Sub-scale	α
Perceived Usability	0.80
Perceived Usefulness	0.95
Perceived Content Appropriateness	0.92

To calculate subscale reliability, the entire questionnaire was broken down into parts that had Likert scale items (usability, usefulness and content appropriateness). Age appropriateness and satisfaction were Likert-type data consisting of one item and hence not applicable for the alpha test. The following presents the results of the alpha test for each of the subscales (table 3.2):

- The α score for perceived usability was 0.80, which indicates high reliability.
- The α score for perceived usefulness was 0.95, which indicated high reliability.
- The α score for perceived content appropriateness was 0.92, which indicates high reliability.

It is important to note that the reliability testing is limited by this study's small sample size. It is stated that given the alpha cutoff, one can generally accept that the questionnaire is reliable but cannot claim its validity (Rattray & Jones, 2007). Validating also involves comparing the instrument with others that measure the same thing (Rattray & Jones, 2007). It is difficult to find or create multiple surveys that do the same thing (O'keefe, 2002) and to the best of the author's research, a survey on the perceived effectiveness of software did not exist. It should be noted, however, that the questions used in this survey are not created from scratch and are borrowed from other pre-validated sources. Performing comprehensive validity testing with a much larger sample size is recommended for future work.

3.4 Data Collection and Sources

The data was collected from both ranking-based and open-ended questions presented to volunteering participants in our sample. The questions assessed if the new tool was perceived to be effective to TECHFIT teachers in identifying their students potential strengths and weaknesses in the programming subset of CT. The questions in the survey ask the participants to rate the tool on the basis of its usefulness, usability, content appropriateness, age appropriateness and satisfaction along with other open-ended questions about the tool. The survey was sent to TECHFIT teachers from 2014 to 2019 with a request for their acceptance in reviewing the proposed CT assessment system. Another email was sent with a link to the tool as well as the above questionnaire to all the teachers who accepted to participate in the study.

No identifiable data was collected. An email was sent to the sample describing the goal of the study, how their voluntary participation can contribute to it and their ability to opt-out at any time. The survey was created and administered using Qualtrics. Only the researcher had access to the data and all the data collected is kept confidential. The study was exempt from the university Institutional Review Board (IRB). The Purdue IRB exempt information can be found in Appendix B.

3.5 CT Tool

The developed web tool is a platform for teachers to log on and assess their students' CT skills. Students can take validated CT quizzes administered through the tool and get instant feedback on the CT concepts they lack. Teachers can view reports and examine their students' performance. For TECHFIT teachers, it is a way of identifying the top performers in the class, identify each student's weak areas, and even track each student's progression throughout the program. For TECHFIT researchers, it can be a great way of assessing curriculum effectiveness by measuring pre and post-test results.

3.5.1 Database design

The tool uses MS SQL Server as the back-end database to support the application. The scope of the application uses the following 13 tables:

1. ct_questions.CT_Concepts - Contains information about different CT concepts.
2. ct_questions.CT_Contributors - Contains information about all the user of the tool.
3. ct_questions.CT_Habits - Contains information about different CT habits.
4. ct_questions.CT_Practices - Contains information about different CT practices.
5. ct_questions.CT_Question_MC - Contains information about MCQ based questions.
6. ct_questions.CT_Question_SA - Contains information about single answer based questions.
7. ct_questions.CT_Question_QuestionConcept - Contains information about different CT concepts a question addresses.
8. ct_questions.CT_Question_QuestionHabit - Contains information about different CT habits a question addresses.
9. ct_questions.CT_Question_QuestionPractice - Contains information about different CT practices a question addresses.
10. ct_questions.CT_Question_Questions - Contains information about all questions in the repository.
11. ct_questions.CT_Question_Quiz - Contains information about each unique quiz.
12. ct_questions.CT_Question_School - Contains information about different schools.
13. ct_questions.CT_Question_Student_Test_History - Contains information about test results of students.

3.5.2 Login and registration

A new user can request access from the admin using the signup page. The password is hashed using SHA256 that takes the password string and returns the hash as a 64-character, hexadecimal-formatted string. Figure 3.2 shows the login page.

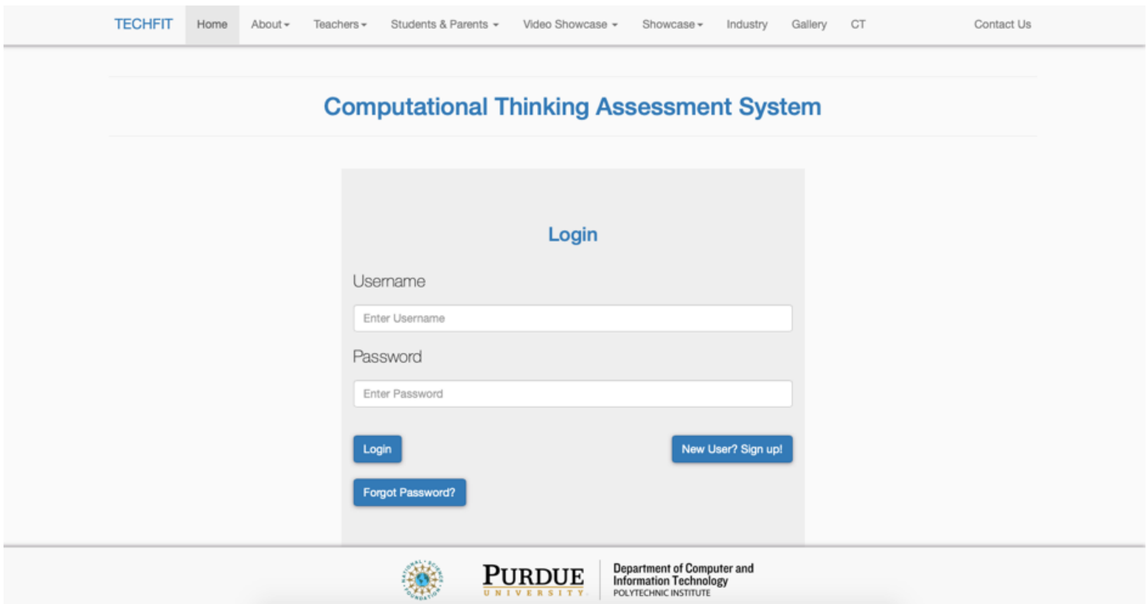


Figure 3.2. CT system login screen

There is also a forgot password feature that allows users to reset their password through a one-time password (OTP) that is sent to their registered email ID (Figure 3.3).

The tool has the following access types that allow a varied level of security and features:

- Administrator: Complete access to View/Add/Delete/Rate questions and User Management.
- Teacher: Access to View/Add/Rate questions, build and assign quizzes to students, view class and student performance.
- Student: Take quizzes, view their own performance, view/edit student profile.
- Expert: Access to View/Rate questions.

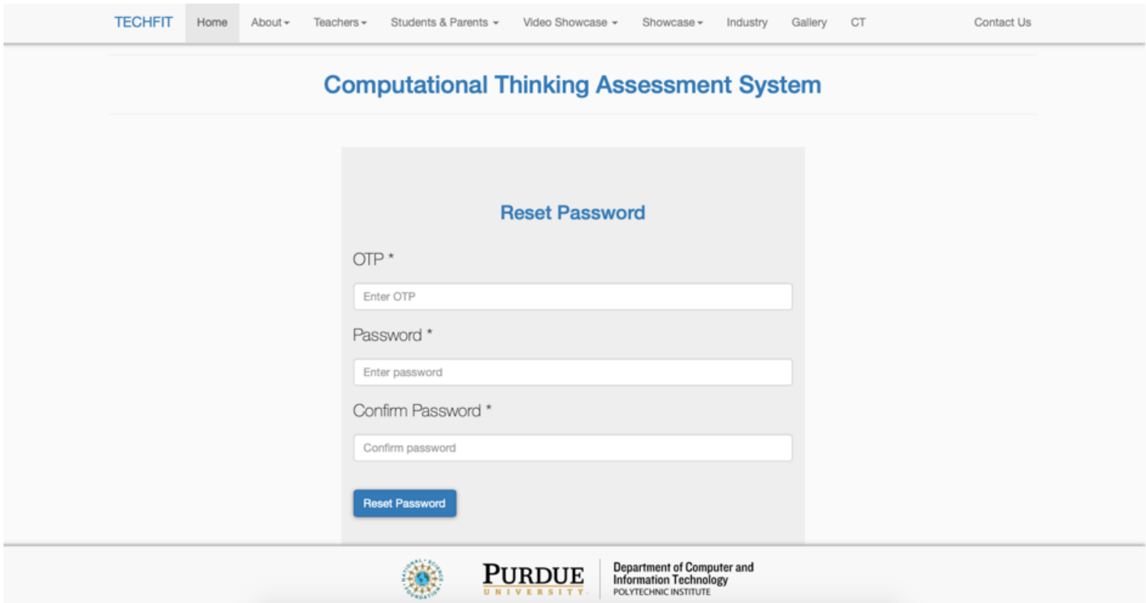


Figure 3.3. CT system forgot password

- Contributor: Access to View and Add questions.
- Viewer: Access to View questions only.

The features of the tool are explained in more detail in the following sections.

3.5.3 Teacher's perspective

The teacher can view questions in the tool using a number of different ways. They can either click on a specific concept or practice and view questions that address that selected concept or practice (Figure 3.4). They can also view all questions or view all open/close-ended questions (Figure 3.5).

Searching for a question using specific keywords is also an added feature of the tool (Figure 3.6).

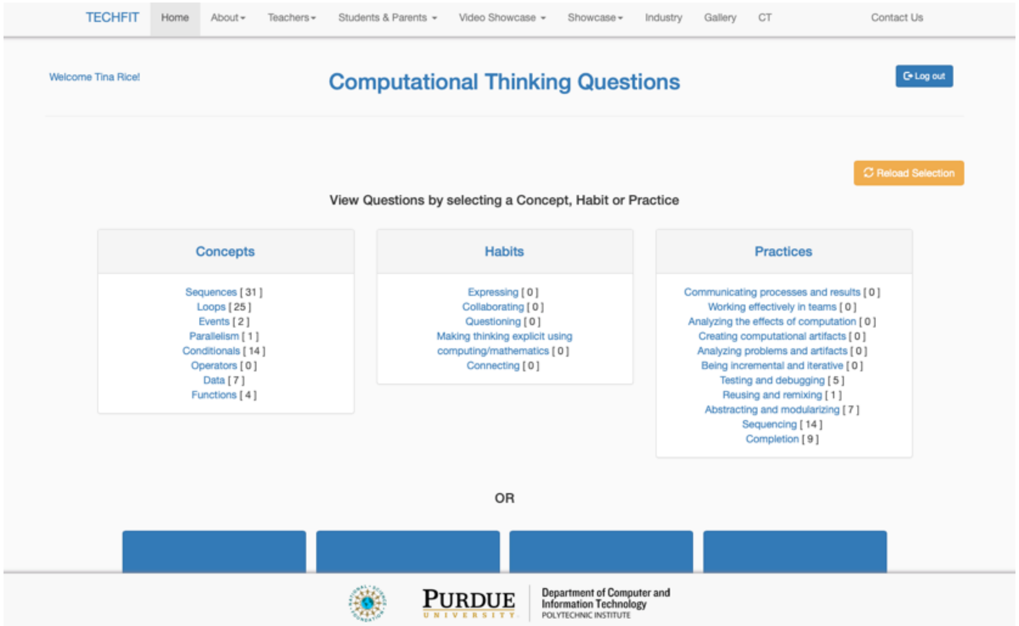


Figure 3.4. CT system view questions by CT concept/practice

The teachers are given the option to mark a question for deletion. If they pick the option, they are required to input a reason for flagging the question and then an email notification is sent to the administrator of the tool asking them to review the flagged question. The final decision to delete the question is up to the administrator's discretion. The tool also allows teachers to rate the questions based on their alignment to certain concepts and practices (Figure 3.7).

The tool contains pre-validated questions that have been proven to be effective and reliable in assessing CT skills, it allows the teachers to add any other questions they wish to use to quiz their students (Figure 3.8). There are two categories of questions: short answer and MCQ. Both question categories have the ability to add code snippets or other screenshots to the questions. Each question is tagged with one or more CT concepts or practices, so each students CT skills can be tracked.

The teacher has several options including viewing questions, rating questions, adding questions, viewing class/student performance and building quizzes (Figure 3.9.

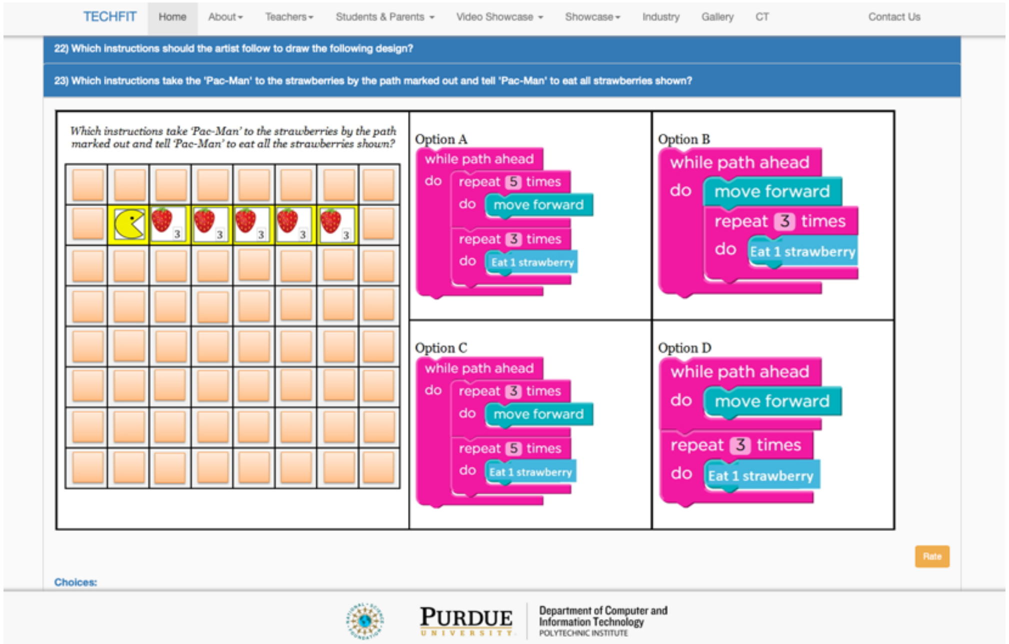


Figure 3.5. CT system question view

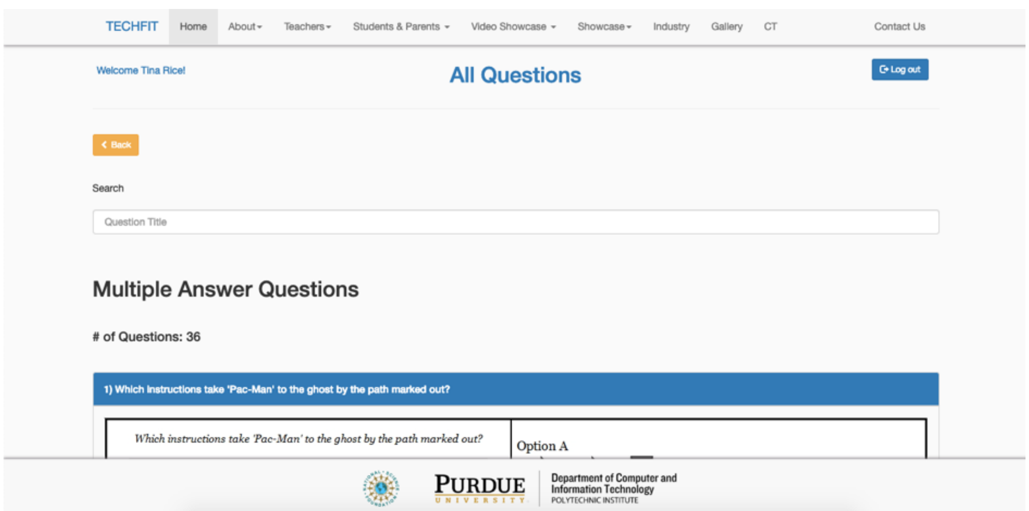


Figure 3.6. Search for a question

The class performance page shows a quick snapshot of the class average along with the top performers of your class with their average score across all the tests they have taken (Figure 3.10). Detailed information about a particular student can be viewed that shows the trend of his performance, information about their last quiz including insight on where they lack and also shows a tab of all their test scores (Figure 3.11).

