University of Southampton

Faculty of Social Sciences

Southampton Education School

The relationship between computational thinking performance and general achievement of secondary school students in Kazakhstan

by

Yerkhan Mindetbay

Thesis for the degree of PhD

September 2021

University of Southampton

Abstract

Faculty of Social Sciences

Southampton Education School

Doctor of Philosophy

The relationship between computational thinking performance and general achievement of secondary school students in Kazakhstan

by

Yerkhan Mindetbay

Computational thinking, a form of thinking and problem solving, is defined as a mental process for abstracting problems and formulating solutions. Computational thinking is considered to be an essential skill for everyone and has become the centre of attention in education settings. There is a limited number of tools to measure computational thinking skills by multiple-choice questions, and limited research on the relationship between computational thinking and other domains. The purpose of this research is to investigate the relationship between computational thinking performance, perception of computational thinking skills and school achievement of secondary school students. Computational thinking performance of secondary school students in Kazakhstan is measured by using a bespoke multiple-choice test, which focuses on the following elements of computational thinking: logical thinking, abstraction and generalisation. The perceptions of computational thinking skills are self-reported using a pre-existing questionnaire, which covers the following factors: creativity, algorithmic thinking, cooperation, critical thinking and problem-solving. The General Knowledge Test results that contain scores for 14 different subjects are used as indicators of students' school achievement, with further sub-scores for the science subjects, language subjects and humanities. The sample group of 775 grade eight students are drawn from 28 secondary schools across Kazakhstan. The validity and reliability of the multiple-choice questions are established by using Item Response Theory models. The item difficulty, discrimination and guessing coefficients are calculated; and the item characteristic curves for each question and test information functions for each quiz are obtained. As a result, the multiple-choice questions are concluded as a valid and reliable tool to measure the computational thinking performance of students. Multiple regression is used to examine the relationship between computational thinking performance, perception of computational thinking and school achievement sub-scores. The results of the data analysis show that science subjects, language subjects and perception of computational thinking skills are significant predictors for computational thinking performance, showing a moderate relationship between computational thinking performance and school achievement. However, no significant relationship is found between humanities subject scores and computational thinking performance. This study also adds to the literature for the studies that investigate the relationship between computational thinking skills and other variables. This research contributes to the development of validated tools to measure computational thinking performance by using multiple-choice questions. This study investigates the relationship between computational thinking performance and general school achievement of secondary school students, and its findings shed light on the measurement of children's cognitive development. The findings can help in designing better curricula by adjusting

subjects that enhance children's higher-order thinking abilities. The findings obtained in this research also adds to the literature for the studies that investigate the relationship between computational thinking skills and other variables.

Table of Contents

Table of Tables

Table of Figures

Research Thesis: Declaration of Authorship

Print name: Yerkhan Mindetbay

Title of thesis: The relationship between computational thinking performance and general achievement of secondary school students in Kazakhstan

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as:-
   Mindetbay, Y., Bokhove, C., & Woollard, J. (2019a). The Measurement of Computational Thinking Performance Using Multiple-choice Questions. In *Proceedings of International Conference on Computational Thinking Education 2019* (pp. 176–179). Hong Kong: The Education University of Hong Kong.
8. Mindetbay, Y., Bokhove, C., & Woollard, J. (2019b). What is the Relationship between Students' Computational Thinking Performance and School Achievement? *International Journal of Computer Science Education in Schools*, *March* (Working Group), 3–19. doi:https://doi.org/10.21585/ijcses.v0i0

Signature: Date: 27.05.2021

Acknowledgements

My sincere gratitude goes to my supervisors, Dr Christian Bokhove and Dr John Woollard, for their support, guidance and encouragement, who helped me to become a better version of myself throughout my PhD journey.

I extend my thanks to all my friends and colleagues at Southampton Education School.

I would also like to thank all the participants and their teachers for taking part in my research study.

My special thanks to the Ministry of Education and Science of the Republic of Kazakhstan for providing me with the Bolashak scholarship.

Last, but by no means least, I would like to thank my family, my beautiful wife Gulnar, for her support, patience and ultimate belief in my quest to complete a Doctoral thesis, and my sons, Sani and Ali for adding meanings to my life.

Definitions and Abbreviations

BIL..Bilim Innovation Lyceums are a set of selective schools under the supervision of Bilim Innovation Foundation in Kazakhstan.

CTS.......................................Computational Thinking Scale is an adapted scale for secondary school students developed for determining the levels of computational thinking skills. CTS is a five-point Likert type scale that consists of 22 items, which covers five factors such as creativity, algorithmic thinking, cooperation, critical thinking and problem-solving. CTS questionnaire is a self-reported measurement.

GKT ....General Knowledge Test is a multiple-choice test that covers 14 subjects: algebra, geometry, physics, chemistry, biology, computer science, English language, Kazakh language, Kazakh literature, Russian language, Turkish language, world history, history of Kazakhstan and geography, which is taken every term that is four times a year at Bilim Innovation Lyceums in Kazakhstan.

ICILS.....................................The International Computer and Information Literacy Study is a computer-based international assessment of 8th-grade students' capacities to use information communications technologies (ICT) productively. ICILS is sponsored by the International Association for the Evaluation of Educational Achievement (IEA).

iSurvey.................................iSurvey is a survey generation and research tool for distributing online questionnaires (www.isurvey.soton.ac.uk).

MCQ ....Multiple-choice questions are objective assessment questions in which respondents are asked to select only the correct answer from the given choices as a list.

NTC......................................The National Testing Centre is a national organisation under the Ministry of Education and Science of the Republic of Kazakhstan, established in 1992. The main goal of the centre is the organising and conducting of all types of external examinations and monitoring in the education system of the Republic of Kazakhstan.

UNT ....The Unified National Test is one of the forms of qualifying examinations for admission to organisations of higher and (or) postgraduate education.