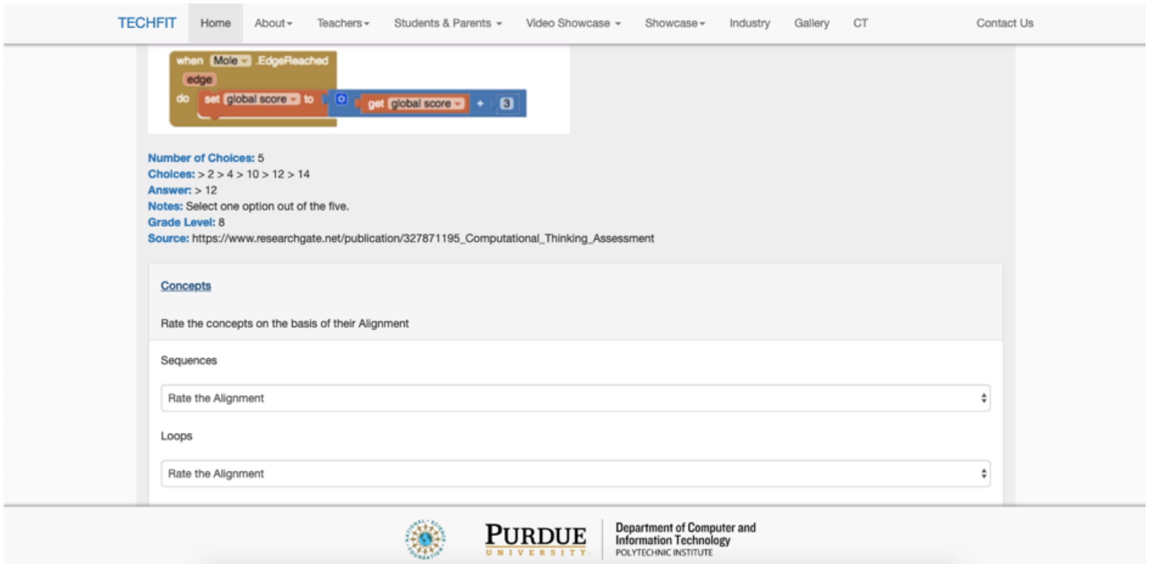


Figure 3.7. Rate a question

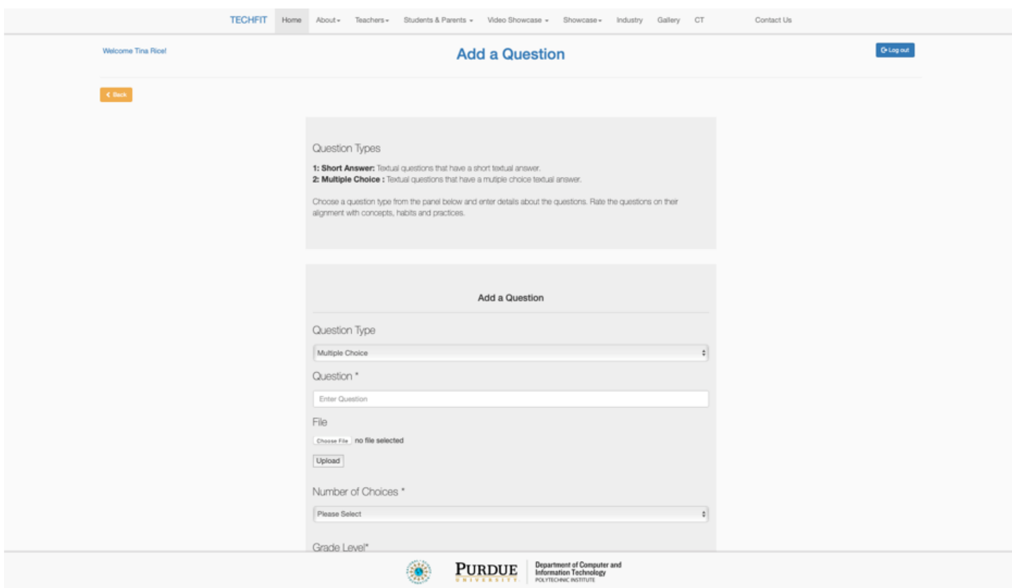


Figure 3.8. Add a question

The teachers also have the option to build a quiz. This feature allows them to select certain concepts and practices and then build a quiz containing only those dimensions. Such a feature can be useful when they wish to exclude topics that have not been taught yet. After building, the quiz can either be assigned to the entire class or particular students can be selected and assigned the quiz. This is depicted in Figure 3.12.

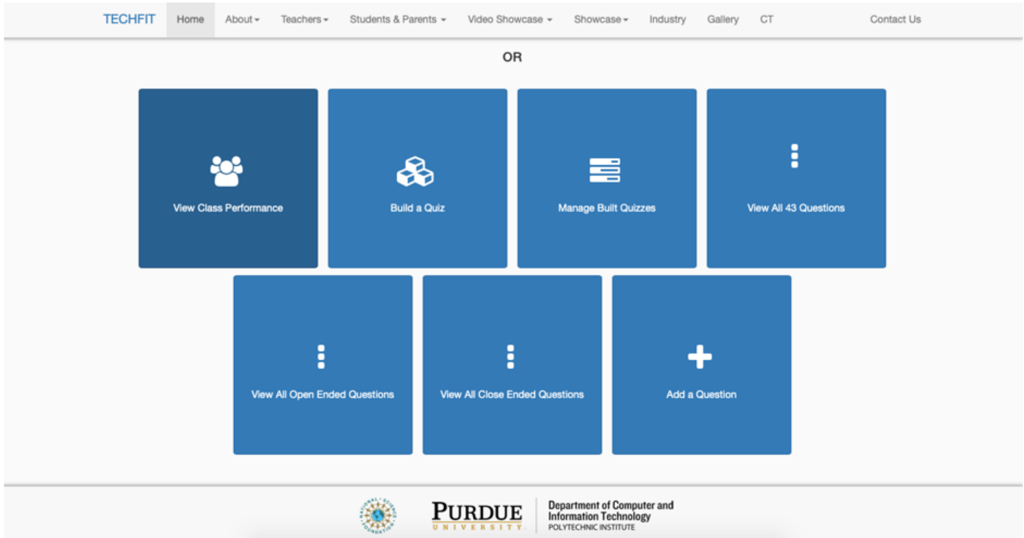


Figure 3.9. Teacher landing screen

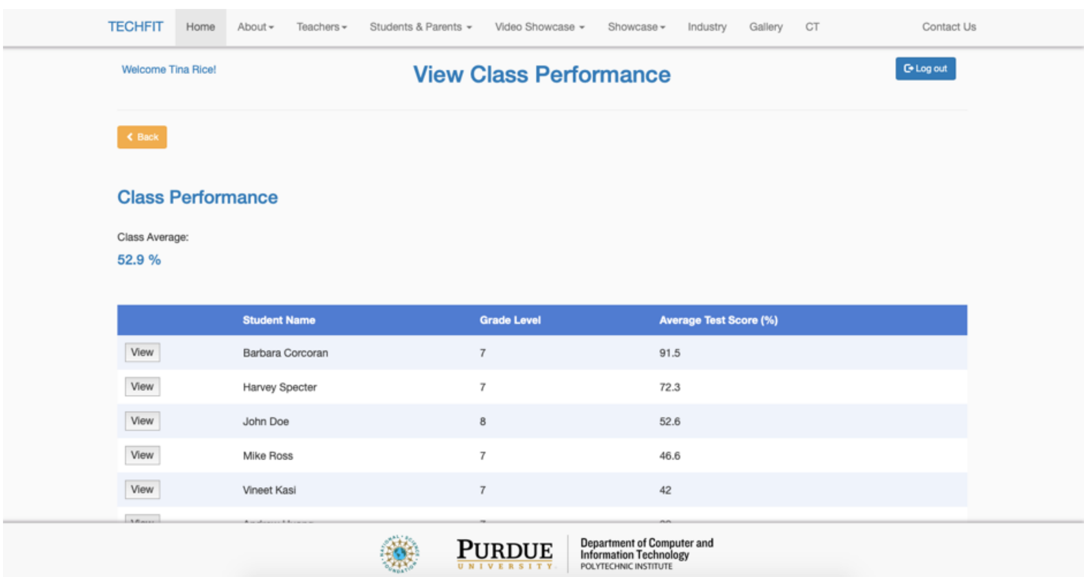


Figure 3.10. View class performance

Manage built quizzes allows the teacher to modify certain parameters of the quiz they have already created including the quiz duration if they feel the existing duration is too long or short. The teachers can also add or remove students to which the quiz is assigned to (Figure 3.13).

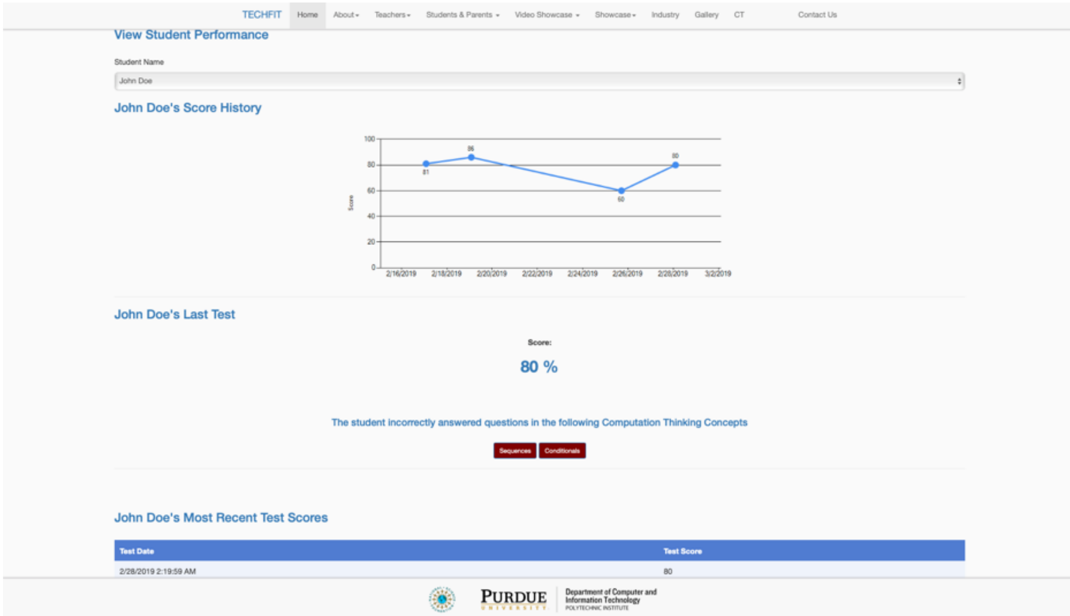


Figure 3.11. View student performance

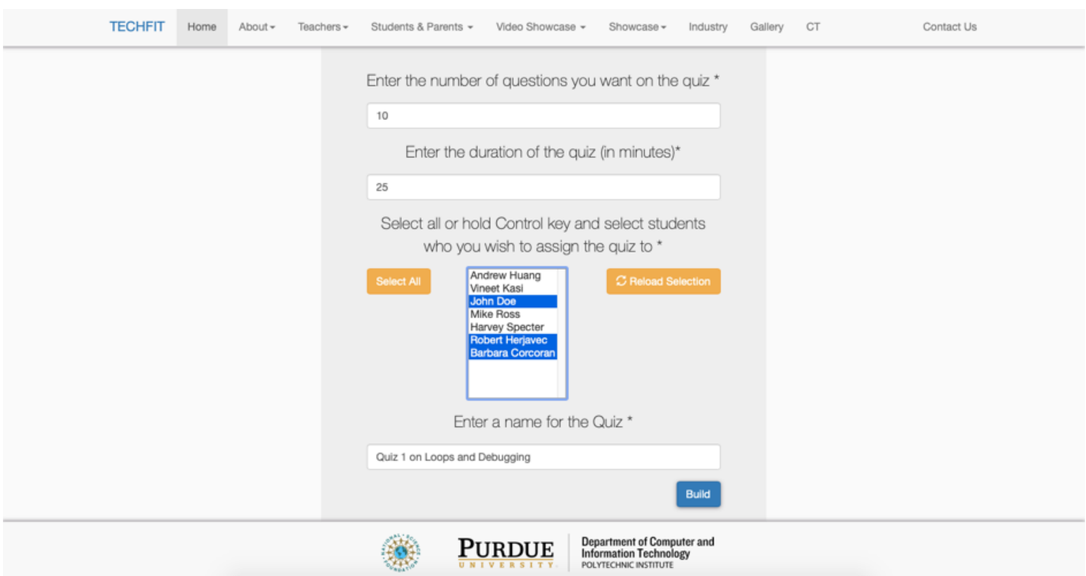


Figure 3.12. Build a quiz

3.5.4 Student's perspective

The student after logging on to the platform is presented with four options: Take a random quiz, take an assigned quiz, view past performance and edit student profile (Figure 3.14).

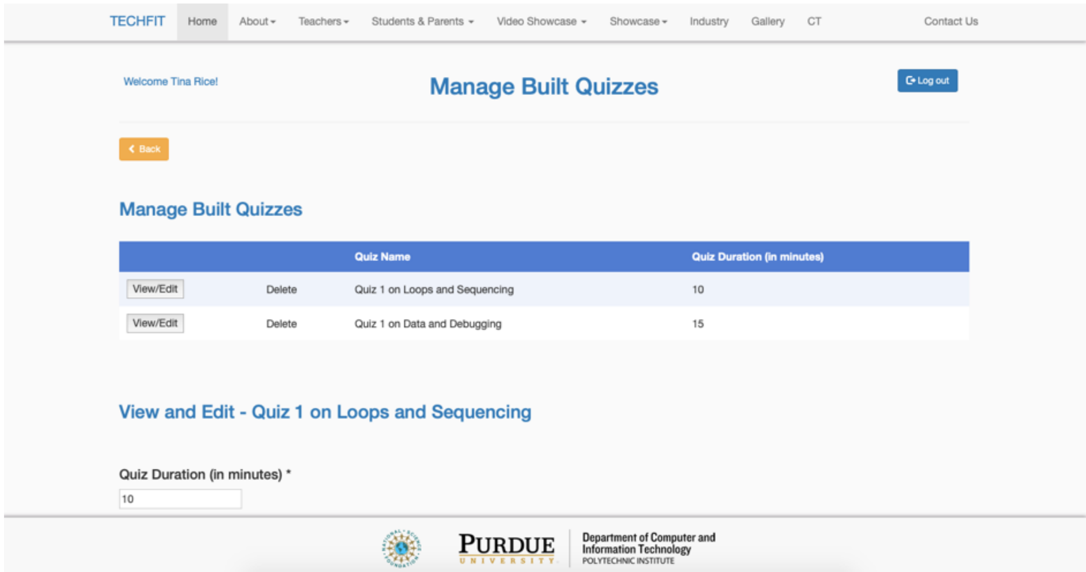


Figure 3.13. Manage built quizzes

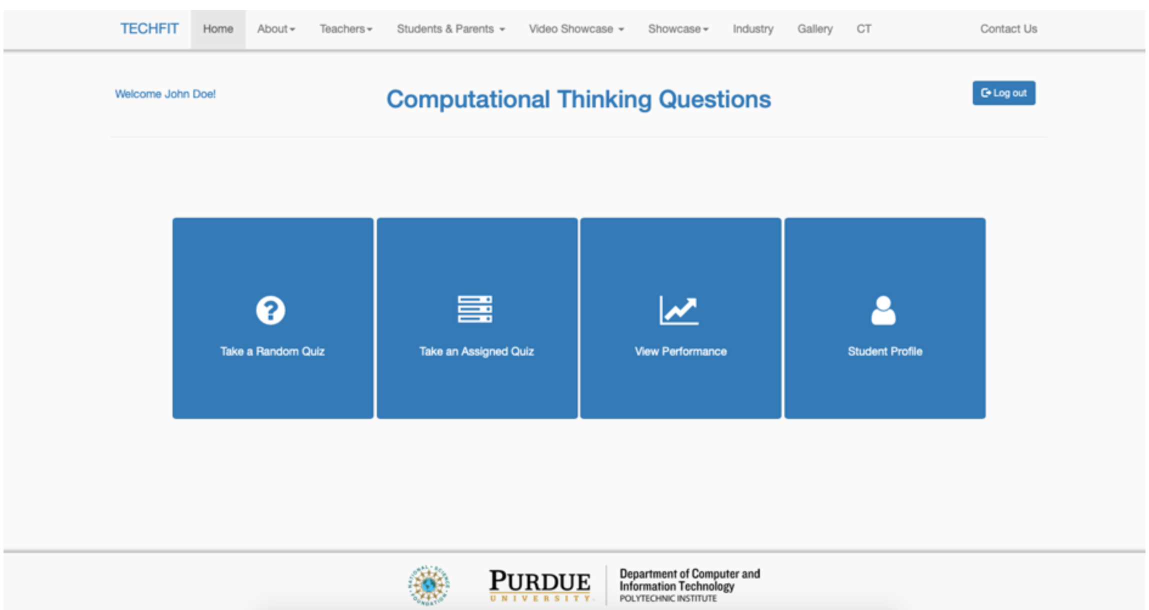


Figure 3.14. Student landing screen

They can either take a random quiz generated by the system picking top rated questions appropriate for the student's grade level (Figure 3.15) or they can take an assigned quiz containing a mix of CT concepts and practices picked by the teachers (Figure 3.16). Students are provided immediate feedback on how they performed with insight on what CT concepts and practices they lacked on the test. This provides an excellent way for them to know their weak areas and focus on improvement. Students can also view their past performance through the view past performance tab on the home page. The page shows the student his detailed past performance on all tests taken.

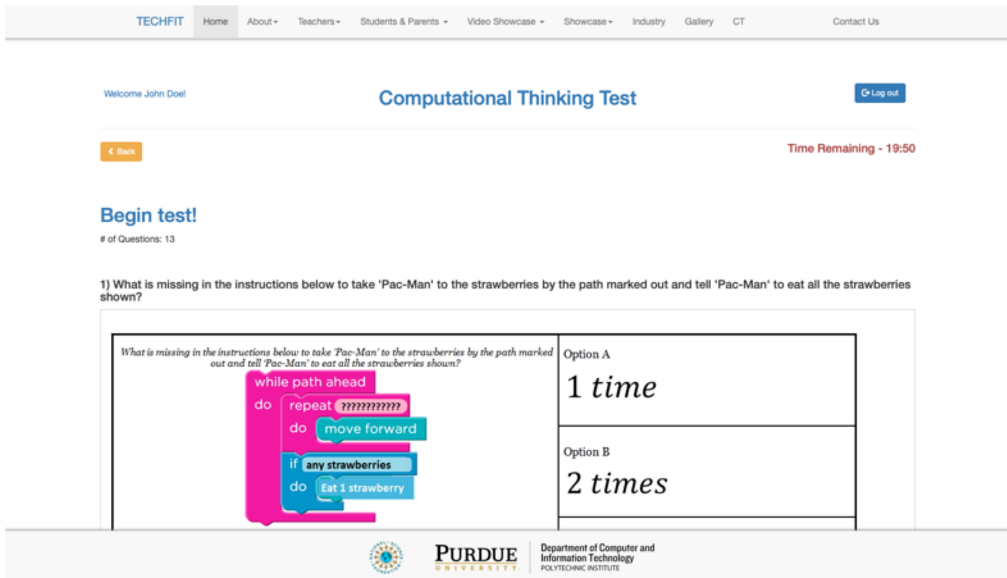


Figure 3.15. Take a random quiz

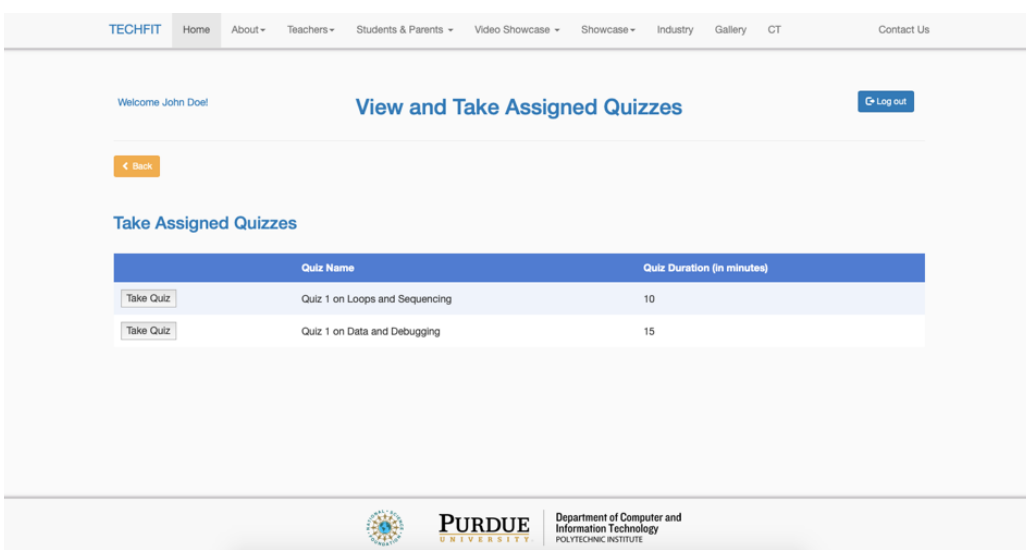


Figure 3.16. View and take assigned quizzes

The student profile section allows the student to view and edit their personal details along with their password on the tool (Figure 3.17).

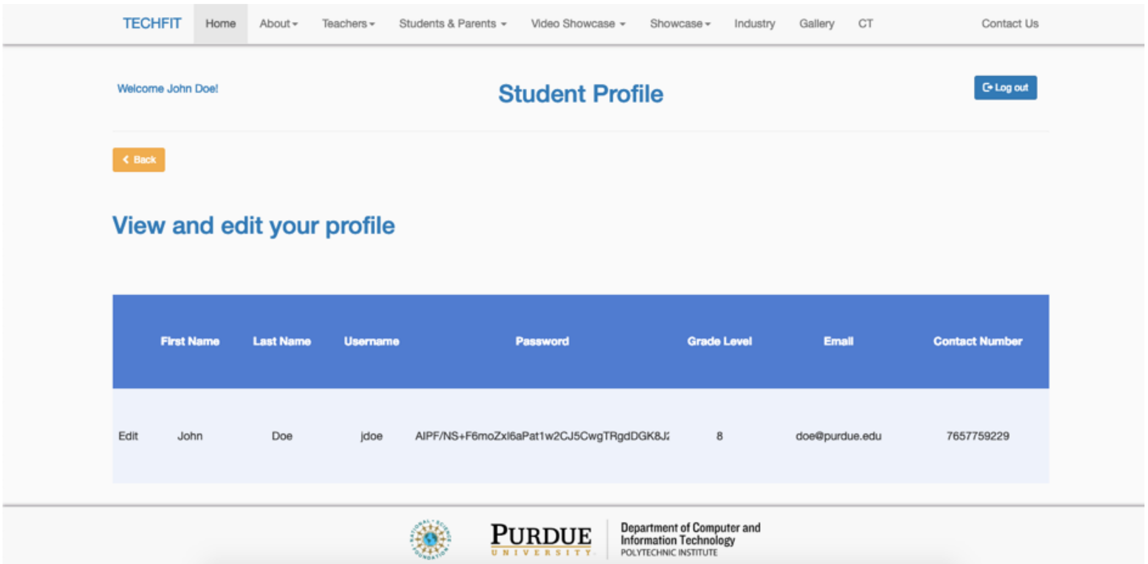


Figure 3.17. View and edit student profile

3.5.5 Admin's perspective

The tool has an admin role that allows the assigned administrator of the tool has access to the admin panel (Figure 3.18), which has options to modify and add schools (Figure 3.19), modify CT concepts and practices along with core user management. The administrator also has the capability to delete a question from the tool.

3.5.6 Expert's perspective

Experts all around the world can collaborate and rate the questions in the tool based on their alignment to certain CT concepts and practices.

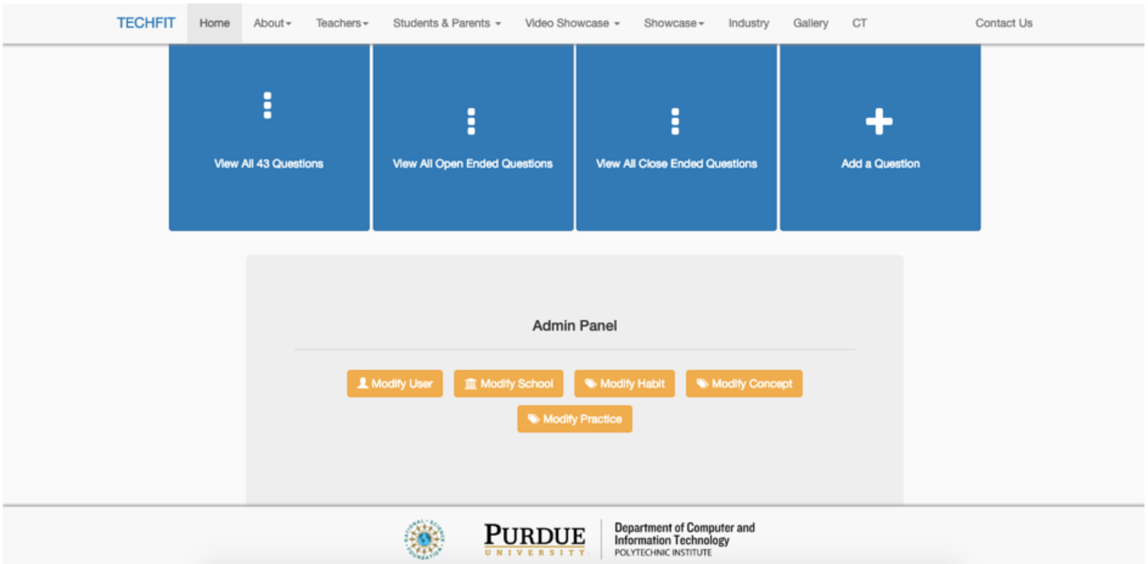


Figure 3.18. Administrator panel

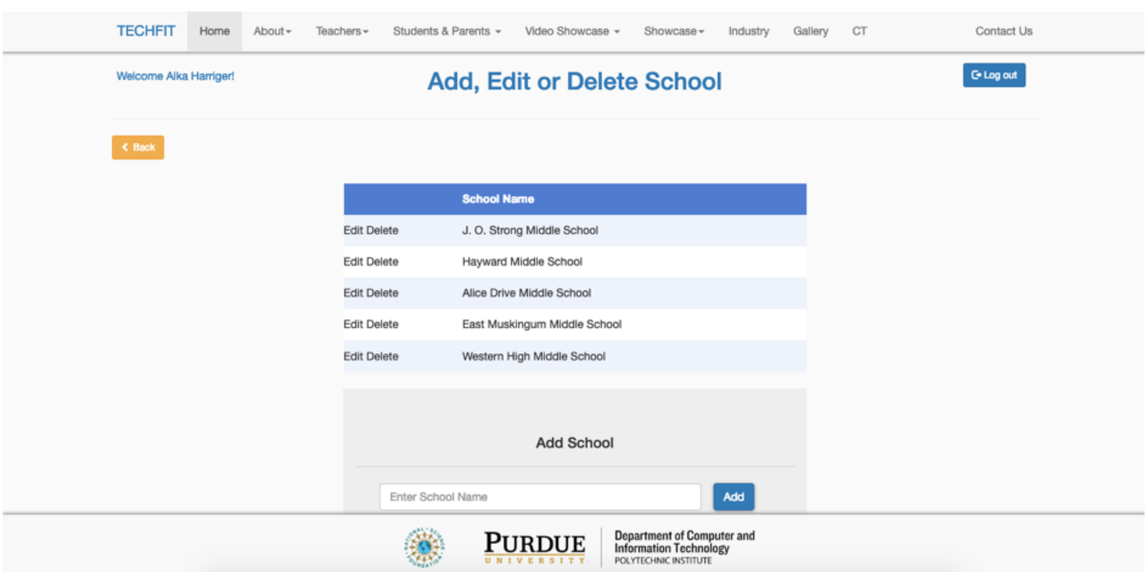


Figure 3.19. Modify and add schools

CHAPTER 4. RESULTS

After data collection, the data was analyzed to conclude whether the proposed collaborative platform for CT assessment was perceived to be helpful to TECHFIT teachers in identifying their students potential strengths and weaknesses in the programming subset of CT. This conclusion is drawn from both a quantitative and qualitative analysis. The analysis includes descriptive statistics and information extracted from the qualitative responses of the sample.

The survey was open for a period of two weeks in which there were 12 teachers who responded to the survey out of which 4 had incomplete responses and were therefore removed from the final data set. Initial analysis of the data in the form of Response Averages (RA) for all the items on the questionnaire was conducted. ”The RA to each question produces a rough idea (or a good indicator) of the aggregate direction towards the two ends of the scale for each question” (Cheung, 1997, p. 16).

Likert type questions in a survey are analyzed depending on whether they are Likert type or Likert scale data. Likert type items are single questions using the Likert response alternatives and Likert scale is a combination of 4 or more Likert type questions into a composite variable (Clason & Dormody, 1994). Likert type items includes median, mode and frequencies and analyzing Likert scale data include means and standard deviations. Since Likert scale data are summed into a score, they are treated at the interval level scale. Further analysis of Likert scale items includes non-parametric methods Spearman’s test for association and Wilcoxon’s signed rank test for hypothesis testing. (Boone & Boone, 2012)

While a one sample t-test for hypothesis testing was an approach, authors including Meek, Ozgur, and Dunning (2007) state that Bowerman, O’Connell, Murphree, Huchendorf, and Porter (2003); Doane, Seward, et al. (2011); Keller (2005) have recommended using the Wilcoxon signed rank test for ordinal data in situations where the sample size is small as this test doesn’t assume a normal distribution and is fairly robust to outliers. While t-tests are proven to be more powerful when the distributions are normally distributed (Conover, 1980), Wilcoxon’s signed rank test is more powerful with non-normal data (Sawilowsky & Blair, 1992). Also, for small sample sizes, the efficiency of the Wilcoxon's signed rank test to the t-test is almost 95% (Siegel, 1956, p. 83). In a similar study by Meek et al. (2007) where the authors compared the two approaches with different Likert scales and sample sizes, it was found that there was very little precision difference between the two. Nanna and Sawilowsky (1998) in their study using actual data from 7 point Likert scale data found the Wilcoxon's rank sum test to be more powerful for a variety of sample sizes and significance levels. Mogey (1999) in her article agree with using Wilcoxon signed rank test as an inferential technique.

Key results from each of the sections of the survey are in the subsections below.

4.1 Quantitative Data Analysis

Questions 1 through 22 adapted through different pre-validated survey instruments are based on Likert data. The questions were used with their original Likert range although the extremes for questions 1-8 were flipped for survey consistency and to not confuse the user. The response from these questions was used to conduct the quantitative analysis.

4.1.1 Age appropriateness

Question 1 covering age appropriateness was based on a 4 point scale where 1 is poor, 2 is good, 3 is very good and 4 is excellent. Table 4.1 shows the RA for these items and figure 4.1 reveals the frequencies and central tendency of the Likert type data of item 1 (question 1) to be 3. The Wilcoxon's signed rank test was performed for the following hypothesis.

H_0: The population mean is 2.5 (middle score for a 4-point Likert scale)

H_α: The population mean is greater than 2.5.

The test revealed the alternate to be significant with a p-value of 0.079 for α of 10%. This implies that the teachers agree with the statement and found the tool's ability to offer different levels (ranging from simpler questions to more complex) to assess skills very good.

Table 4.1. Response average of question 1

Question	Q1) The web tool's ability to offer different levels (ranging from simpler questions to more complex) to assess skills
Minimum	2.00
Maximum	4.00
Median	3.00
Mean	3.00
Std. Dev.	0.82
Count	9

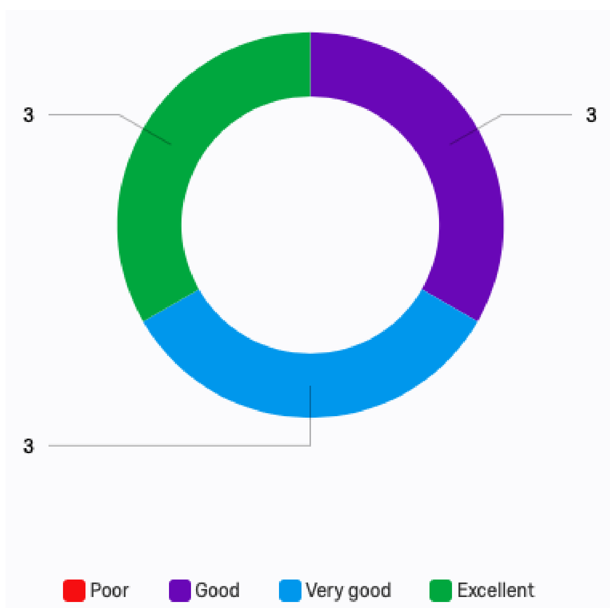


Figure 4.1. Frequency distribution and central tendency of question 1

4.1.2 Motivation level of student

Table 4.2. Response averages of item 2 (questions 2-8)

Question	Min.	Max.	Mean	Std. Dev.	Count
Q2) The web tool's ability to increase student's confidence related to the skill.	2.00	3.00	2.63	0.48	8
Q3) The web tool's ability to allow the teacher to assess the student's progress.	2.00	3.00	2.88	0.33	8
Q4) The web tool's ability to show the student the relationships between and among the concepts he or she is learning.	2.00	3.00	2.38	0.48	8
Q5) The web tool's ability to provide the teacher with information related to the student's performance.	2.00	3.00	2.75	0.43	8
Q6) The web tool's ability to provide the teacher with a basis for curriculum improvement.	2.00	4.00	2.63	0.70	8
Q7) The web tool's ability to show the student immediate feedback related to his or her progress.	1.00	3.00	2.25	0.66	8
Q8) To what degree did the web tool provide opportunity for the student to utilize his or her CT skills.	2.00	4.00	2.63	0.70	8

Questions 2 through 8 covering motivation level were based on a 4 point scale where 1 is poor, 2 is good, 3 is very good and 4 is excellent. Table 4.2 shows the RA along with the standard deviations for these items and figure 4.2 shows the frequency distribution. The average mean of the Likert scale data item 2 (question 2-8) was 2.59. The Wilcoxon's signed rank test showed that at a 5% significance level, our data did not

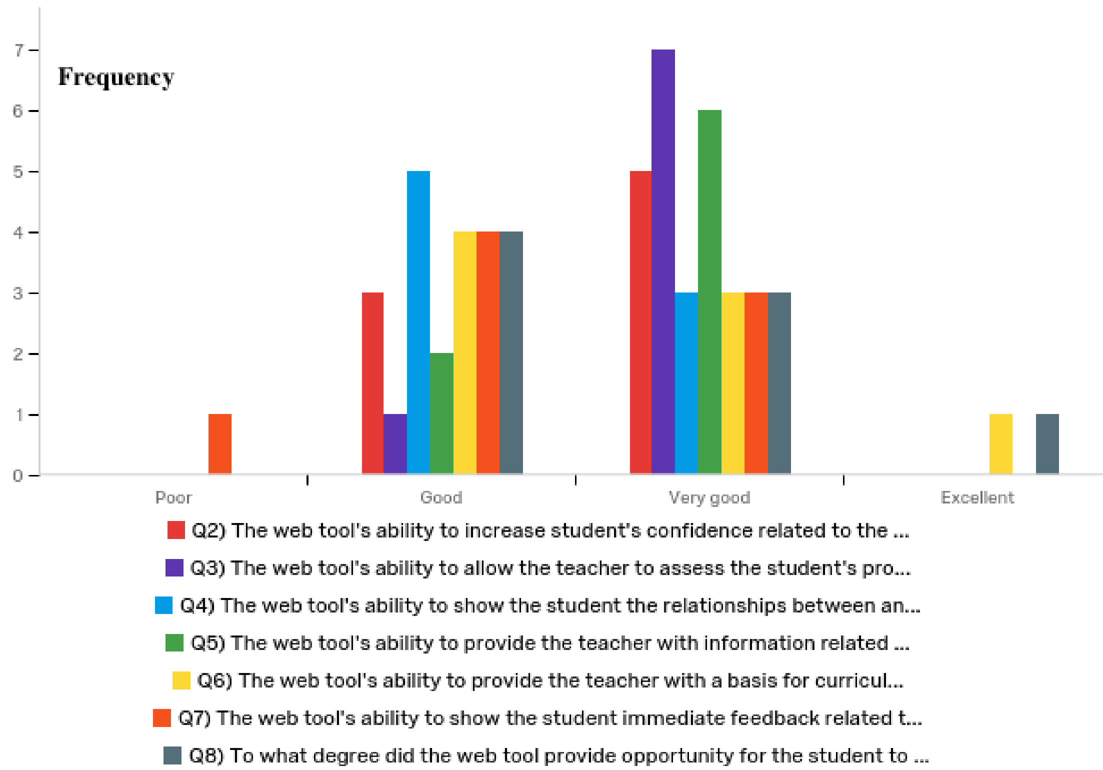


Figure 4.2. Frequency distribution of questions 2-8

have enough evidence to say that the mean is greater or lesser than the hypothesized mean of 2.5. However, the data shows significant results of individual questions 3 (at α of 0.05) and 5 (at α of 0.10) which implies that the respondents found the CT tool better than good for the statements respectively:

- The web tool has the ability to allow the teacher to assess the student's progress. (mean: 2.88, statistically significant with p-value: 0.02)

- The web tool has the ability to provide the teacher with information related to the student's performance. (mean: 2.75, statistically significant with p-value: 0.09)

4.1.3 System usability

Table 4.3. Response averages of item 3 (questions 9-16)

Question	Min.	Max.	Mean	Std. Dev.	Count
Q9) I think that I would like to use this web tool frequently.	1.00	5.00	3.75	1.20	8
Q10) I found this web tool unnecessarily complex.	1.00	3.00	1.88	0.78	8
Q11) I thought this web tool was easy to use.	1.00	5.00	3.63	1.32	8
Q12) I think that I would need assistance to be able to use this web tool.	1.00	2.00	1.50	0.50	8
Q13) I found the various functions in this web tool were well integrated.	2.00	4.00	3.75	0.66	8
Q14) I thought there was too much inconsistency in this web tool.	1.00	3.00	1.63	0.70	8
Q15) I would imagine that most people would learn to use this web tool very quickly.	3.00	5.00	4.25	0.83	8
Q16) I found this web tool very cumbersome/awkward to use.	1.00	4.00	1.88	0.93	8

Questions 9 through 16 covering usability were based on a 5 point Likert scale where 1 is strongly disagree, 2 is disagree, 3 is neither agree nor disagree, 4 is agree and 5 is strongly agree. Table 4.3 shows the RA along with the standard deviations for these questions. These questions taken from the SUS have an inverse scale every alternate question. The information from the survey is aligned with this and can be seen visually in figure 4.3.