Chapter 1 Introduction

1.1 Statement of Problem

In recent years, the integration of computational thinking into the school curriculum has been actively discussed among educational institutions and academia (Furber, 2012; Heintz, Mannila, & Farnqvist, 2016; Wing, 2011). This movement is being driven by educators and scholars who advocate the educational benefits of teaching programming which enhance the development of a key ability for the 21st century, computational thinking skills (Moreno-leon, Román-González, & Robles, 2018). The term gained wide popularity among researchers and institutions with the publication of a paper “Computational Thinking” by Wing in 2006. Institutions, educators and researchers work on promoting computational thinking skills via teaching coding skills, beginning from primary school up to the university level (Selby, 2014; Vahldick, Mendes, Marcelino, & Roberto, 2016). Computational thinking affects the research areas of both humanity and the natural sciences (Bundy, 2007). Especially in every science and mathematics-related field, such as bioinformatics, computational statistics, computational linguistics, digital humanities, and neuroinformatics, has witnessed a growth in computational methods and solutions. Wing (2010) claims that a person with computational thinking skills has an advantage over others without. Programming is considered by some as an important activity that supports learning to think computationally (Kong & Li, 2016; Lye & Koh, 2014). There are growing numbers of technologies that support teaching and learning computer programming, such as Logo (Papert, 1980), Lego Mindstorms (Petre & Price, 2004), Alice (Werner, Denner, & Campe, 2015) Scratch (Moreno-Leon, Robles, & Román-González, 2016), Greenfoot (Cooper & Dann, 2015; Grover & Pea, 2013a; Utting, Cooper, & Kölling, 2010), Arduino kits and Raspberry Pi kits (Buechley, Eisenberg, Catchen, & Crockett, 2008; Grover & Pea, 2013a; Leonard et al., 2016; Sobota, Pisl, Balda, & Schlegel, 2013; Vallance & Towndrow, 2016). There are few studies closely related to the assessment of computational thinking and its relationship with other indicators (Brennan & Resnick, 2012; Durak & Saritepeci, 2017; Koh, Basawapatna, Nickerson, & Repenning, 2014; Moreno-Leon et al., 2016; Werner, Denner, Campe, & Kawamoto, 2012; Yadav, Zhou, Mayfield, Hambrusch, & Korb, 2011). When wisely used, computational tools and computational skills can support learning mathematics and science subjects and their content (Sengupta, Kinnebrew, Basu, Biswas, & Clark, 2013; Wilensky, Brady, & Horn, 2014). Assessment of computational thinking following curriculum and standards needs to be studied extensively (Ambrosio, Xavier, & Georges, 2015; Grover, 2015; Jamil, 2017b). A valid assessment is required to measure students’ computational thinking skills (Grover & Pea, 2013a). A lack of appropriate measuring tools that could assist both learners and educators in assessing the learning process (Moreno-Leon et al., 2016) is an important issue to be considered. Therefore, the question “How to evaluate if these computational thinking skills help students perform better?” arises (Grover & Pea, 2013a). Particularly, when evaluating programming activities, Weinberg (1998) states that most studies on the psychology of programming focus on “what” to measure, rather than “how” to measure. The importance of “what” to measure uncovers more significant issues. In the present Kazakhstani new curriculum, “Computational thinking” is presented as one separate section among Computer systems, Information processes and Health and safety (National Academy of Education, 2016). In this updated curriculum, what concepts of computational thinking should be measured and in what way they should be assessed are important points to identify. Valid assessment tools can provide solid evidence to see to what extent computational thinking skills of secondary school students are related to their school performance. A limited number of studies investigates the relationship between computational thinking and other variables. Some critics over computational thinking state that exaggerated claims such as the universal value of computational thinking should be softened (Denning, 2017), as there is no evidence of transfer of knowledge from computer science to daily lives (Guzdial, 2015). Transfer can be defined as the learning process when a person learns to use previously acquired knowledge/skills/competence/expertise in a new situation(Eraut, 2004). The question of transfer from one area to another has been a topic of dispute for a long time. The transfer can be near or far, where near is a transfer to a similar context and far is a transfer to a different context (Perkins & Solomon, 1992). Although, how far the learned skills can be transferred is understood as the degree of difference between the initial and the final tasks, the transfer distance cannot be reduced to a matter of degree of difference between these two (De Corte, 2003). The ability to see the similarities, patterns and differences is a key aspect of transfer (Marton, 2006). Perkins and Solomon (1992) describe the transfer of learning as a positive transfer or a negative; positive transfer is seen when learning in one area strengthen the performance in a different area and a negative transfer is when learning weakens the performance. How computational thinking is correlated with other areas is being studied in various sample groups. There is a correlation found between computational thinking skills of university students and their performance at a computer science course Gouws et al. (2013). However, little research has been carried out to examine the relationship between computational thinking skills and other domains, such as personality, non-cognitive psychological construct, gender, thinking styles and academic success at mathematics (Durak & Saritepeci, 2017; Román-González, Pérez-González, & Jiménez-Fernández, 2016). The practical implementation of computational thinking, its assessment in a learning environment (Brennan & Resnick, 2012) and linking them with the topics in curricula is crucial in bringing computational thinking into the classroom (Yadav, Hong, & Stephenson, 2016a). Studying the relationships between problem-solving, computational thinking, and programming may help improve curriculum, classroom activities, and finally clarify the understanding of the 21st skills (Selby, 2012). There is a need to work on the pedagogical and educational roles of computational thinking, specifically, what should be taught, and how to assess computational thinking (Tedre & Denning, 2016), particularly in different educational levels (Mannila et al., 2014). Most of the research studies carried out until today are small scale or descriptive qualitative and there is a need for large-scale studies about computational thinking (Czerkawski, 2018).