The Wilcoxon's signed rank test was performed for the following hypothesis.

H_0: The population mean is 3 (neither agree nor disagree)

H_α: The population mean is greater than 3 (agrees).

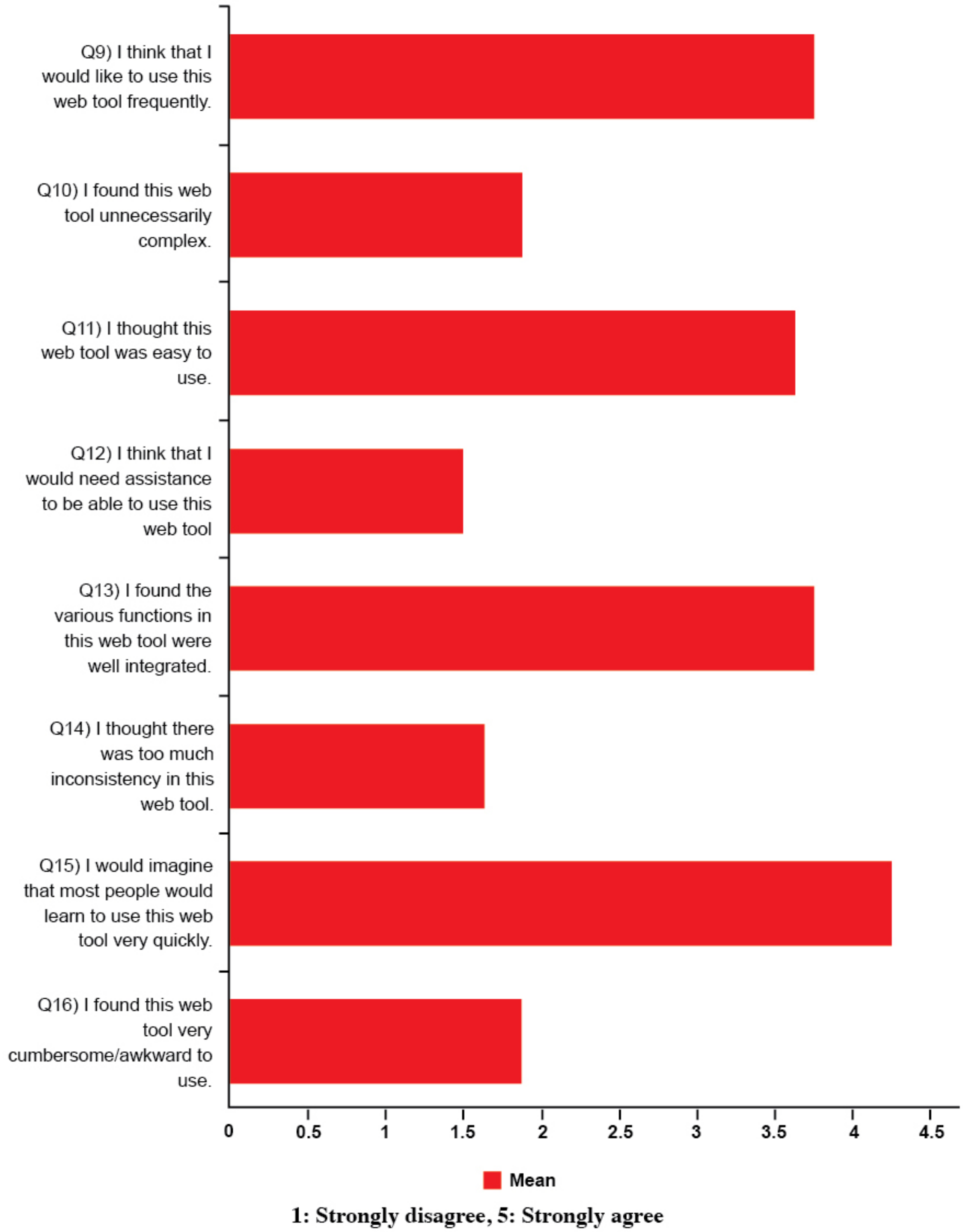


Figure 4.3. Response average of item 3 (questions 9-16)

The test revealed the alternate to be significant with a p-value of 0.01 for α of 5%. This implies that the teachers agree that the tool is usable.

The individual tests at α of 5% showed that the respondents agree with the following statements:

- I found the various functions in this web tool were well integrated. (mean: 3.75, statistically significant with p-value: 0.02054)
- I would imagine that most people would learn to use this web tool very quickly. (mean: 4.25, statistically significant with p-value: 0.01527)

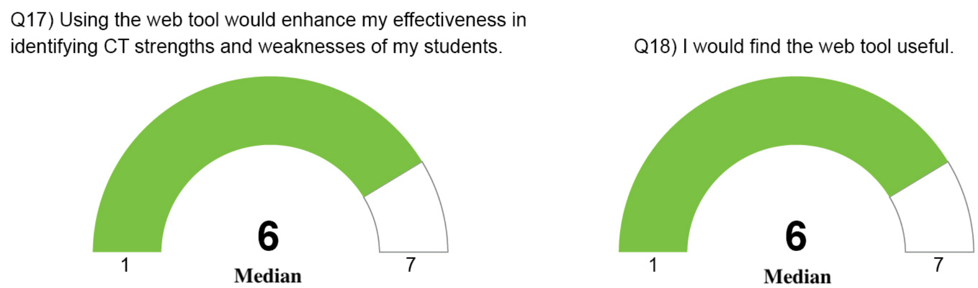
The individual tests at α of 5% showed that the respondents disagree with the following statements:

- I found this web tool unnecessarily complex. (mean: 1.88, statistically significant with p-value: 0.0016)
- I think that I would need assistance to be able to use this web tool. (mean: 1.50, statistically significant with p-value: 0.005988)
- I thought there was too much inconsistency in this web tool. (mean: 1.63, statistically significant with p-value: 0.009611)
- I found this web tool very cumbersome/awkward to use. (mean: 1.88, statistically significant with p-value: 0.01818)

4.1.4 Usefulness

Questions 17 and 18 asked the user to rate two statements based on their perceived usefulness. These questions were based on a 7 point Likert scale where 1 was unlikely and 7 was likely. Descriptive statistics are shown in figures 4.4 show that the RA for both these statements was 6.

The Wilcoxon's signed rank test was performed for the following hypothesis.



Question	Min.	Max.	Mean	Std. Dev.	Count
Q17) Using the web tool would enhance my effectiveness in identifying CT strengths and weaknesses of my students.	4.00	7.00	6.00	1.00	8
Q18) I would find the web tool useful.	3.00	7.00	6.00	1.22	8

Figure 4.4. Response averages of item 4 (questions 17-18)

H_0: The population mean is 4 (neither likely nor unlikely)

H_α: The population mean is greater than 4 (likely).

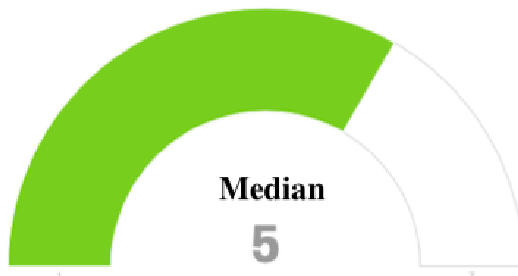
The test revealed the alternate to be significant with a p-value of 0.009 for α of 5%. This implies that the teachers agree that the tool is perceived to be useful to them.

The individual tests at α of 5% showed that the respondents agree with the following statements:

- Using the web tool would enhance my effectiveness in identifying CT strengths and weaknesses of my students. (mean: 6, statistically significant with p-value: 0.01)
- I would find the web tool useful. (mean: 6, statistically significant with p-value: 0.0096)

4.1.5 Satisfaction

Questions 19 asked the user to rate a statement based on their perceived satisfaction. This question was based on a 7 point Likert scale where 1 was strongly disagree and 7 was strongly agree. Descriptive statistics in figure 4.5 show that the median for the Likert type data was 5.



Question	Min.	Max.	Mean	Std. Dev.	Count
Q19) I felt a sense of satisfaction about this web tool.	2.00	7.00	5.00	1.41	8

Figure 4.5. Response average of item 5 (question 19)

Figure 4.6 shows that 76% of the respondents agree that they felt a satisfaction about this CT web tool.

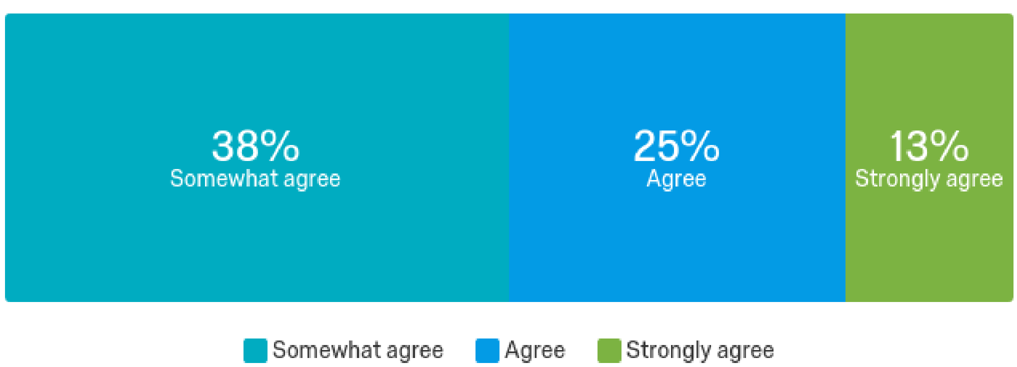


Figure 4.6. Frequency distribution of item 5 (question 19)

4.1.6 Content appropriateness

Questions 20 through 22 asked the user to rate three statements based on their perceived content appropriateness. These questions were based on a 7 point Likert scale where 1 was strongly disagree and 7 was strongly agree. Descriptive statistics in table 4.4 shows the response averages.

Table 4.4. Response averages of item 6 (questions 20-22)

Question	Min.	Max.	Mean	Std. Dev.	Count
Q20) The subject content is appropriate for delivery on the web.	4.00	7.00	5.63	0.99	8
Q21) The quiz in the web-based materials enhances the learning process.	2.00	7.00	5.13	1.36	8
Q22) The web tool can be used to supplement traditional classroom approach.	3.00	7.00	5.63	1.41	8

The Wilcoxon's signed rank test was performed for the following hypothesis.

H_0: The population mean is 4 (neither agree nor disagree)

H_α: The population mean is greater than 4 (agrees).

The test revealed the alternate to be significant with a p-value of 0.0176 for α of 5%. This implies that the teachers agree that the tool's content is appropriate.

The individual question tests at α of 5% showed that the respondents agree with the following statements:

- The subject content is appropriate for delivery on the web. (mean: 5.63, statistically significant with p-value: 0.01)
- The quiz in the web-based materials enhances the learning process. (mean: 5.13, statistically significant with p-value: 0.05)
- The web tool can be used to supplement traditional classroom approach. (mean: 5.63, statistically significant with p-value: 0.02)

4.1.7 Correlation analysis

For further analysis, associations were explored by running correlation analysis. Spearman's correlation analysis is preferred for ordinal data as the data in this study did not meet all the assumptions for the Pearson's correlation test. The Spearman's test does not assume normality and is also relatively robust to outliers (Schober, Boer, & Schwarte, 2018).

The correlation analysis results are shown in Figure 4.7. Item 1 is for questions that assess the perceived age appropriateness, item 2 is for the questions that assess the perceived motivation level, item 3 is for the questions that assess the perceived usability, item 4 is for the questions that assess the perceived usefulness, item 5 is for the questions that assess the perceived satisfaction and item 6 is for the questions that assess the perceived content appropriateness. This figure contains the Spearman's ρ coefficients and the figure 4.8 contains the p-values for those values. Generally, using convention (Schober et al., 2018), an absolute ρ value between 0.00-0.10 is negligible, 0.10-0.39 is weak, 0.40-0.69 is moderate, 0.70-0.89 is strong and 0.90-1.00 is very strong correlation.

4.1.7.1 Perceived age appropriateness and usefulness

The analysis reveals a statistically significant (p-value=0.0379 at α=0.05) strong positive monotonic correlation (ρ=0.73) between perceived age appropriateness and perceived usefulness. Perceived age appropriateness was measured using the statement ”The web tool's ability to offer different levels (ranging from simpler questions to more complex) to assess skills” (Q1). Perceived usefulness was assessed by statements including ”Using the web tool would enhance my effectiveness in identifying CT strengths and weaknesses of my students” (Q17) and ”I would find the web tool useful” (Q18). This implies that respondents who agree with the CT tool being able to adapt according to the appropriate age tend to agree with the tool's perceived usefulness as well.

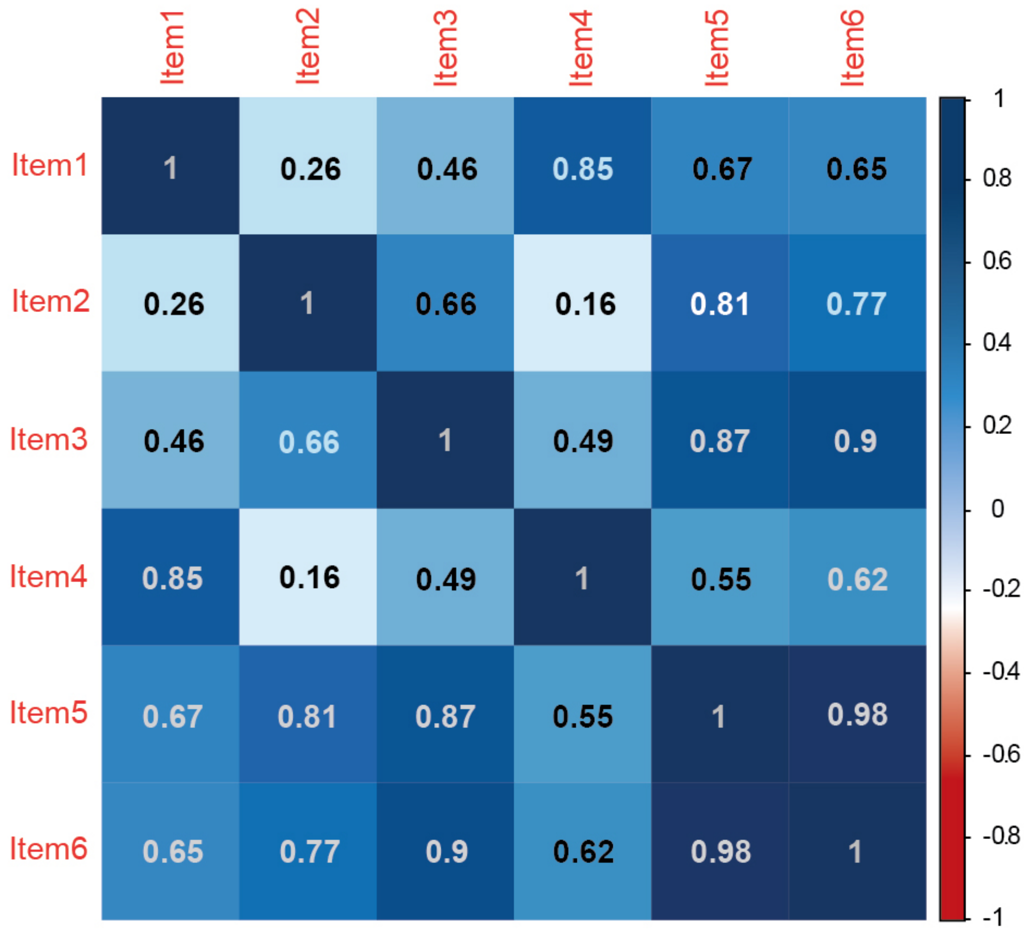


Figure 4.7. Correlation analysis coefficients

P

	Item1	Item2	Item3	Item4	Item5	Item6
Item1		0.5397	0.2847	0.0379	0.0639	0.0802
Item2	0.5397		0.0220	0.3257	0.0195	0.0427
Item3	0.2847	0.0220		0.0513	0.0007	0.0007
Item4	0.0379	0.3257	0.0513		0.0146	0.0083
Item5	0.0639	0.0195	0.0007	0.0146		0.0000
Item6	0.0802	0.0427	0.0007	0.0083	0.0000	

Figure 4.8. Correlation analysis p-values

4.1.7.2 Perceived motivation level and usability

The analysis reveals a statistically significant (p-value=0.0220 at α=0.05) strong positive monotonic correlation (ρ=0.78) between perceived motivation level and perceived usability. Perceived motivation level was assessed by statements including ”The web tool has the ability to allow the teacher to assess the students progress” (Q3) and ”The web tool has the ability to provide the teacher with information related to the students performance” (Q5). Perceived usability was measured by statements such as ”I found the various functions in this web tool were well integrated” (Q13), ”I would imagine that most people would learn to use this web tool very quickly” (Q15). Majority of the respondent’s statistically disagreed with statements like ”I found this web tool unnecessarily complex” (Q10), ”I think that I would need assistance to be able to use this web tool” (Q12), ”I thought there was too much inconsistency in this web tool” (Q14) and ”I found this web tool very cumbersome/awkward to use” (Q16). This implies that respondents who agree with the CT tool’s ability to increase student confidence, assess student performance and provide the teacher with appropriate information related to student performance tend to agree with the tool’s perceived usability as well.

4.1.7.3 Perceived motivation level and satisfaction

The analysis reveals a statistically significant (p-value=0.0195 at α=0.05) strong positive monotonic correlation (ρ=0.79) between perceived motivation level and perceived satisfaction. Perceived motivation level was assessed by statements including ”The web tool has the ability to allow the teacher to assess the students progress” (Q3) and ”The web tool has the ability to provide the teacher with information related to the students performance” (Q5). Perceived satisfaction was measured by the statement ”I felt a sense of satisfaction about this web tool” (Q19). This implies that respondents who agree with the CT tool’s ability to increase student confidence, assess student performance and provide the teacher with appropriate information related to student performance tend to perceived to be satisfied with it as well.

4.1.7.4 Perceived motivation level and content appropriateness

The analysis reveals a statistically significant (p-value=0.0427 at α=0.05) strong positive monotonic correlation (ρ=0.72) between perceived motivation level and perceived content appropriateness. Perceived motivation level was assessed by statements including ”The web tool has the ability to allow the teacher to assess the students progress” (Q3) and ”The web tool has the ability to provide the teacher with information related to the students performance” (Q5). Perceived content appropriateness was measured by the statement ”The subject content is appropriate for delivery on the web” (Q20), ”The quiz in the web-based materials enhances the learning process” (Q21) and ”The web tool can be used to supplement traditional classroom approach” (Q22). This implies that respondents who agree with the CT tool’s ability to increase student confidence, assess student performance and provide the teacher with appropriate information related to student performance tend to perceived the tool’s content to be appropriate.

4.1.7.5 Perceived usability and satisfaction

The analysis reveals a statistically significant (p-value=0.0007 at α=0.05) very strong positive monotonic correlation (ρ=0.93) between perceived usability and perceived satisfaction (figure 4.9). Perceived usability was measured by statements such as ”I found the various functions in this web tool were well integrated” (Q13), ”I would imagine that most people would learn to use this web tool very quickly” (Q15). Majority of the respondent’s statistically disagreed with statements like ”I found this web tool unnecessarily complex” (Q10), ”I think that I would need assistance to be able to use this web tool” (Q12), ”I thought there was too much inconsistency in this web tool” (Q14) and ”I found this web tool very cumbersome/awkward to use” (Q16). Perceived satisfaction was measured by the statement ”I felt a sense of satisfaction about this web tool” (Q19). This implies that respondents who agree with the CT tool’s perceived usability are satisfied with the tool.

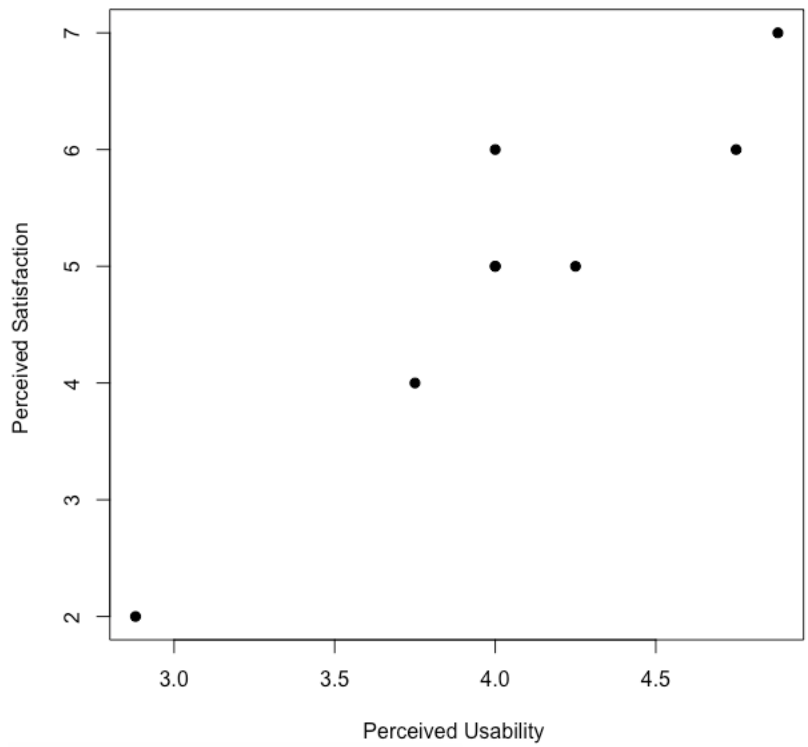


Figure 4.9. Scatter plot - perceived usability vs. perceived satisfaction

<u>4.1.7.6 Perceived usability and content appropriateness</u>

The analysis reveals a statistically significant (p-value=0.0007 at α=0.05) very strong positive monotonic correlation (ρ=0.93) between perceived usability and perceived content appropriateness (figure 4.10). Perceived usability was measured by statements such as ”I found the various functions in this web tool were well integrated” (Q13), ”I would imagine that most people would learn to use this web tool very quickly” (Q15). Majority of the respondent’s statistically disagreed with statements like ”I found this web tool unnecessarily complex” (Q10), ”I think that I would need assistance to be able to use this web tool” (Q12), ”I thought there was too much inconsistency in this web tool” (Q14) and ”I found this web tool very cumbersome/awkward to use” (Q16). Perceived content appropriateness was measured by the statement ”The subject content is appropriate for

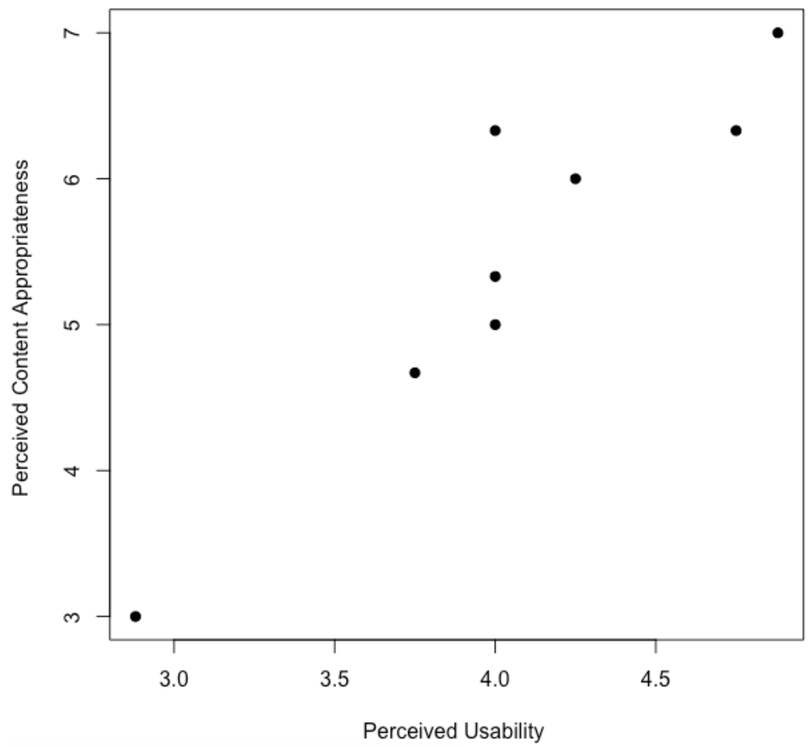


Figure 4.10. Scatter plot - perceived usability vs. content appropriateness

delivery on the web" (Q20), "The quiz in the web-based materials enhances the learning process" (Q21) and "The web tool can be used to supplement traditional classroom approach" (Q22). This implies that respondents who agree with the CT tool's perceived usability found the tool's content to be appropriate.

4.1.7.7 Perceived usefulness and satisfaction

The analysis reveals a statistically significant (p-value=0.0146 at α=0.05) strong positive monotonic correlation (ρ=0.81) between perceived usefulness and perceived satisfaction. Perceived usefulness was assessed by statements including "Using the web tool would enhance my effectiveness in identifying CT strengths and weaknesses of my students" (Q17) and "I would find the web tool useful" (Q18). Perceived satisfaction was measured by the statement "I felt a sense of satisfaction about this web tool" (Q19). This implies that respondents who agree with the CT tool to be perceived useful feel satisfied using the tool.

4.1.7.8 Perceived usefulness and content appropriateness

The analysis reveals a statistically significant (p-value=0.0083 at α=0.05) strong positive monotonic correlation (ρ=0.84) between perceived usefulness and perceived content appropriateness. Perceived usefulness was assessed by statements including ”Using the web tool would enhance my effectiveness in identifying CT strengths and weaknesses of my students” (Q17) and ”I would find the web tool useful” (Q18). Perceived content appropriateness was measured by the statement ”The subject content is appropriate for delivery on the web” (Q20), ”The quiz in the web-based materials enhances the learning process” (Q21) and ”The web tool can be used to supplement traditional classroom approach” (Q22). This implies that respondents who agree with the CT tool to be perceived useful feel that the tool’s content is appropriate.

4.1.7.9 Perceived satisfaction and content appropriateness

The analysis reveals a statistically significant (p-value=0.0146 at α=0.05) strong positive monotonic correlation (ρ=0.81) between perceived satisfaction and perceived content appropriateness. Perceived satisfaction was measured by the statement ”I felt a sense of satisfaction about this web tool” (Q19). Perceived content appropriateness was measured by the statement ”The subject content is appropriate for delivery on the web” (Q20), ”The quiz in the web-based materials enhances the learning process” (Q21) and ”The web tool can be used to supplement traditional classroom approach” (Q22). This implies that respondents who are satisfied using the tool find the content to be appropriate as well.

4.2 Qualitative Analysis

In combination with the statistical results, open-ended responses from survey supported the above results and provided further detail and feedback about the CT tool. The respondents of the survey were asked 3 open-ended questions. Q23 and Q25 (”Any positive aspects of the tool you wish to highlight?”, ”Please provide any additional comments about the tool”) asked them to add any positive comments they had about the tool. A total of 7 respondents answered Q23 and 5 answered Q25. Some of the most evocative comments from the survey are below.

- ”Massive potential for assessing, testing, probing, and analyzing students ability to computationally think...as well as break this skill down into smaller sub skills. Also does a good job, particularly in early stage development, of providing individual student information, scores, and skill + and -. This is a huge plus for real and meaningful classroom implementation.”
- ”Ease”
- ”Teachers can add their own questions. Immediate feedback and graphs of how you did.”
- ”Visually appealing (colors, characters). Easy to read format and answer choice.”
- ”I liked that I could add questions or make a quiz from the item bank.”
- ”It was good that it started off obvious and then got more challenging.”
- ”Well done.”
- ”I think this would be a great pre and post test for our students. I don’t mind saying I missed 4 questions and will try to understand what I need to know better.”
- ”I think these questions are thought provoking and discussion of the answers will provide great insight into their programming skills.”

Figure 4.11 portrays the word cloud for open-ended comments entered for Q23 and reveals that the most liked features of the tool were the adaptive question difficulty, immediate feedback to students about their weakness, and the user interface. Frequency analysis of the keywords revealed: ease (2), love (2), add questions (2), feedback(1), visually appealing (1), thought provoking (1), great(1), and well done (1).

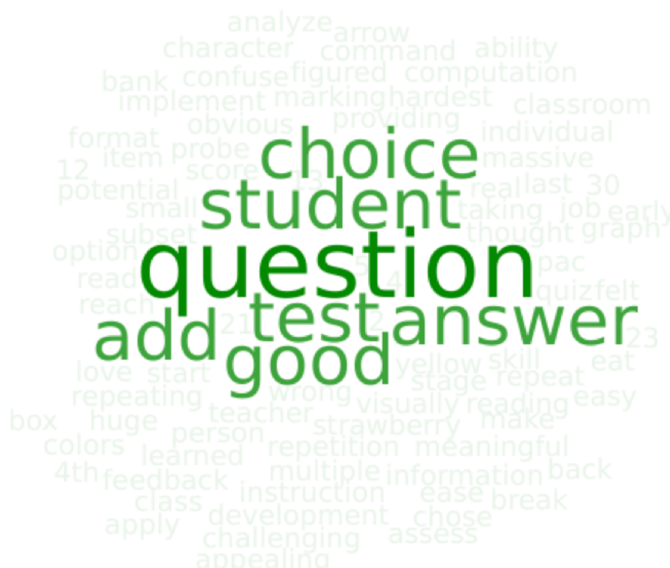


Figure 4.11. Positive aspect word cloud

Q24 (”Any negative aspects of the tool you wish to highlight?”) and Q25 (”Please provide any additional comments about the tool”) asked them to add any negative comments they had about the tool. A total of 5 respondents answered these questions. Some of the most evocative comments from the survey are below.

- ”Wouldn’t call it a negative...but a definite growth area as the tool is deployed more in real classroom situations...the question set needs to continue to grow and a large enough question base for each sub skill needs to exist so that students don’t interact with the same question over and over...thus causing problems with knowledge bias as students simply memorize proper responses. There is an avenue for users to add to the question base ..but how are these questions going to be ”graded” and evaluated before fully being implemented on the tool (pretty sure this has been figured out already but it was def. something I thought of). Perhaps in the future also add on the ability to see TOT (Time on task) that each student spends on each assigned CT assessment. This would provide valuable context when evaluating student performance.”

- ”I think it is a good web tool. The students will not be able to answer these questions in 10 minutes though. They do not even know what a sprite is when the class begins so it will take them some time to figure out a repetitive movement and how to create that effect.”

After analyzing the comments, one of the teachers had concerns about the questions being repeated. This issue has been tackled by randomly picking questions every time a new quiz is picked by the student. Of course, randomness is directly linked with the number of questions in the tool. The time on task is a parameter that might be useful to the teachers and is something that should be a part of the future work of this study. The issue of insufficient time for completing a quiz has been solved with the creation of a build a quiz feature that will allow the teachers to pick specific CT concepts or practices and build a quiz using only the ones they want. This feature allows the teacher to enter the number of questions they want on the quiz along with the duration of the quiz.

4.3 Summary

The study used a questionnaire with questions that were taken from popular pre-validated instruments that are known to be reliable. The overall and subgroup reliability of the constructed questionnaire was also measured and found to be reliable. The mixed methods analysis of the data collected revealed that the proposed CT tool was perceived to be effective for the TECHFIT teachers in identifying their student’s CT skills. The respondents agree that they are satisfied with the tool and the tool is perceived to be age appropriate, useful, usable and content appropriate. At α=5%, some of the key statements the respondents significantly agreed with are:

- The web tool has the ability to allow the teacher to assess the student's progress. (mean: 2.88 on a scale of 4 where 3 is very good, statistically significant with p-value: 0.02)
- The web tool has the ability to provide the teacher with information related to the student's performance. (mean: 2.75 on a scale of 4 where 3 is very good, statistically significant with p-value: 0.09)
- I found the various functions in this web tool were well integrated. (mean: 3.75 on a scale of 5, statistically significant with p-value: 0.02054)
- I would imagine that most people would learn to use this web tool very quickly. (mean: 4.25 on a scale of 5, statistically significant with p-value: 0.01527)
- Using the web tool would enhance my effectiveness in identifying CT strengths and weaknesses of my students. (median: 6 on a scale of 7, statistically significant with p-value: 0.01)
- I would find the web tool useful. (median: 6 on a scale of 7, statistically significant with p-value: 0.0096)
- The subject content is appropriate for delivery on the web. (mean: 5.63 on a scale of 7, statistically significant with p-value: 0.01)
- The quiz in the web-based materials enhances the learning process. (mean: 5.13 on a scale of 7, statistically significant with p-value: 0.05)
- The web tool can be used to supplement traditional classroom approach. (mean: 5.63 on a scale of 7, statistically significant with p-value: 0.02)

Correlation analysis showed significant very strong positive monotonic correlations between perceived usability and perceived satisfaction; and perceived usability and perceived content appropriateness. There were also significant strong positive correlations found between perceived age appropriateness and perceived usefulness; perceived motivation level and perceived usability; perceived motivation level and perceived satisfaction; perceived motivation level and perceived content appropriateness; perceived usefulness and perceived satisfaction; perceived usefulness and perceived content appropriateness; and perceived satisfaction and perceived appropriateness.

The qualitative results are mostly positive. The teachers liked the usability of the tool, the way the difficulty of the questions adapted to the grade of the student taking the quiz and the fact that immediate feedback about the student's weak CT areas was given after taking a quiz. Teachers commented that the questions were thought-provoking and perceived it to be a promising tool that is capable of being used in pre-post conditions to gauge CT skills. Being able to add questions to the question repository of the tool was another feature that was highlighted in the positive comments. An unexpected outcome was an unsolicited request by one of the teacher-subjects to begin using the tool immediately in her TECHFIT class. Although the tool will not be released for such use until this summer, the request demonstrates that the objective of TECHFIT teachers using the tool to assess their students' CT skills will be achieved, at least for this particular teacher (D. Shuler, personal communication, March 25, 2019). Some teachers in their feedback have given areas of improvement, some of which have already been incorporated in the tool as discussed in the section above.

CHAPTER 5. CONCLUSION AND FUTURE WORK

The proposed CT tool was built as a platform for CT assessment with a validated repository of validated questions. Experts can collaborate and rate questions, adding to their reliability. The responsive, web-based tool makes it easily accessible to anyone using a browser and eliminates the operating system dependence. The tool helps assess teachers their students' CT skill using quizzes which make it a quick way to gauge their understanding while the tool provides instant feedback about CT areas they lacked. The tool has the advantage of combining multiple sources to add questions from, making it source independent and allowing different kinds of questions to be populated into the tool's repository. The quizzes contain questions that adapt according to the age level of the student taking the quiz. An important feature of the tool is that it allows the teachers to build quizzes by selecting specific CT concepts and practices, the number of questions and the duration of the quiz. They can then assign the quiz to the entire class or a subset.

To answer the research question of finding out the TECHFIT teachers' perceptions of the effectiveness of the CT assessment system in identifying their students' CT strengths and weaknesses in the programming subset of CT, a survey was conducted. The analysis of the mixed methods study uses the statistical backing in combination with the comprehensive feedback gathered through the survey and reveals that the TECHFIT teachers perceive the tool to be effective in identifying their students' CT skills in the programming subset of CT. Through the statistically significant quantitative analysis, it is shown the respondents agree that they are satisfied with the tool and the tool is perceived to be age appropriate, useful, usable and content appropriate. There were also strong correlations found between the perceived age appropriateness and usefulness, perceived motivation level and usability, perceived motivation level and satisfaction, perceived motivation level and content appropriateness, perceived usability and satisfaction, perceived usability and content appropriateness, perceived usefulness and satisfaction, perceived usability and content appropriateness, perceived satisfaction and content appropriateness. The qualitative analysis revealed that the feedback from the respondents was heavily positive and the some of the most liked features of the tool were the adaptive question difficulty, immediate feedback to students about their weakness, and the user interface. Teachers found the questions thought-provoking and perceived it to be a promising tool that is capable of being used in pre-post conditions to gauge CT skills.