Although the advantage of computational thinking skills in computer science-related fields is studied and reported, critics are stating that computational thinking only works in computations and such a big claim that computational thinking is good for everyone (Wing, 2006) is not justified (Denning, 2017). It might be assumed that students' school achievement is somehow related to their thinking skills or general intelligence. However, it would not be appropriate to claim that there is such a relationship without any research findings. Thus, there is a need for a study that investigates the relationship between computational thinking skills and other subjects, areas and domains. The biggest questions around computational thinking are What should be taught on different levels? (Mannila et al., 2014) How should it be taught?, and How to measure it? (Denning, 2017; Mannila et al., 2014; Tedre & Denning, 2016). Computational thinking has become the centre of attention in the largescale assessment such as PISA (OECD, 2018) and ICILS (2020). This study investigates not only "How to measure computational thinking skills?" but also the relationship between school achievement. Chuang et al. (2015) stated that by the time students are in grade 7, they develop their logical thinking and problem-solving skills. This 7th grade age range reflects the final stage of cognitive development proposed by Piaget, the Formal Operational Stage that claims 11 years old and over children develop their ability to logically test hypothesis and think about abstract concepts (Flavell, 1963; Piaget, 1958). There is also a consensus on computational thinking strands that data representation and abstraction are to be taught in grade 7-9, as children learn to program (Chuang et al., 2015). Likewise, according to the informatics curriculum in Kazakhstan, 7-8 grade secondary school students are introduced to CISCO's "Get connected" course, logic gates, algorithms with flowcharts using a program FCPro, block programming with Scratch, and spreadsheets. There are only a few studies found that explore the relationship between computational thinking and the general academic achievement of secondary school students. This research study investigates how Kazakhstani 8th grade students' computational thinking performance is related to their school achievement. This thesis explores the relationship between students' school achievement in three subcategories of subjects: the sciences, languages and humanities, and students' perception of their computational thinking skills. The study seeks to what extent there is an interaction between students' computational thinking perceptions/performance and their academic achievement. The findings of this research help to better understand the relationship between computational thinking skills and school achievement. The implications of this study can help in designing better curricula to support children's cognitive development by adjusting subjects that enhance children's higher-order thinking abilities. There is a lack of validated instruments to measure computational thinking by using non-tool-specific instruments. And this research contributes to the development of validated measurement tools for computational thinking skills by showing the validity and reliability of using multiple-choice questions. The findings of this study also contribute to the Quantum project by providing evidence for the future evaluation of the assessment processes.

1.2 Research Questions

This research study aims to examine the computational thinking performance of 8th-grade secondary school students and to explore the relationship between computational thinking performance and general school achievement. The study findings provide an overall picture of students' computational thinking level, the measurement of the computational thinking performance using multiple-choice questions, students' perception of computational thinking and the relationship between the school achievement.

To measure the computational thinking performance of the students by multiple-choice questions an objective measurement is required, therefore the initial research question is formulated.

1. How to measure the computational thinking performance of secondary school students by multiple-choice questions?

2. The second research question addresses the relationship between computational thinking performance and the general school achievement of students. To investigate the relationship between the computational thinking performance of secondary school students in Kazakhstan and their general achievement, the following research questions are formulated.

    a. To what extent is there a relationship between the computational thinking performance as measured by the multiple-choice questions and the science-subjects achievement of the secondary school students in Kazakhstan as measured by the General Knowledge Test?

    b. To what extent is there a relationship between the computational thinking performance as measured by the multiple-choice questions and the language-subjects achievement of the secondary school students in Kazakhstan as measured by the General Knowledge Test?

c. To what extent is there a relationship between the computational thinking performance as measured by the multiple-choice questions and the humanities-subjects achievement of the secondary school students in Kazakhstan as measured by the General Knowledge Test?

Another important aspect is the perception of computational thinking skills and it is addressed in the last research question 3.

3. Is the perception of the computational thinking skills a predictor of the computational thinking performance of the secondary school students?

1.3 Context

Kazakhstan is a democratic and constitutional republic ("The Constitution of the Republic of Kazakhstan") with a diverse cultural heritage located in the heart of the Eurasian continent. It is the world's largest landlocked country, at 2 724 900 km^2, larger than all Western Europe, which borders five countries - Russia, China, Kyrgyzstan, Uzbekistan, and Turkmenistan. Kazakh and Russian are two major spoken and written languages, and there are over 100 ethnic groups in Kazakhstan. According to the statistics by the Agency for Strategic planning and reforms of the Republic of Kazakhstan, Bureau of National statistics the population is 18.7 million.

Since 1991, when Kazakhstan declared its independence, many reforms in the education system have taken place to improve the status, quality and structure of schooling (Beisenbayeva & Gelişli, 2016). Education, health and well-being of Kazakhstani citizens have been among the seven long-term priorities of the Kazakhstan-2030 Strategy for development addressed by the president of the Republic of Kazakhstan, Nursultan Nazarbayev, in October 1997 (Nazarbayev, 1997). The official language in Kazakhstan is Kazakh. However, as Kazakhstan was a part of the former USSR, the Russian language is also officially used alongside the Kazakh language in the state organisations and bodies of local self-government ("The Constitution of the Republic of Kazakhstan"). Therefore, the education system consists of bilingual curricula – Kazakh and Russian at primary and secondary schools and Higher Education Institutions. According to the data of the Agency of Statistics from 2018, the total number of students reached 2 972 200, among them, the number of students at Kazakh language schools is – 19 608 000 (66 %), whereas, the number of students at Russian language schools is – 9 095 000 (30.6 %). The rest of the students are 1 019 000 (3.6 %) at ethnic minority schools who also use their ethnic languages for classroom instruction: Uzbek, Uyghur, Tadjik, etc. (Ministry of National Economy of the Republic of Kazakhstan Committee on Statistics, 2018). The language of instruction makes a difference in students' school achievement in Kazakhstan (OECD, 2015). The State Programme of Education Development (SPED) for 2011- 2020 is acting as a foundation document driving education reform in the country (Almaty, 2004; Burkhalter & Shegebayev, 2012; MESRK, 2014). This state program is the basis for implementing the public policy in the education system that guarantees the educational modernisation of the country (Ministry of Education and Science of the Republic of Kazakhstan, 2010). In Kazakhstan, the academic year begins at the beginning of September and ends at the end of May; the overall period is nine months long with holidays included (except summer holiday) and divided into four terms. The secondary education system in Kazakhstan includes preschool and secondary levels as shown in Table 1. The main criteria for distributing students by grade is age. Secondary education consists of primary education (Grade 1-4), basic secondary education (Grade 5-9) and upper secondary education (Grade 10-11). One-year pre-primary education and secondary education are free and compulsory (EPDC, 2014; Ministry of Education and Science of the Republic of Kazakhstan, 2014; The Ministry of Education and Science of the Republic of Kazakhstan-UNESCO Cluster Office, 2004). In order to maximise the match of the resources to individual student learning needs, students are organised into learning groups within schools and grades. After students are distributed into grades according to their ages at schools, students choose one of the two main groups according to the instruction language, the Kazakh and the Russian. Once students reach the upper secondary education level (Grade 10-12), the formation of classes follows one of the two direction. These two directions are the natural-mathematical and social-humanitarian (Ministry of Education and Science of the Republic of Kazakhstan, 2014). Optional subjects and their number of hours per term differ according to the chosen direction; the natural-mathematical direction is more science-oriented and the social-humanitarian direction is humanities oriented. In Kazakhstan, schooling is mandatory for all children from the age of 6-7 to 15-16 and it is provided free of charge ("The Constitution of the Republic of Kazakhstan"). The law on Education regulates social relations in the field of education for citizens of the Republic of Kazakhstan, foreigners, and persons without citizenship permanently residing in the territory of Kazakhstan.

Table 1 Secondary education in Kazakhstan

Levels	Age range
• **Primary (1-4 grades)** • **Basic (5-9 grades)** • **Upper (10-11 grades)**	• 6 (or 7) • 10 (or 11)-14 (or15) • 16-18

1.3.1 National examinations

The External Assessment of Academic Achievement (EAAA) is currently the main assessment tool in the Kazakhstani education system for measuring students' learning outcomes (OECD, 2020b). Although the EAAA has been designed to cover some of the issues that the previous national assessment system could not address, there are concerns that the EAAA still lacks to fit the requirements of the current national curriculum; it also cannot be used inappropriately as an indicator of the quality of schools in Kazakhstan. Therefore, improvements or changes in the national assessment system of the Kazakhstani education system is expected by the Ministry of Education and Science of the Republic of Kazakhstan (OECD, 2020a). Formal education in Kazakhstan practices several ways of assessing learning outcomes such as state examination at the end of the 9th grade, the Unified National Test (UNT) and the Final Attestation. By the end of the 9th grade, students must take a state examination at their local schools in four subjects. Three of these subjects are compulsory and one is optional. The compulsory subjects are literature, math, and Kazakh or Russian language depending on the language instruction. For the classes with Russian or other language instruction, the Kazakh language is mandatory, likewise, the Russian language is mandatory for the classes with the Kazakh language of instruction. The optional subject can be chosen among the following: physics, chemistry, biology, geography, geometry, informatics, foreign language and literature, world history, and the history of Kazakhstan (EGOV.KZ, 2021). Informatics is among the optional subjects in the state examination system by the end of 9th grade. Further, in upper-level education informatics is no longer found among the subjects in the state examination system. The National Testing Centre (NTC) monitors the performance of pupils before they take the Unified National Test after the completion of 11th-grade. The National Testing Centre conducts online trial tests for 9th-grade students. There are two subjects: compulsory Kazakh language and one optional subject. The optional subject can be chosen from one of the following subjects: world history, history of Kazakhstan, biology, geography, Kazakh literature (with the Kazakh language of instruction) / Russian literature (with the Russian language of instruction), algebra, chemistry, physics, Russian language (with the Kazakh language of instruction) / Russian language (with the Russian language of instruction), English (National Testing Centre, 2017a). Informatics subject is not on the list of optional subjects for this final stage of examination.

Students must undertake the Unified National Testing (UNT), a test for obtaining state grants, to enter higher education institutes in Kazakhstan. To be eligible for higher education grants, test-taking students must get a minimum of 50 points out of 140. Since 2017 students who finish secondary school take the new format Unified National Test at the test centres. Up until the introduction of the new format of the UNT, the old UNT results played a major role in school rankings, which led to various disputes and criticism of the validity of the UNT, which raised the question “Is the UNT system supporting the national curriculum or is the national curriculum chasing the UNT to fit its assessment system?”. Another issue that was raised regarding the old UNT system is the fact that school administrations, teachers and parents perceived the UNT only as an entrance examination to HEIs. The introduction of the new structure and admission process provides a better assessment system and eliminated these misperceptions (Approved by Decree of the Government of the Republic of Kazakhstan of 21 December 2007 № 1253, №568 change of 3 May 2012). Alimkulov, the director of the National Testing Centre reported that the new UNT kept its previous multiple-choice format with more emphasis on logical reasoning and strived for parity with the international standards (Parkhomenko, 2017). The UNT system introduced additional multiple-response items to the previous multiple-choice items, in which there might be more than one correct answer for each question. As the content of the UNT changes regularly these changes do not allow for reliable comparisons over time. The number of test-takers at the UNT varies each year. Each year, almost 30 per cent of the final-year school students do not take part in the UNT (Ministry of Science and Education, 2017; National Testing Centre, 2017b). According to Kazakhstan’s Committee on Statistics, 80 000 students passed the UNT and 53 785 students were given state grants to continue their education in higher education institutes in Kazakhstani universities in 2020. The UNT is also limited in assessing test-taking students’ achievement concerning Kazakhstani standards as it does not suit the national curriculum, and other changes are expected in the UNT system such as new types of items in examinations (OECD, 2020a). Therefore, it can be stated that although the UNT as an assessment system has a major role, its results cannot reflect objective conclusions about the quality of secondary education in Kazakhstan.