Overall, the findings of this study regarding the proposed CT assessment tool are positive and educators/researchers are encouraged to increase the question bank by developing validated questions that are reliable in assessing CT skills. Further study with larger sample size is recommended to improve the accuracy of the tests. The future vision of this work includes expanding the tool into a go-to platform for CT assessment where CT experts can collaborate and add valid, reliable questions and anyone and not just teachers and students wanting to assess their CT skills can do so by using the web application.

REFERENCES

Advancing computational thinking across k-12 education. (2019, Mar). Retrieved from https://digitalpromise.org/2017/12/06/advancing-computational-thinking-across-k-12-education/

Ahamed, S. I., Brylow, D., Ge, R., Madiraju, P., Merrill, S. J., Struble, C. A., & Early, J. P. (2010). Computational thinking for the sciences: a three day workshop for high school science teachers. In *Proceedings of the 41st ACM technical symposium on Computer science education* (pp. 42–46).

Aho, A. V. (2012, July). Computation and Computational Thinking. *Comput. J.*, *55*(7), 832–835. Retrieved from http://dx.doi.org/10.1093/comjnl/bxs074 doi: 10.1093/comjnl/bxs074

Aiken, J. M., Caballero, M. D., Douglas, S. S., Burk, J. B., Scanlon, E. M., Thoms, B. D., & Schatz, M. F. (2013). Understanding student computational thinking with computational modeling. In *AIP Conference Proceedings* (Vol. 1513, pp. 46–49).

Ambrósio, A. P., Xavier, C., & Georges, F. (2014). Digital ink for cognitive assessment of computational thinking. In *Frontiers in education conference (FIE), 2014 IEEE* (pp. 1–7).

Anderson, N. D. (2016). A call for computational thinking in undergraduate psychology. *Psychology Learning & Teaching*, *15*(3), 226-234. doi: 10.1177/1475725716659252

Atkinson, D., & Yeoh, S. (2008). Student and staff perceptions of the effectiveness of plagiarism detection software. *Australasian Journal of Educational Technology*, *24*(2).

Atmatzidou, S., & Demetriadis, S. (2016). Advancing students computational thinking skills through educational robotics: A study on age and gender relevant differences. *Robotics and Autonomous Systems*, *75*, 661–670.

Barba, L. (2016). *Computational thinking: I do not think it means what you think it means.* Berkeley Institute for Data Science. Retrieved from `https://bids.berkeley.edu/news/computational-thinking-i-do-not-think-it-means-what-you-think-it-means`

Barefoot, C. A. S. (2014). *Computational thinking.* Retrieved from `https://barefootcas.org.uk/barefoot-primary-computing-resources/concepts/computational-thinking/`

Barr, D., Harrison, J., & Conery, L. (2011). Computational thinking: A digital age skill for everyone. *Learning & Leading with Technology*, *38*(6), 20–23.

Basawapatna, A., Koh, K. H., Repenning, A., Webb, D. C., & Marshall, K. S. (2011). Recognizing computational thinking patterns. In *Proceedings of the 42nd ACM technical symposium on Computer science education* (pp. 245–250).

Basu, S., Kinnebrew, J. S., & Biswas, G. (2014). Assessing student performance in a computational-thinking based science learning environment. In *International conference on intelligent tutoring systems* (pp. 476–481).

Bati, K. (2018). Computational thinking test (ctt) for middle school students. *Mediterranian Journal of Educational Research*, *12*(23), 89–101.

Bell, T., Witten, I. H., Fellows, M., Adams, R., & McKenzie, J. (2005). Computer science unplugged: An enrichment and extension programme for primary-aged children.

Berland, M., & Wilensky, U. (2015). Comparing virtual and physical robotics environments for supporting complex systems and computational thinking. *Journal of Science Education and Technology*, *24*(5), 628–647.

Bers, M. U., Flannery, L., Kazakoff, E. R., & Sullivan, A. (2014). Computational thinking and tinkering: Exploration of an early childhood robotics curriculum. *Computers & Education*, *72*, 145–157.

Boone, H. N., & Boone, D. A. (2012). Analyzing likert data. *Journal of extension*, *50*(2), 1–5.

Bowerman, B. L., O'Connell, R. T., Murphree, E., Huchendorf, S. C., & Porter, D. C. (2003). *Business statistics in practice*. McGraw-Hill/Irwin New York.

Brennan, K., & Resnick, M. (2012). New frameworks for studying and assessing the development of computational thinking. In *Proceedings of the 2012 annual meeting of the American Educational Research Association, Vancouver, Canada* (Vol. 1, p. 25).

Brooke, J., et al. (1996). Sus-a quick and dirty usability scale. *Usability evaluation in industry*, *189*(194), 4–7.

Brown, N. C. C., Kölling, M., Crick, T., Peyton Jones, S., Humphreys, S., & Sentance, S. (2013). Bringing computer science back into schools: lessons from the uk. In *Proceeding of the 44th ACM technical symposium on Computer science education* (pp. 269–274).

Cetin, I. (2016). Preservice teachers introduction to computing: Exploring utilization of scratch. *Journal of Educational Computing Research*, *54*(7), 997–1021.

Chang, V. (1999). Evaluating the effectiveness of online learning using a new web based learning instrument. In *Proceedings western australian institute for educational research forum.*

Cheung, E. (1997). A qualitative and statistical analysis of students perceptions in internet learning. *Journal of Systemics, Cybernetics and Informatics*, *1*(2).

Clason, D. L., & Dormody, T. J. (1994). Analyzing data measured by individual likert-type items. *Journal of agricultural education*, *35*(4), 4.

Code.org 2018 annual report. (n.d.). Retrieved from `https://code.org/about`

Committee, K.-. C. S. F. S., et al. (2016). K-12 computer science framework.

Computational thinking. (n.d.). Learning.com. Retrieved from `https://catalog.learning.com/search?unit=Computational%20Thinking`

Computational thinking for every student. (n.d.). Retrieved from `https://www.myskillsfuture.sg/content/portal/en/career-resources/career-resources/lmi-industry-news/computational-thinking-for-every-student.html`

Computational thinking quiz. (n.d.). BBC Bitesize. Retrieved from `https://www.bbc.com/bitesize/topics/z7tp34j`

Computational thinking will be vital for the future job market. (2018, Jul). Retrieved from `https://www.enterprisetimes.co.uk/2018/07/04/computational-thinking-will-be-vital-for-the-future-job-market/`

Conover, W. J. (1980). Practical nonparametric statistics.

Cortina, J. M. (1993). What is coefficient alpha? an examination of theory and applications. *Journal of applied psychology*, *78*(1), 98.

Cronbach, L. (1951). *Coefficient alpha and the internal consistency of tests. psychometrica, 16, 297-334.*

CSTA and ISTE. (2011). *Operational definition of computational thinking for K-12 education.* Retrieved from `http://www.iste.org/docs/ct-documents/computational-thinking-operational-definition-flyer.pdf`

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 319–340.

Denner, J., Werner, L., Campe, S., & Ortiz, E. (2014). Pair programming: Under what conditions is it advantageous for middle school students? *Journal of Research on Technology in Education*, *46*(3), 277–296.

Digital literacy quiz. (2016). Quizziz. Retrieved from `https://quizizz.com/admin/quiz/576e6f4391cb32ef5fc69a36/computational-thinking`

Doane, D. P., Seward, L. W., et al. (2011). *Applied statistics in business and economics*. New York, NY: McGraw-Hill/Irwin,.

Engaging students through computational thinking. (2018, Sep). Retrieved from `https://www.gettingsmart.com/2018/09/engaging-students-through-computational-thinking/`

Field, A. (2009). *Discovering statistics using spss*. Sage publications.

Furner, J. M., & Daigle, D. (2004). The educational software/website effectiveness survey. *International Journal of Instructional Media*, *31*(1), 61–78.

Grover, S., Pea, R., & Cooper, S. (2015). Designing for deeper learning in a blended computer science course for middle school students. *Computer Science Education*, *25*(2), 199–237.

Harriger, A., & Harriger, B. (2017). Supporting students stem innovations with industry partners.

Henderson, P. B., Cortina, T. J., & Wing, J. M. (2007). Computational thinking. *ACM SIGCSE Bulletin*, *39*(1), 195–196.

Hunt, K. A., & Riley, D. (2014). *Computational thinking for the modern problem solver*. Chapman and Hall/CRC.

Ioannidou, A., Bennett, V., Repenning, A., Koh, K. H., & Basawapatna, A. (2011). Computational thinking patterns. *Online Submission*.

Israel, M., Pearson, J. N., Tapia, T., Wherfel, Q. M., & Reese, G. (2015). Supporting all learners in school-wide computational thinking: A cross-case qualitative analysis. *Computers & Education*, *82*, 263–279.

Jennings, J. (2018, May). *Why we need computational thinking.* Advancing K12. Retrieved from `https://www.skyward.com/discover/blog/skyward-blogs/skyward-executive-blog/may-2018/why-we-need-computational-thinking`

Jun, S., Han, S., Kim, H., & Lee, W. (2014). Assessing the computational literacy of elementary students on a national level in korea. *Educational Assessment, Evaluation and Accountability*, *26*(4), 319–332.

K3 computational thinking starter quiz. (2018). Tes. Retrieved from `https://www.tes.com/teaching-resource/ks3-computational-thinking-starter-quiz-11864915`

Kalelioğlu, F. (2015). A new way of teaching programming skills to k-12 students: Code. org. *Computers in Human Behavior*, *52*, 200–210.

Keller, G. (2005). Statistics for management and economics ed: 9e.

Ketenci, T., Calandra, B., Margulieux, L., & Cohen, J. (n.d.). The relationship between learner characteristics and student outcomes in a middle school computing course: An exploratory analysis using structural equation modeling. *Journal of Research on Technology in Education*.

Ketenci, T., Calandra, B., Margulieux, L., & Cohen, J. (2019). The relationship between learner characteristics and student outcomes in a middle school computing course: An exploratory analysis using structural equation modeling. *Journal of Research on Technology in Education*, 1–14.

Kim, B., Kim, T., & Kim, J. (2013). and-pencil programming strategy toward computational thinking for non-majors: Design your solution. *Journal of Educational Computing Research*, *49*(4), 437–459.

Kline, P. (2013). *Handbook of psychological testing*. Routledge.

Koh, K. H., A., Bennett, V., & Repenning, A. (2010). Towards the automatic recognition of computational thinking for adaptive visual language learning. In *Visual languages and human-centric computing (VL/HCC), 2010 IEEE symposium* (pp. 59–66).

Lee, I., Martin, F., & Apone, K. (2014). Integrating computational thinking across the k–8 curriculum. *Acm Inroads*, *5*(4), 64–71.

Lee, Y., Kozar, K. A., & Larsen, K. R. (2003). The technology acceptance model: Past, present, and future. *Communications of the Association for information systems*, *12*(1), 50.

Lin, J. M.-C., & Liu, S.-F. (2012). An investigation into parent-child collaboration in learning computer programming. *Journal of Educational Technology & Society*, *15*(1).

Lye, S. Y., & Koh, J. H. L. (2014). Review on teaching and learning of computational thinking through programming: What is next for k-12? *Computers in Human Behavior*, *41*, 51–61.

McGrew, K. S. (2009). *CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research.* Elsevier.

Meek, G. E., Ozgur, C., & Dunning, K. (2007). Comparison of the t vs. wilcoxon signed-rank test for likert scale data and small samples. *Journal of modern applied statistical methods*, *6*(1), 10.

Meyers, E., & Huang, H. (2019). Learning to code, coding to learn: youth and computational thinking. *Call for papers, Information and Learning Science, available at: www. emeraldgrouppublishing. com/products/journals/call_for_papers. htm.*

Mogey, N. (1999). So you want to use a likert scale. *Learning technology dissemination initiative*, *25*.

Moran, G., & Moran, G. (2016, Jun). *These will be the top jobs in 2025 (and the skills you'll need to get them).* Fast Company. Retrieved from `https://www.fastcompany.com/3058422/these-will-be-the-top-jobs-in-2025-and-the-skills-youll-need-to-get-them`

Moreno-León, J., & Robles, G. (2015). Analyze your scratch projects with Dr . scratch and assess your computational thinking skills..

Mühling, A., Ruf, A., & Hubwieser, P. (2015). Design and first results of a psychometric test for measuring basic programming abilities. In *Proceedings of the workshop in primary and secondary computing education* (pp. 2–10).

Nanna, M. J., & Sawilowsky, S. S. (1998). Analysis of likert scale data in disability and medical rehabilitation research. *Psychological Methods*, *3*(1), 55.

NRC, et al. (2010). *Report of a workshop on the scope and nature of computational thinking*. National Academies Press.

O'keefe, D. J. (2002). Persuasion. *The International Encyclopedia of Communication*.

ÖZKÖK, A. (2013). Reliability and validity of the turkish version of the web-based learning environment instrument (weblei). *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, *28*(28-2), 335–347.

Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. Basic Books, Inc.

Papert, S., & Harel, I. (1991). Situating constructionism. *Constructionism*, *36*(2), 1–11.

Pawley, A. L. (2009). Universalized narratives: Patterns in how faculty members define engineering. *Journal of Engineering Education*, *98*(4), 309–319.

Rattray, J., & Jones, M. C. (2007). Essential elements of questionnaire design and development. *Journal of clinical nursing*, *16*(2), 234–243.

Razzouk, R., & Shute, V. (2012). What is design thinking and why is it important? *Review of Educational Research*, *82*(3), 330-348. doi: 10.3102/0034654312457429

Reed, D. A., Bajcsy, R., Fernandez, M. A., Griffiths, J.-M., Mott, R. D., Dongarra, J., … others (2005). *Computational science: Ensuring america's competitiveness* (Tech. Rep.). President's Information Technology Advisory Committee, Arlington, VA.

Robinson, M. A. (2018). Using multi-item psychometric scales for research and practice in human resource management. *Human Resource Management*, *57*(3), 739–750.

Román-González, M., Pérez-González, J.-C., & Jiménez-Fernández, C. (2017). Which cognitive abilities underlie computational thinking? criterion validity of the computational thinking test. *Computers in Human Behavior*, *72*, 678–691.

Román-González, M., Pérez-González, J.-C., Moreno-León, J., & Robles, G. (2018). Can computational talent be detected? predictive validity of the computational thinking test. *International Journal of Child-Computer Interaction*, *18*, 47–58.

Sauro, J. (2011). *A practical guide to the system usability scale: Background, benchmarks & best practices*. Measuring Usability LLC Denver, CO.

Sawilowsky, S. S., & Blair, R. C. (1992). A more realistic look at the robustness and type ii error properties of the t test to departures from population normality. *Psychological bulletin*, *111*(2), 352.

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: appropriate use and interpretation. *Anesthesia & Analgesia*, *126*(5), 1763–1768.

Seiter, L., & Foreman, B. (2013). Modeling the learning progressions of computational thinking of primary grade students. In *Proceedings of the ninth annual international ACM conference on International computing education research* (pp. 59–66).

Shute, V. J., Sun, C., & Asbell-Clarke, J. (2017). Demystifying computational thinking. *Educational Research Review*.

Siegel, S. (1956). Nonparametric statistics for the behavioral sciences.

Smith, D. C., Cypher, A., & Tesler, L. (2000). Programming by example: novice programming comes of age. *Communications of the ACM*, *43*(3), 75–81.

Snapp, M. (n.d.). *The workplace of today is not what it was 35 years ago. computational thinking is the new basic requirement.* Retrieved from `https://www.the74million.org/article/the-workplace-of-today-is-not-what-it-was-35-years-ago-computational-thinking-is-the-new-basic-requirement/`

Sneider, C., Stephenson, C., Schafer, B., & Flick, L. (2014). Computational thinking in high school science classrooms. *The Science Teacher*, *81*(5), 53.

Weintrop, D., & Wilensky, U. (2015). Using commutative assessments to compare conceptual understanding in blocks-based and text-based programs. In *11th Annual ACM Conference on International Computing Education Research, 2015.*

Werner, L., Denner, J., Campe, S., & Kawamoto, D. C. (2012). The fairy performance assessment: measuring computational thinking in middle school. In *Proceedings of the 43rd ACM technical symposium on Computer Science Education* (pp. 215–220).

Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, *49*(3), 33–35.

Wing, J. M. (2008). Computational thinking and thinking about computing. *Philosophical transactions of the royal society of London A: mathematical, physical and engineering sciences*, *366*(1881), 3717–3725.

Wing, J. M. (2010). *Research notebook: Computational thinkingwhat and why? the link magazine, spring.* Carnegie Mellon University, Pittsburgh.

Yadav, A., Hong, H., & Stephenson, C. (2016). Computational thinking for all: pedagogical approaches to embedding 21st century problem solving in k-12 classrooms. *TechTrends*, *60*(6), 565–568.

Yu, D. (2014). The Research on Teaching Reform of the University Computer Basic Course In Police Active Colleges. In *Proceedings of the 2nd International Conference on Education Technology and Information System (ICETIS 2014)* (pp. 583–586).

APPENDIX A. SURVEY INSTRUMENT

The section includes the survey questions taken from Qualtrics. The first page of the survey (figure A.1) introduced the study and also outlined the voluntary nature of participation as aligned by the granted IRB exempt (Appendix B).

PURDUE
UNIVERSITY

Introduction

Introduction

This anonymous survey seeks to gauge the effectiveness of the Computational Thinking Web Assessment Tool as part of the master's thesis work of Arjun Shakdher under the direction of Prof. Alka Harriger. It contains the following 6 sections: Age Appropriateness, Motivation level, System Usability, Usefulness, Satisfaction, and Content Appropriateness.

Your participation is completely voluntary. You can opt out at any time. All questions are optional. The data collected is non-identifiable. Only the researcher will have access to the data and all the data collected will be kept confidential.
The project's research records may be reviewed by the National Science Foundation, and by departments at Purdue University responsible for regulatory and research oversight.

How long will it take for completing the survey?
The survey contains 22 multiple choice questions and 3 short answer questions. The survey must be completed in one session. We estimate that completing the survey will take 7-10 minutes.

Who can I contact If I have questions about the survey?
If you have any questions, comments or concerns, you can contact Arjun Shakdher at

arjunshakdher@purdue.edu or +1 (765) 775-9226.