1.3.2 Bilim Innovation Lyceums

With the support of the two presidents, Nursultan Nazarbayev (Kazakhstan) and Turgut Ozal (Turkey), two Kazakh-Turkish Lyceums (known as KTL) were opened in 1992 (Gaipov, Yaylaci, Çiğ, & Guvercin, 2013). Later, the International Public Foundation was established in 1997, which was rebranded in 2016 as the Bilim Innovation Foundation. The schools that are under the leadership of the Bilim Innovation Foundation are known as Bilim Innovation Lyceums (BIL). There are over 30 Bilim Innovation Lyceums located in all regions of Kazakhstan (Abylkasymova, Nurmukhamedova, Nurbaeva, & Zhumalieva, 2016; McCarthy, 2016). As these Bilim Innovation Lyceums are selective schools, pupils who want to enter these schools apply to these BILs by the end of the 6^{th} grade; and follow the admission procedures. The number of admission places are limited and differs for each Bilim Innovation Lyceum (Gaipov et al., 2013). The entrance examination to Bilim Innovation Lyceums consists of two phases. Those applicants who successfully pass the first and the second phase of the entrance examination are admitted to Bilim Innovation Lyceums. The following characteristics distinguish BILs from other public schools: An integrated curriculum, which embraces both national and international education courses. The rich environment of innovative technology and its use, and science-oriented curriculum enriches the BILs education system. Classes are conducted in Kazakh, Russian and English languages at Bilim Innovation Lyceums ("Bilim Innovation Lyceum," 2021). Some other distinctive features of the Bilim Innovative Lyceums are the integration of the content of curricula subjects of the natural-mathematical cycle and interdisciplinary integration, an expanded educational programs and the use of innovative technologies. These schools are leaders in utilising new technology and techniques, such as Flipped Classroom; and BILs run various Project-Based Learning (PBL) and other innovative techniques in education as a means to gain 21st Century skills (Koshegulova & Mindetbay, 2020). PBL helps BILs students to gain skills such as collaboration, communication, critical thinking and creativity. As reported by the vice-president of Human Resources and Professional Development at Bilim Innovation Foundation, Kojabekov (as cited in McCarthy, 2016), among these four essential skills, critical thinking skills help students to fully perform different tasks in their future life. Kojabekov mentioned the PISA 2015 test results, where Kazakhstan was around 40th and emphasizing that the memorisation of facts would not help students in their career, once again underlining the importance of critical thinking skills (McCarthy, 2016). From ministry to local schools, several bodies are responsible for enhancing the critical thinking skills of children (Burkhalter & Shegebayev, 2012). And the PISA 2018 results show that Kazakhstan is 69th in reading, mathematics and science that is lower than the OECD average. Particularly, in reading only 36% of participants reached Level 2 proficiency, which means they can identify the main idea in given moderate length text with some complex criteria. And only a few percentages of Kazakhstani participants could take place among the top performers in reading, showing that they reached Level 5-6, which means students can understand longer texts, deal with abstract concepts and distinguish the facts and opinions. In mathematics, 51% of Kazakhstani participants reached Level 2, showing they can interpret and are familiar with how to represent mathematically simple situations without relying on direct instructions. And only 2% of participants reached Level 5, which show they can choose, compare and evaluate problem-solving strategies, and also model complex situations mathematically. In science, 40% of Kazakhstani participants reached Level 2, which means these participant students can identify the correct explanation for scientific phenomena that they are familiar with. Only a negligible percentage of participants were among the top performers in science reaching Levels 5-6, which shows that these participant students can apply their knowledge of science to a wider variety of situations even in unfamiliar circumstances (OECD, 2019). These PISA 2018 results point to the areas where students need help to improve. Any assessment, whether it is formative or summative, should assist teachers to support students' learning.

It is stated that there is no external assessment of pupils' performance before they take the Unified National Test except the External Assessment of Academic Achievement (EAAA) (OECD & World Bank, 2014). However, there is a centralised system of assessment of performance at Bilim Innovation Lyceums (BILs) administered by the Bilim Innovation Foundation. All BIL students from 7th grade up to 10th grade are obliged to take multiple-choice tests from all fourteen subjects except such subjects as physical education (PE) and industrial arts. The 11th graders in BILs have special preparation tests for the Unified National Test. These tests are taken by the end of each quarter and the Bilim Innovation Lyceums compete with each other to get the highest ranking in terms of achievements of students on these test results (Gaipov et al., 2013). This test is to measure the general knowledge of students from all subjects, therefore called the Test of General Knowledge. The General Knowledge Test (GKT) is unique to Bilim Innovation Lyceums and there is no such systematic and unified evaluation in other state schools in Kazakhstan. Chapter-3 discusses the GKT content and test-taking procedures more precisely.

The history of education for gifted children in Kazakhstan has a significantly long history beginning from the Soviet Union periods. One of the first schools that started to work with gifted children is the republican specialised physical and mathematics school named after Orymbek Zhautykov, which is well known for its achievement both nationally and internationally (Yakavets, 2013). Apart from that, after the collapse of the USSR, the first Kazakh–Turkish Lyceums (KTL) started to function in 1992. Since then, Kazakh Turkish Lyceums started to open in other regions of the country. These were both mixed and gender-segregated schools. In 2016, in honour of the 25th anniversary of Kazakhstan's independence, KTL was rebranded to BIL, which stands for Bilim Innovation Lyceum. "Bilim" means knowledge in the Kazakh language. The admission examinations to BILs consist of mathematics, logic, history and Kazakh language to measure students' general logic and mathematics level. Mainly, BILs admissions begin by the end of the academic year for 6th-grade students; and that means the first-year students at BILs are 7th graders. The main differences between Bilim Innovation Lyceums and other public schools are the multi-language education, which covers Kazakh, Russian, English and Turkish languages and the education program that focuses on mathematics and science subjects (Yakavets, 2013). Apart from that, the Ministry of Education and Science established the Republican Research and Practical Centre "Daryn" in 1998. The Daryn centre is a body that functions as an umbrella for the other selective schools with gifted children.

1.4 The original contribution

Conducting a research study that contributes to the development of validated measurement tools for computational thinking performance using multiple-choice questions helps shed light on how to measure computational thinking skills of secondary school students. Such measurement tools and the findings of the analysis of the relationships can help in designing better curricula to support children's cognitive development by adjusting subjects that enhance children's problem-solving skills. As the study also investigates the relationship between computational thinking performance and general school achievement of secondary school students it demonstrates what interrelationship and correlations exist between them. This study also informs all educators, teachers, school administrators, Bilim Innovation Lyceums, and stakeholders, the education board and the Ministry of Education and Science of the Republic of Kazakhstan of the measurement methods and relationships between thinking skills of students. The findings of this study provide recommendations and implications from computational thinking performance test to guide school agendas for school administrators as the study investigates the relationship between computational thinking and general school achievement. This research also contributes to the body of knowledge about the measurement of computational thinking by multiple-choice items, as there is not much research about measuring computational thinking by using non-tool-specific instruments. It aims to demonstrate the validity and reliability of using multiple-choice questions to measure computational thinking performance. The findings of this research specifically provide feedback on Bilim Innovation Lyceums (BIL) students' performance of computational thinking, while also raising awareness of teachers about the concepts of computational thinking. The study offers a better understanding of the relationship between computational thinking skills and school achievement, thus emphasizing the role of science and language subjects in schools. This research also contributes to the Quantum project (question bank) and provides a piece of evidence for the future evaluation of the assessment processes. It also provides recommendations for constructing good quality multiple-choice questions in general and specifically to measure computational thinking performance. The findings obtained in this research also adds to the literature for the studies that investigate the relationship between computational thinking skills and other variables. A statistically significant relationship between any subject score and computational thinking scores, the recommendations and implications of this study can be taken into consideration when constructing and developing new curricula and measurement of computational thinking skills.