Figure A.1. Survey introduction page

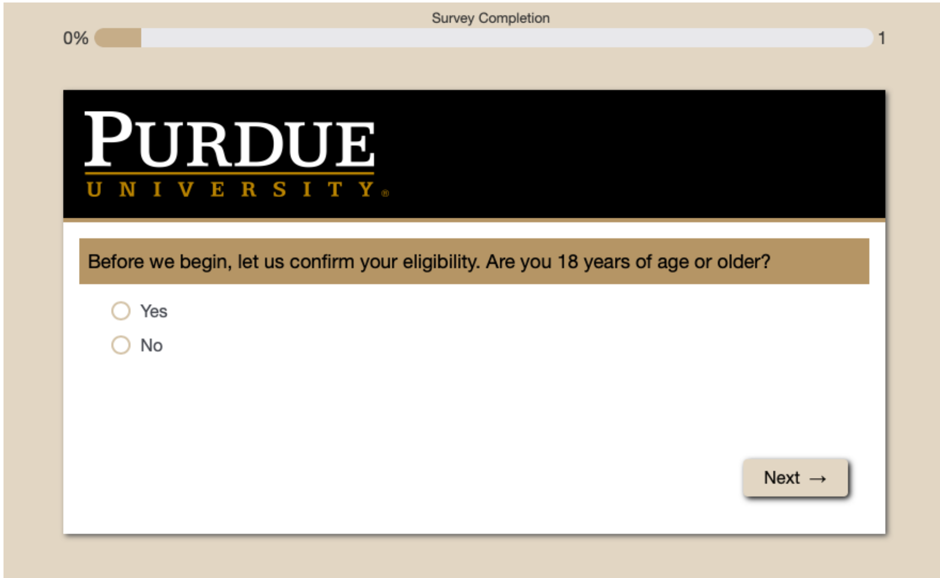


Figure A.2. Survey question 0: Age eligibility

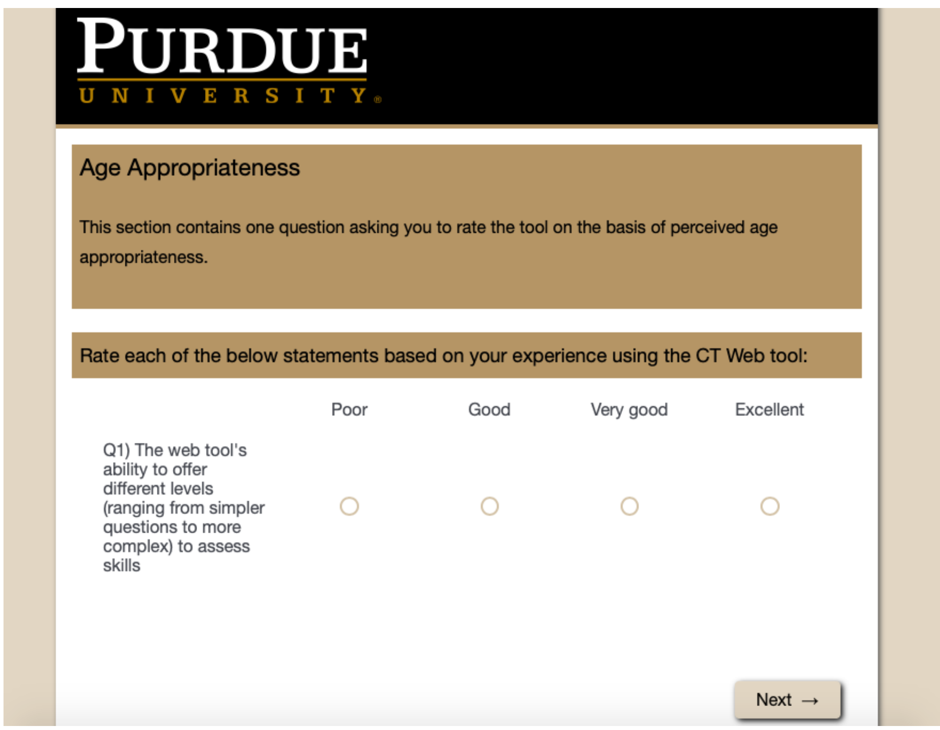


Figure A.3. Survey question 1: Perceived age appropriateness

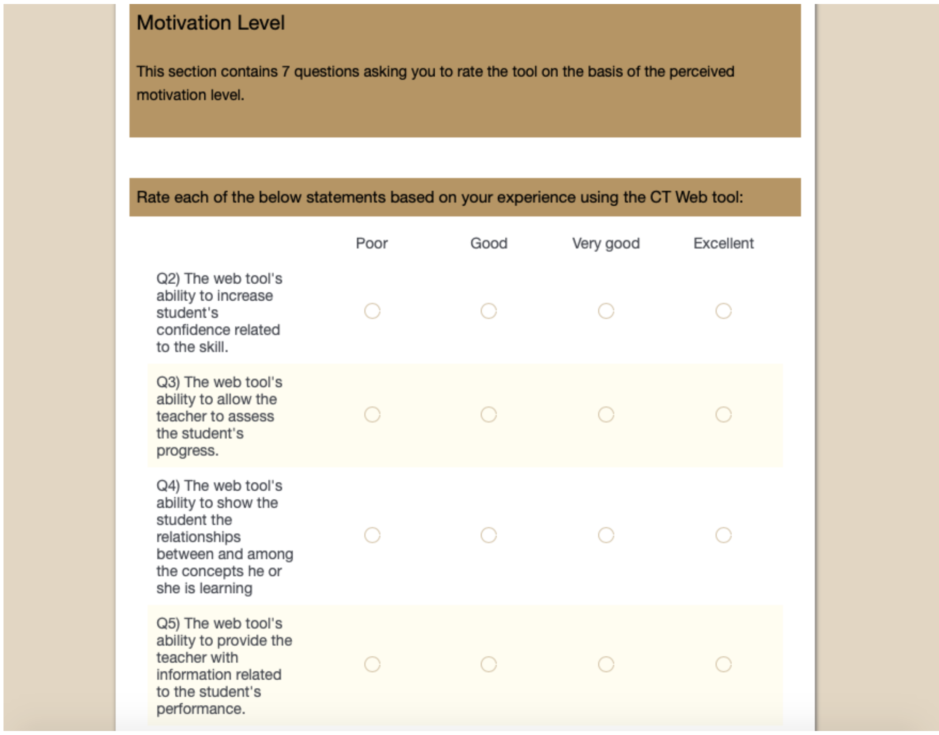


Figure A.4. Survey question 2-5: Perceived motivation level

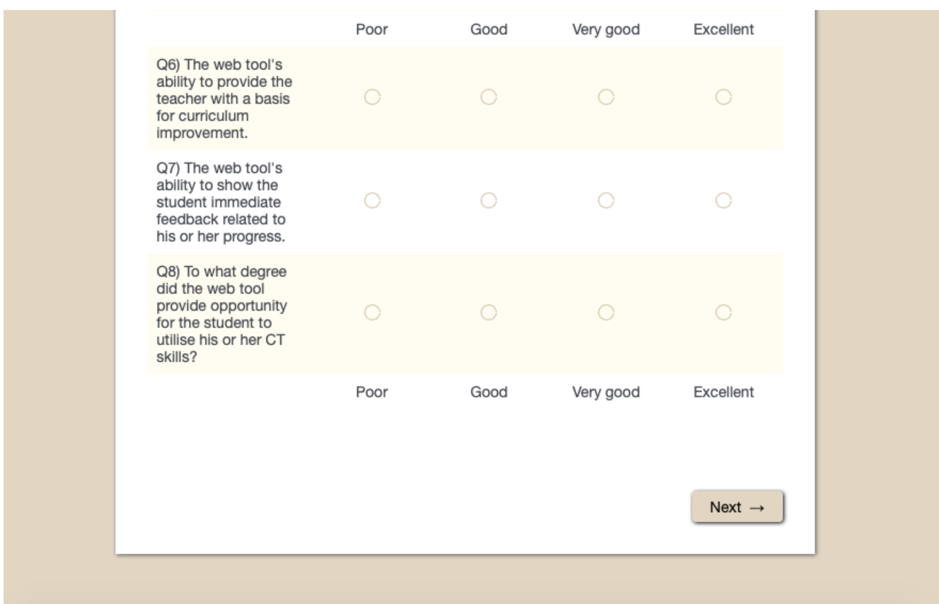
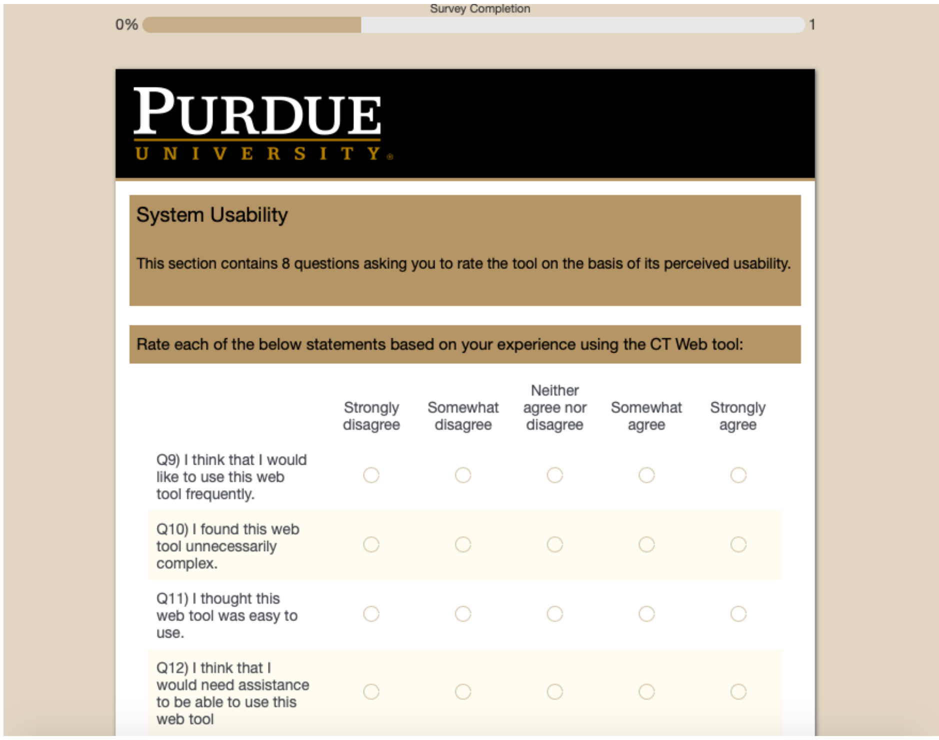


Figure A.5. Survey question 6-8: Perceived motivation level


Survey Completion

0% 1

PURDUE UNIVERSITY

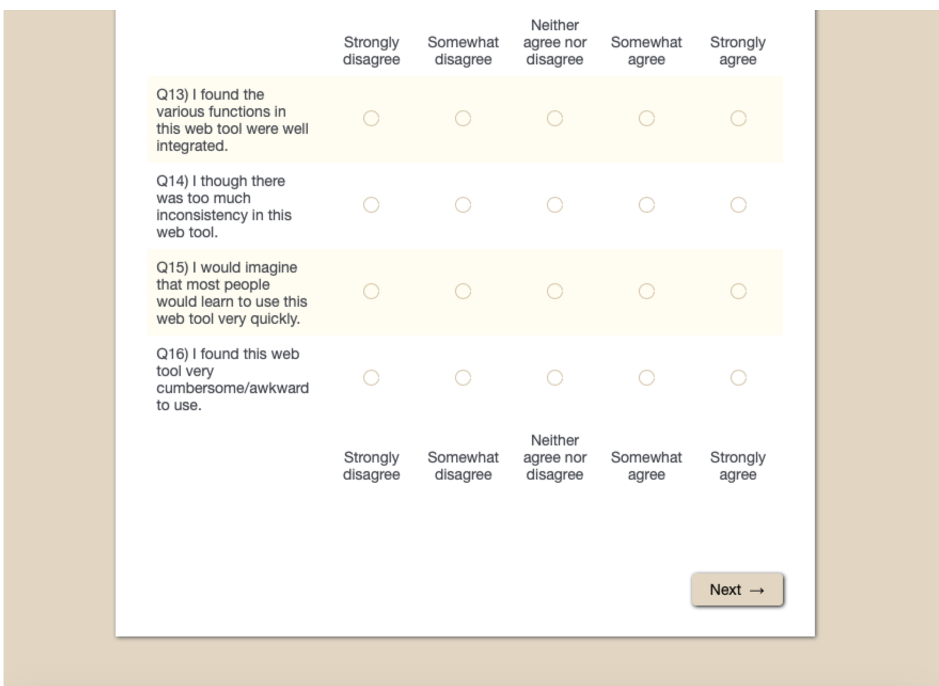
System Usability

This section contains 8 questions asking you to rate the tool on the basis of its perceived usability.

Rate each of the below statements based on your experience using the CT Web tool:

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
Q9) I think that I would like to use this web tool frequently.	○	○	○	○	○
Q10) I found this web tool unnecessarily complex.	○	○	○	○	○
Q11) I thought this web tool was easy to use.	○	○	○	○	○
Q12) I think that I would need assistance to be able to use this web tool	○	○	○	○	○

Figure A.6. Survey question 9-12: Perceived usability


	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
Q13) I found the various functions in this web tool were well integrated.	○	○	○	○	○
Q14) I though there was too much inconsistency in this web tool.	○	○	○	○	○
Q15) I would imagine that most people would learn to use this web tool very quickly.	○	○	○	○	○
Q16) I found this web tool very cumbersome/awkward to use.	○	○	○	○	○
	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree

Next →

Figure A.7. Survey question 13-16: Perceived usability

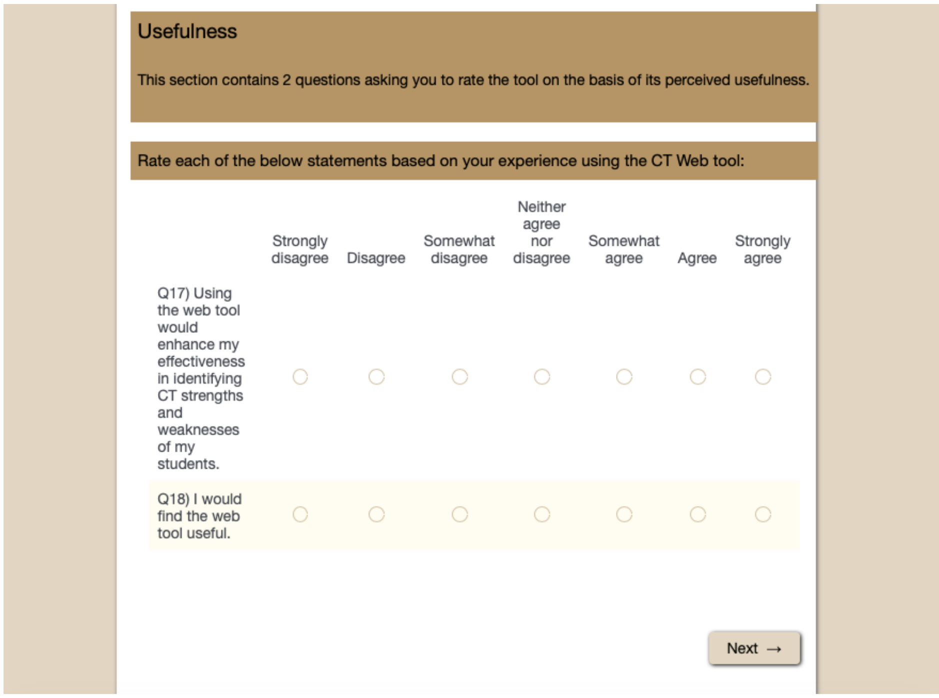


Figure A.8. Survey question 17-18: Perceived usefulness

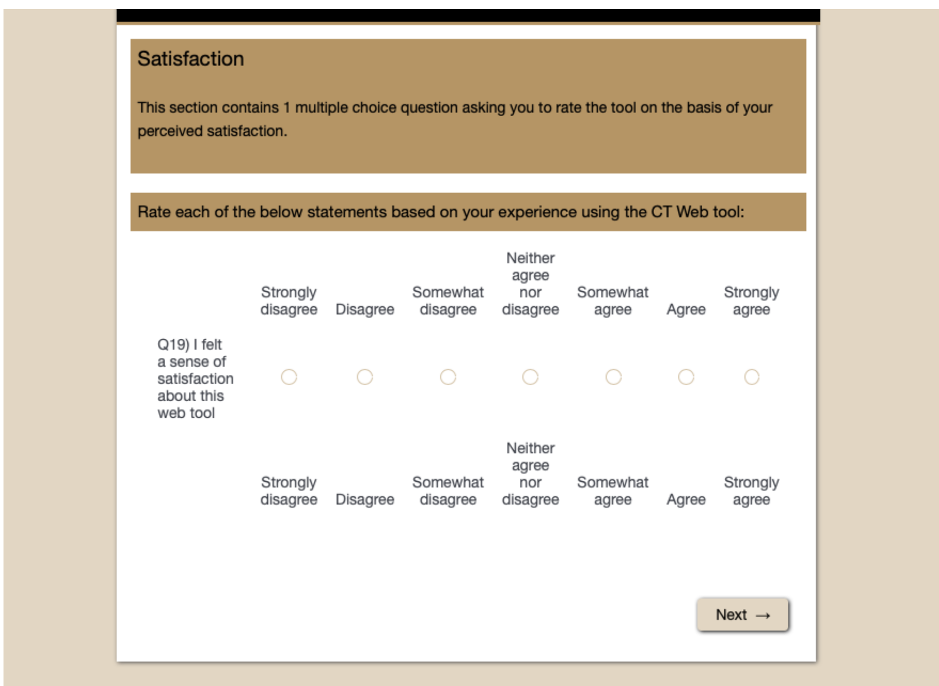
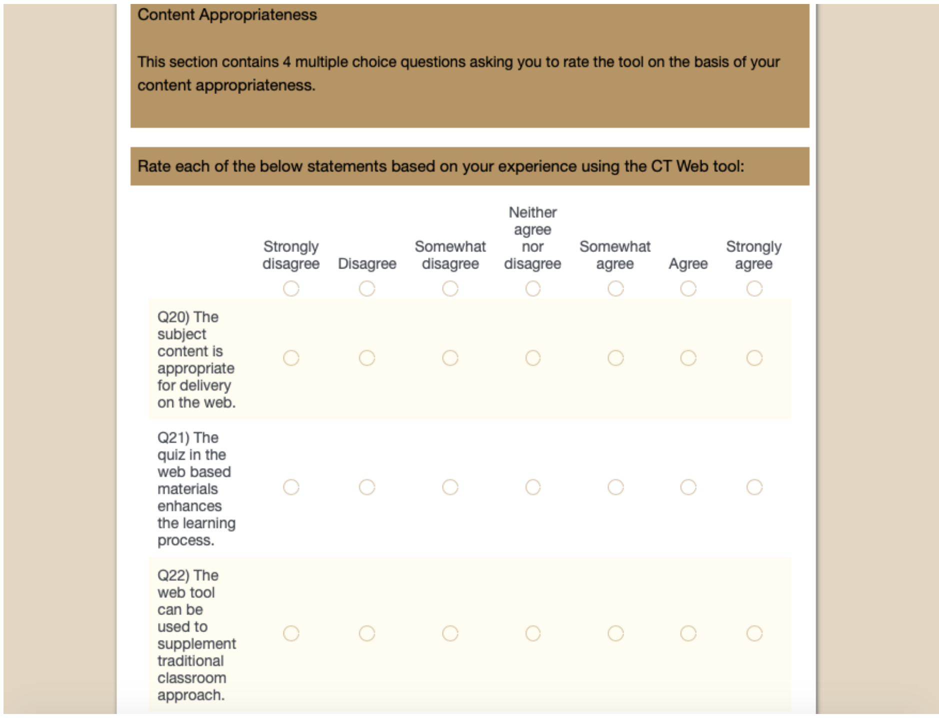


Figure A.9. Survey question 19: Perceived satisfaction



Content Appropriateness

This section contains 4 multiple choice questions asking you to rate the tool on the basis of your content appropriateness.