1.5 Overview of the chapters

Chapter 2 focuses on the literature review. It discusses thinking in general, the development of computational thinking, and the descriptions of computational thinking by different organisations such as Google, ISTE, CSTA, Royal Society, CAS and the University of Canterbury. Then, the role of computational thinking and computer science in the UK and Kazakhstan, as well as studies and projects on the assessment of computational thinking are presented. The evaluation of computational thinking and the research works that study the relationship between different domains are discussed. Chapter 2 also covers assessments in more general terms, particularly regarding response items, and summarises the advantages and disadvantages of multiple-choice items. Chapter 3, the methodology chapter describes the research methodology used in this study. The research paradigm and ethical considerations are discussed. The quantitative research approach, research design and research questions are explained. As the research participants, the 8th-grade students from Bilim Innovation Lyceums are introduced, particularly these Bilim Innovation Lyceums and their features are introduced as well. As primary data answers from the multiple-choice questions designed by the author and a pre-existing computational thinking scale questionnaire (Korkmaz, Çakır, & Özden, 2015) are used, while as for secondary data the results from four General Knowledge Tests are used. The variables and their definitions are provided. The validity and reliability of used instruments are discussed. The pilot study process and the main phase of data collection are explained under the data collection section. Chapter 4, the results chapter, presents the findings from the quantitative analysis obtained from 775 participants. The following are presented in the results chapter: descriptive statistics, inferential statistics, Cronbach's alpha as a reliability analysis of the computational thinking scale questionnaire, correlation and regression analysis. As for the reliability and validity of the multiple-choice questions used to measure the computational thinking performance of students, the Item Response Theory (IRT) is used. Chapter 5, the discussion and conclusion chapter summarises the findings and answers each research question. These conclusions are then discussed in light of the existing literature. Finally, the limitations of the research study and recommendations for future research are presented.

Chapter 2 Literature review

As this research covers several topics like computational thinking, assessment and multiple-choice questions, this chapter starts by discussing 'thinking' in general, after that focusing on computational thinking. Then, programs and tools for assessing computational thinking are presented, particularly the programs for delivering computational thinking and the studies with instruments used to measure computational thinking. General aspects of assessment are also discussed in the last section including multiple-choice questions used to measure computational thinking performance of secondary school students.

2.1 Thinking

People in every moment of their lives are involved in the process of thinking. A big stream of information from outside is converted into nervous impulses then transmitted to the human brain, which the brain processes this information to form thoughts (Spielman, Jenkins, & Lovett, 2020)**.** The way to understand what is in one's mind is by transforming these thoughts into words and speech, or transferring them into actions (Athreya & Mouza, 2017) or forms such as sounds, shapes, pictures, numbers, emotions, tastes, touches and smells (Ham, 2018). Athreya and Mouza (2017) do not consider every thought passing through our mind as an element of thinking, stating that only conscious actions involve thinking, not intuitive ones. Several philosophers and scientists presented various definitions for the term thinking. Dewey (1910) defined thinking as 'that operation in which present facts suggest other facts (or truths) in such a way as to induce belief in the latter upon the ground or warrant of the former' (p 3). Baron (1993) defines thinking as 'a mental activity that is used to resolve a doubt about what to do, what to believe, or what to desire or seek' (p 193). Bartlett in his paper in 1958 (cited in Athreya & Mouza 2017) defined thinking as 'the use of information about something present, to get somewhere else' (p 27). Two powerful factors affecting thoughts are emotions and memories (Spielman et al., 2020). Based on the view and stance of Dewey, Bartlett, and Baron, Athreya & Mouza (2017) define thinking as a higher-order cognitive function used in the process of making choices and judgments. They also state that, as in Bloom's revised taxonomy (L. W. Anderson et al., 2001), series of sub-processes such as lower-order thinking and higher-order thinking construct the process of thinking and describe them as follows:

- Lower-order thinking is involved in retrieving information from the memory and learned knowledge through gathering, storing, measuring and observing.
- Higher-order thinking is more complicated than the lower one and is carried out by integrating data, analysing information and tinkering to gain new knowledge.

Bloom's taxonomy is a classification system that defines different levels of thinking, learning, and understanding. Bloom (1956) introduced six levels of the cognitive domain, from recognition of facts, as the lowest level, through getting more complex and to the higher-order abstract mental levels in his original taxonomy. The revision of Bloom’s taxonomy was developed by L. W. Anderson et al. (2001) with several modifications to the original one; and it distinguishes ‘knowing what’, the content of thinking, from ‘knowing how’, the procedures used in problem-solving. The Knowledge Dimension refers to the ‘knowing what’ part and consists of the following four categories: factual, procedural, conceptual, and metacognitive. Factual knowledge includes very basic information. Conceptual knowledge includes the relationships among smaller parts of a bigger structure, which provide them function together, such as classifications and categories. Procedural knowledge is simple how to do something, which includes algorithms, rules, methods and techniques, as well as knowledge about when to use these techniques. Metacognitive knowledge refers to knowledge of cognitive processes and information about how to utilize these processes effectively, in other words, it is the awareness of a person's cognition and particular thinking processes. The Cognitive Process Dimension of Bloom’s revised taxonomy by Anderson et al. (2001) consists of the following levels: Remember, Understand, Apply, Analyse, Evaluate, and Create as demonstrated in Figure 1, where the top levels being the higher-order of metacognitive skills and the bottom is the lower that provide the foundation for higher-order thinking skills.

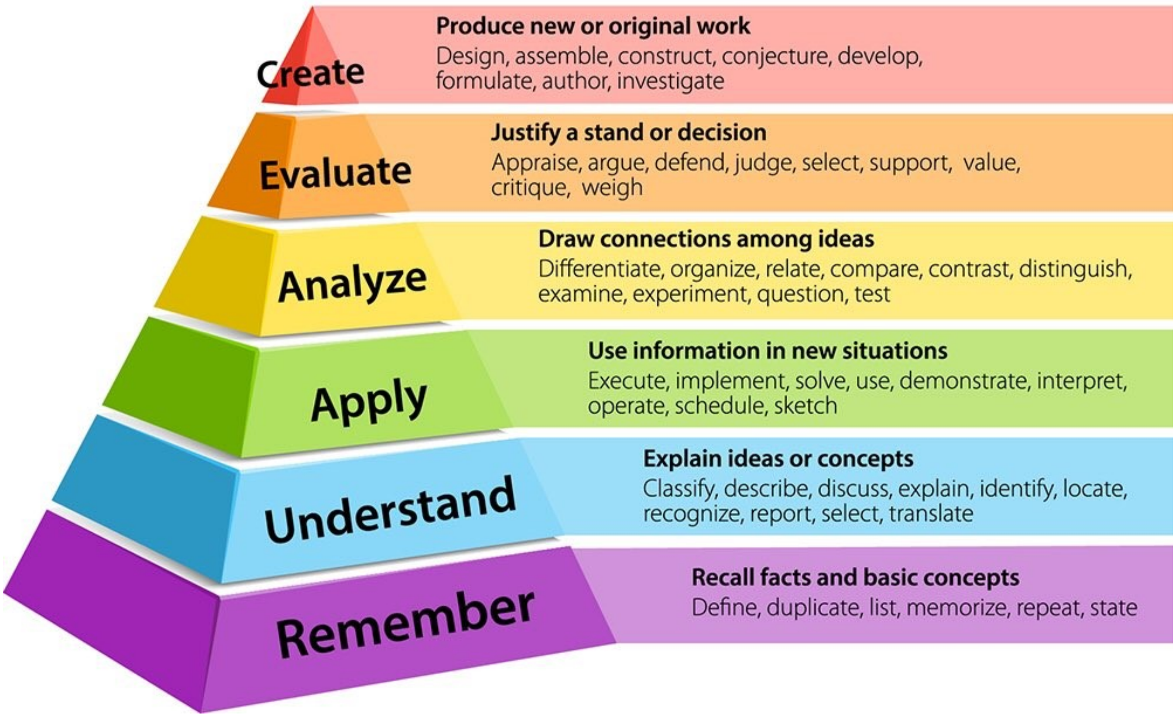


Figure 1 Bloom's revised taxonomy

According to Bloom's revised taxonomy, remembering, as the base of all, is recognizing and recalling relevant information from memory. Understanding shows the ability to construct extract meanings by reading and learning; interpreting, exemplifying, classifying, summarizing, and explaining are the sub-processes in the understanding level. Apply refers to using previously learned knowledge in similar or new situations. Analyse consists of the processes of breaking down the knowledge into subparts and understanding how these subparts are related to its overall structure. Analyse is where the processes of differentiating, organizing, and attributing take place. Evaluate level includes checking and judging. Creating is putting previously learned knowledge and skills together to make something new. To be able to create, learners produce and generate ideas. According to Bloom's revised taxonomy, each level of knowledge can demonstrate the level of cognitive process, where a learner can remember factual or procedural knowledge, understand conceptual or metacognitive knowledge, or analyse metacognitive or factual knowledge. Bloom's revised taxonomy presents the hierarchy as a pyramid where the lower-order thinking provides the foundation for higher-order thinking, however, Agarwal (2019) claims that the activities that require lower-order thinking, such as fact quizzes can only enhance lower-order thinking not the higher-order learning; and higher-order retrieval practices mostly enhance higher-order learning. Thinking involves higher-order mental actions, and it drives people to act deliberately based on their goals. By means of thinking and thought, it is possible to organize the objects and experiences in systematized knowledge, then based on this knowledge avoid future possible dangers or see possible opportunities. We do not question if lower-order or higher-order is better than the other, the question is the criteria: for recalling accurate details lower-level learning is more effective, for transferring the learned knowledge to other areas the higher-level strategies better helps. Hattie and Gregory (2018) claim that a person cannot demonstrate higher-order thinking skills, such as problem-solving without acquiring enough level of content knowledge; higher-order thinking requires lower-order thinking and enough knowledge to be effectively utilised. The questions of whether factual knowledge improves higher-order learning and does it differ for different age groups are left open and needs further investigation (Agarwal, 2019). Real-life problems and situations are diverse, and the solution approaches, techniques used are also different depending on the nature of the issues, so the thinking. The questions of whether factual knowledge improves higher-order learning and does it differ for different age groups are left open and needs further investigation. Computational thinking requires not only lower-order thinking but also higher-order cognitive skills. Athreya & Mouza (2017) state that computational thinking is a special branch of science while listing other types of thinking as follows: Reflexive and Low-Level Analytical Thinking can be described as everyday thinking, which is almost reflexive, reactive and automatic, and it is based on perceptions, memory and emotions. To get wiser and well-adapted, it is important to look back on the decision we made and reflect on our reflexive action. In contrast to reflexive thinking, reflective thinking is controlled thinking. It is not automatic but planned. An example of reflective thinking is scientific reasoning, and critical thinking, creative thinking, convergent thinking, divergent thinking, inductive thinking, and deductive thinking can all be under this category. Critical thinking is not simply the ability to distinguish useful and truthful information from useless or fake ones. Critical thinking is analysing and evaluating obtained information thoroughly and systematically with concentration. In addition to objectivity, precision and appropriate use of logic and language, the critical thinking process needs higher-order cognitive skills such as analysis, synthesis, interpretation and comparison. Creative thinking is referred most when to answer a question regarding the future, and it requires imagination and aesthetics to create such artefacts as art, novel mechanical devices, new technology or a new method of solving a problem. When creating an art object, a human mind perhaps is not solving a problem but creating something new using imagination. This imaginative thinking is also creative or artistic thinking. In contrast to artistic thinking, creative problem solving involves inventing something novel and problem-solving, but not producing art objects. Convergent thinking is a cognitive process when the mind gathers information from various sources and angles to a certain conclusion. Convergent thinking is used in a critical thinking process when synthesising evidence from multiple sources into one. Divergent thinking is the cognitive process when the mind starts at a point and looks for different ways to reach a certain point. Divergent thinking is close to creative thinking and it utilises the imagination power of the mind. Inductive thinking is a reasoning process, a generalisation process, which allows the mind to move from particular to general, from a part to a whole. Deductive thinking is the opposite of inductive thinking, which is the reasoning that allows the mind to move from a general to a more specific point. Concrete Thinking refers to a cognitive process when a specific result or goal is to be achieved and reached. When thinking is used as a stepping-stone to further thinking, it shows abstract thinking. Collecting facts, balancing the evidence, testing different hypotheses, and making decisions, developing ideas are examples of abstract thinking. Analytical thinking in cognitive psychology means understanding the whole, its parts, and the interrelationships; and synthetic thinking considering different ideas, connecting the parts to the whole, and making a relationship between parts and ideas. Systems thinking, refers to the study of complex systems through the study of the structure and functions of their components and their interconnections. System thinking is a study of complex systems by examining the structure and functions of the components and their relationships with each other. In system thinking the main focus is on the stable development, resilience, and sustainability of the whole system (Athreya & Mouza, 2017).