Rate each of the below statements based on your experience using the CT Web tool:

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
Q20) The subject content is appropriate for delivery on the web.	○	○	○	○	○	○	○
Q21) The quiz in the web based materials enhances the learning process.	○	○	○	○	○	○	○
Q22) The web tool can be used to supplement traditional classroom approach.	○	○	○	○	○	○	○

Figure A.10. Survey question 20-22: Perceived content appropriateness

PURDUE
UNIVERSITY

Additional Feedback

This section contains 3 open-ended questions asking you to provide any additional feedback you might have.

Q23) Any positive aspects of the tool you wish to highlight?

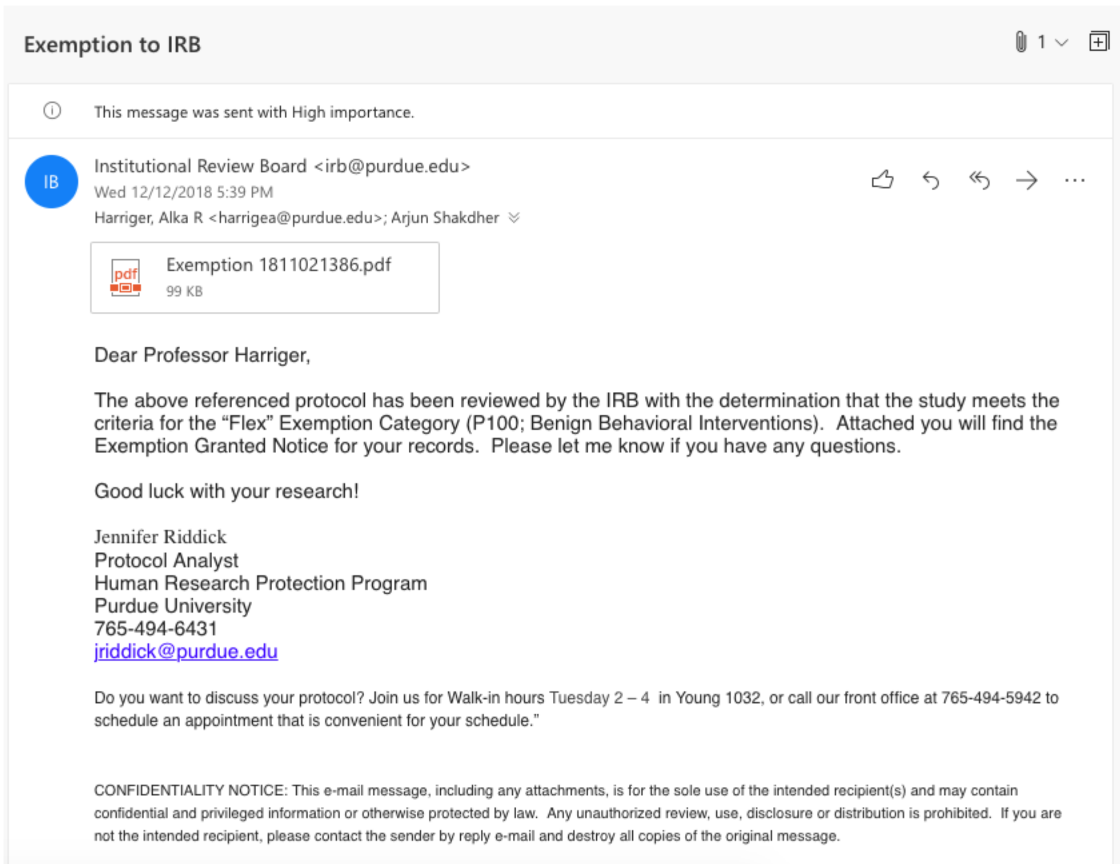
Q24) Any negative aspects of the tool you wish to highlight?

Q25) Please provide any additional comments about this tool.

Figure A.11. Survey question 23-25: Open-ended questions

APPENDIX B. APPROVAL - PURDUE INSTITUTIONAL REVIEW BOARD

The section includes the approval email from IRB (figure B.1) and the exemption approval letter on the next page. Purdue IRB protocol number: 1811021386



Exemption to IRB

This message was sent with High importance.

Institutional Review Board <irb@purdue.edu>
Wed 12/12/2018 5:39 PM
Harriger, Alka R <harrigea@purdue.edu>; Arjun Shakdher

Exemption 1811021386.pdf
99 KB

Dear Professor Harriger,

The above referenced protocol has been reviewed by the IRB with the determination that the study meets the criteria for the "Flex" Exemption Category (P100; Benign Behavioral Interventions). Attached you will find the Exemption Granted Notice for your records. Please let me know if you have any questions.

Good luck with your research!

Jennifer Riddick
Protocol Analyst
Human Research Protection Program
Purdue University
765-494-6431
jriddick@purdue.edu

Do you want to discuss your protocol? Join us for Walk-in hours Tuesday 2 – 4 in Young 1032, or call our front office at 765-494-5942 to schedule an appointment that is convenient for your schedule."

CONFIDENTIALITY NOTICE: This e-mail message, including any attachments, is for the sole use of the intended recipient(s) and may contain confidential and privileged information or otherwise protected by law. Any unauthorized review, use, disclosure or distribution is prohibited. If you are not the intended recipient, please contact the sender by reply e-mail and destroy all copies of the original message.

Figure B.1. Purdue IRB approval email

PURDUE
UNIVERSITY

HUMAN RESEARCH PROTECTION PROGRAM
INSTITUTIONAL REVIEW BOARDS

To:	HARRIGER, ALKA R
From:	DICLEMENTI, JEANNIE D, Chair Social Science IRB
Date:	12/12/2018
Committee Action:(P100)	Determined Exempt, Category **(P100)**
IRB Action Date:	12 / 12 / 2018
IRB Protocol #:	1811021386
Study Title:	Collaborative Platform for Computational Thinking Assessment

The Institutional Review Board (IRB) has reviewed the above-referenced study application and has determined that it meets the criteria for exemption under 45 CFR 46.101(b). Flex” Exemption Category (P100; Benign Behavioral Interventions).

Before making changes to the study procedures, please submit an Amendment to ensure that the regulatory status of the study has not changed. Changes in key research personnel should also be submitted to the IRB through an amendment.

General

- To recruit from Purdue University classrooms, the instructor and all others associated with conduct of the course (e.g., teaching assistants) must not be present during announcement of the research opportunity or any recruitment activity. This may be accomplished by announcing, in advance, that class will either start later than usual or end earlier than usual so this activity may occur. It should be emphasized that attendance at the announcement and recruitment are voluntary and the student’s attendance and enrollment decision will not be shared with those administering the course.
- If students earn extra credit towards their course grade through participation in a research project conducted by someone other than the course instructor(s), such as in the example above, the students participation should only be shared with the course instructor(s) at the end of the semester. Additionally, instructors who allow extra credit to be earned through participation in research must also provide an opportunity for students to earn comparable extra credit through a non-research activity requiring an amount of time and effort comparable to the research option.
- When conducting human subjects research at a non-Purdue college/university, investigators are urged to contact that institution’s IRB to determine requirements for conducting research at that institution.
- When human subjects research will be conducted in schools or places of business, investigators must obtain written permission from an appropriate authority within the organization. If the written permission was not submitted with the study application at the time of IRB review (e.g., the school would not issue the letter without proof of IRB approval, etc.), the investigator must submit the written permission to the IRB prior to engaging in the research activities (e.g., recruitment, study procedures, etc.). Submit this documentation as an FYI through Coeus. This is an institutional requirement.

Categories 2 and 3

- Surveys and questionnaires should indicate
  - only participants 18 years of age and over are eligible to participate in the research; and
  - that participation is voluntary; and
  - that any questions may be skipped; and
  - include the investigator's name and contact information.
- Investigators should explain to participants the amount of time required to participate. Additionally, they should explain to participants how confidentiality will be maintained or if it will not be maintained.
- When conducting focus group research, investigators cannot guarantee that all participants in the focus group will maintain the confidentiality of other group participants. The investigator should make participants aware of this potential for breach of confidentiality.

Category 6

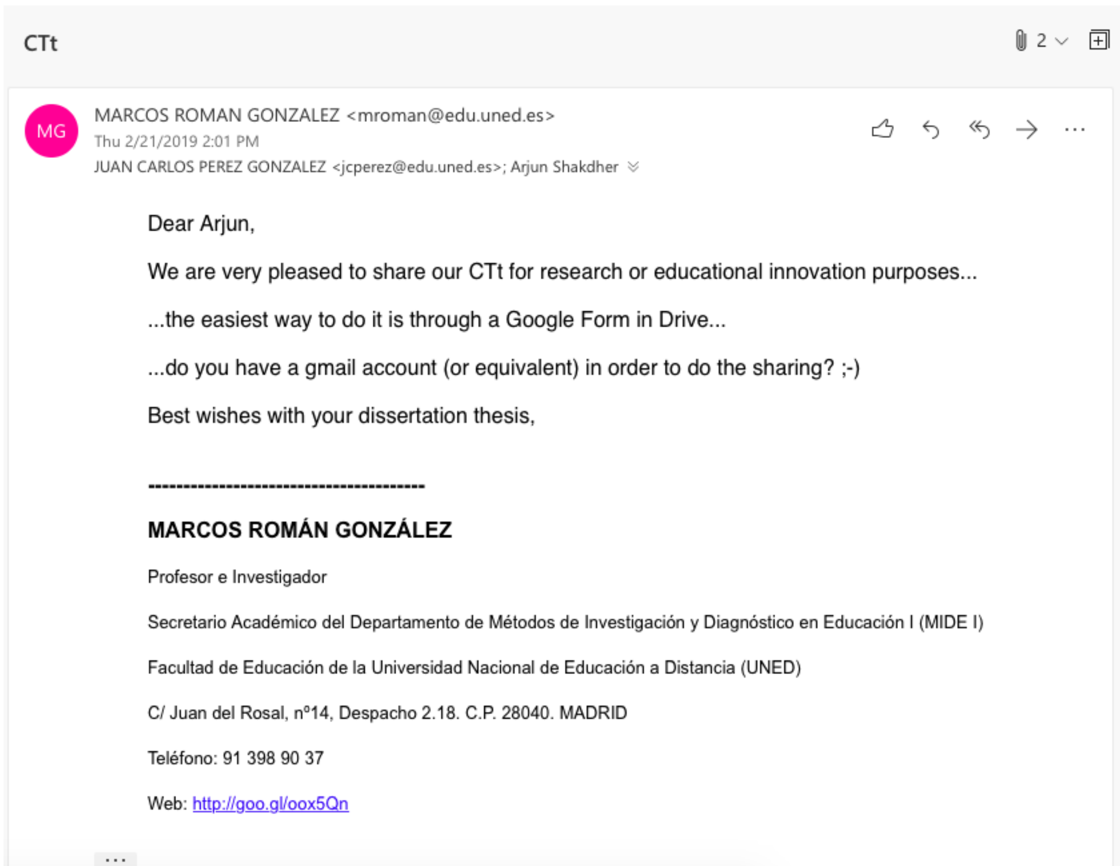
- Surveys and data collection instruments should note that participation is voluntary.
- Surveys and data collection instruments should note that participants may skip any questions.
- When taste testing foods which are highly allergenic (e.g., peanuts, milk, etc.) investigators should disclose the possibility of a reaction to potential subjects.

You are required to retain a copy of this letter for your records. We appreciate your commitment towards ensuring the ethical conduct of human subjects research and wish you luck with your study.

Ernest C. Young Hall, 10th Floor - 155 S. Grant St. - West Lafayette, IN 47907-2114 - (765) 494-5942 - Fax: (765) 494-9911

APPENDIX C. AUTHOR PERMISSIONS FOR QUESTION REPOSITORY

The section includes the email permission from the researchers (figures C.1 and C.2) of various questions used to populate the proposed tool's repository.



CTt

MARCOS ROMAN GONZALEZ <mroman@edu.uned.es>
Thu 2/21/2019 2:01 PM
JUAN CARLOS PEREZ GONZALEZ <jcperez@edu.uned.es>; Arjun Shakdher

Dear Arjun,

We are very pleased to share our CTt for research or educational innovation purposes...

...the easiest way to do it is through a Google Form in Drive...

...do you have a gmail account (or equivalent) in order to do the sharing? ;-)

Best wishes with your dissertation thesis,

MARCOS ROMÁN GONZÁLEZ

Profesor e Investigador

Secretario Académico del Departamento de Métodos de Investigación y Diagnóstico en Educación I (MIDE I)

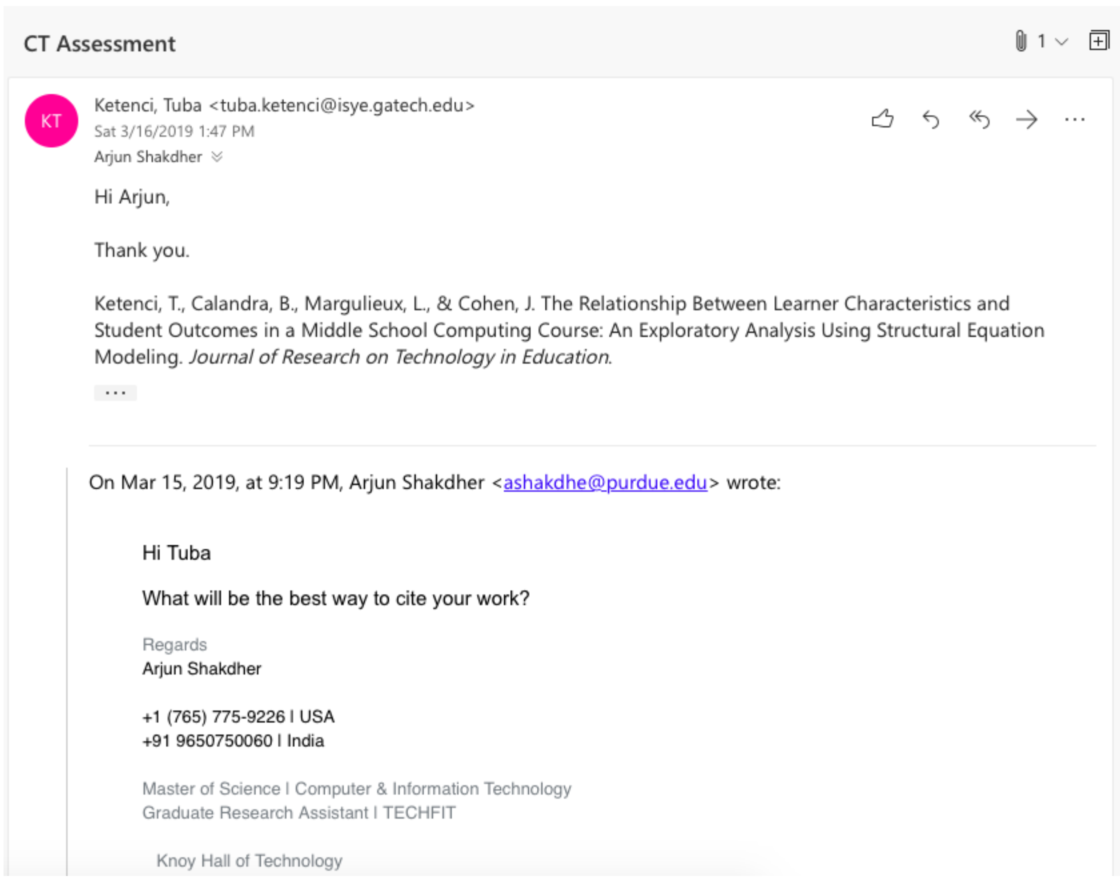
Facultad de Educación de la Universidad Nacional de Educación a Distancia (UNED)

C/ Juan del Rosal, nº14, Despacho 2.18. C.P. 28040. MADRID

Teléfono: 91 398 90 37

Web: http://goo.gl/oox5Qn

Figure C.1. Author permission 1



CT Assessment

Ketenci, Tuba <tuba.ketenci@isye.gatech.edu>
Sat 3/16/2019 1:47 PM
Arjun Shakdher

Hi Arjun,

Thank you.

Ketenci, T., Calandra, B., Margulieux, L., & Cohen, J. The Relationship Between Learner Characteristics and Student Outcomes in a Middle School Computing Course: An Exploratory Analysis Using Structural Equation Modeling. *Journal of Research on Technology in Education*.

...

On Mar 15, 2019, at 9:19 PM, Arjun Shakdher <ashakdhe@purdue.edu> wrote:

Hi Tuba

What will be the best way to cite your work?

Regards
Arjun Shakdher

+1 (765) 775-9226 I USA
+91 9650750060 I India

Master of Science I Computer & Information Technology
Graduate Research Assistant I TECHFIT

Knoy Hall of Technology

Figure C.2. Author permission 2