Recent studies by neuroscientists show two types of networks in our brain, the highly attentive and the resting states (Andrews-Hanna, 2012). The highly attentive state activates when our brain is focused on something and when it is more relaxed it is in the resting state (Oakley, 2014). Oakley (2014) names the thinking process related to these two different types of networks the focused mode and the diffuse mode, which are very close in meaning to de Bono's (1970) descriptions of vertical and lateral thinking. Immordino-Yang et al. (2012) recommend teachers have a balanced practice of switching between a focused and diffuse mode in education for the best achievement of students. Both focused and diffuse modes are necessary for problem-solving, but more importantly, the skill of switching from one to another when necessary is crucial. Lewis and Smith (1993) suggest there is a need for a term that includes problem-solving, critical thinking, creativity, and decision-making. They offer higher-order thinking for such a comprehensive coverage term; and they define it as 'when a person takes new information and information stored in memory and interrelates and/or rearranges and extends this information to achieve a purpose or find possible answers in perplexing situations' (Lewis & Smith 1993, p 136).

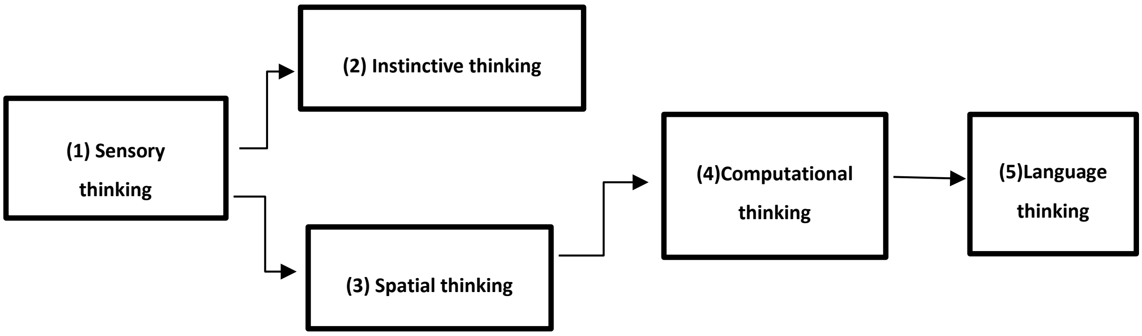


Figure 2 Five ways of thinking hierarchy by Ham (2018)

There are various types of thinking; these types of thinking depend on the nature of the problem or question. Lewis and Smith (1993) and some other scholars suggest using the term higher cognitive skills to refer to all types of thinking (Athreya & Mouza, 2017). Ham (2018) states five ways of thinking: sensory thinking, instinctive thinking, spatial thinking, computational thinking, and language thinking, as shown in Figure 2. Ham (2018) points to the need to distinguish between the terms like ways of thinking, modes of thinking, and styles of thinking, each of which describes different aspects of our mind (p 105). He explains the difference between them in a hierarchical way: all thinking modes are carried out by a specific way of thinking usually expressed by a particular thinking style.

Computational thinking is a complex construct with both lower-order (e.g. following algorithms) and higher-order (e.g. generalisation) thinking. Computational thinking processes in the human mind encapsulate both focused and diffuse modes of thinking. According to Ham (2018), thoughts are blocks of data that occurs through the assembly process in one's head, which is run by different rules and schemas. Computational thinking is the ability of a person to process complex thoughts in his/her head formed by these rules and established schemas. An assignment of discrete thought variables such as the concept of time is where computational thinking starts to form. And the discrete thought variables are assembled with spatial arrangements, where these arrangements become computational rules and build generic schemas. Brichacek (2014) describes computational thinking as a cognitive approach that involves a set of digital age abilities and their use in a proper context; Brichacek defines computational thinking as utilizing the power of computing together with higher-order thinking skills. Ham (2018) states language thinking as a separate category in human thinking and there is a certain relation between using words and speaking language. Language thinking is based on symbols of a particular language, which can be both received and dispersed (Minsky, 1988). Language can be described as a mode of expressing thoughts using the motions of the organs of that person's body. Thought cannot be carried on without the help of language, thus thoughts and language cannot be separated (Vygotsky, 1986). The knowledge about language's neurobiological basis made it possible to propose various models, which can learn grammar rules and analyse sentences that complies with the hierarchical organization and information flow in the human brain (Borensztajn, 2011). Any language as a system contains such subcategories as phonology - the study of the speech sound, morphology - the study of words, syntax - the study of sentences and phrases, and the rules of grammar and semantics - the study of sentence meanings, each of which has several sets of rules; and bonded to speech, writing, and gestures. Phonology and syntax are close to the core of linguistic intelligence, while semantics and pragmatics are related to logical intelligence (Gardner, 2011). In this perspective when looked at from a wider angle, a language can be considered as a computational system. Many learning theories exist in education, some has disappeared, and some new theory proposals are being introduced. A cognitive architecture by Anderson (1993) the Adaptive Control of Thought (ACT-R) is based on the assumption of a unified theory of mind. The ACT-R attempts to explain how human cognition functions when learning occurs, and show how the elements and processes of human memory, thinking, problem-solving, and language work. The ACT-R (R stands for rational) is based on the principle of the rationality of the human mind, which means the human mind is optimized to get maximum with minimum effort. The ACT-R theory allows us to predict typical measures like latency, such as time to perform a task, accuracy level, such as correct or wrong answer, and neurological data such as data from fMRI (Qin, Bothell, & Anderson, 2007). ACT-R models how people recall chunks of information from memory and how they solve problems by decomposition, that is breaking down the problems into smaller parts (sub-goals) and applying knowledge from the working memory.

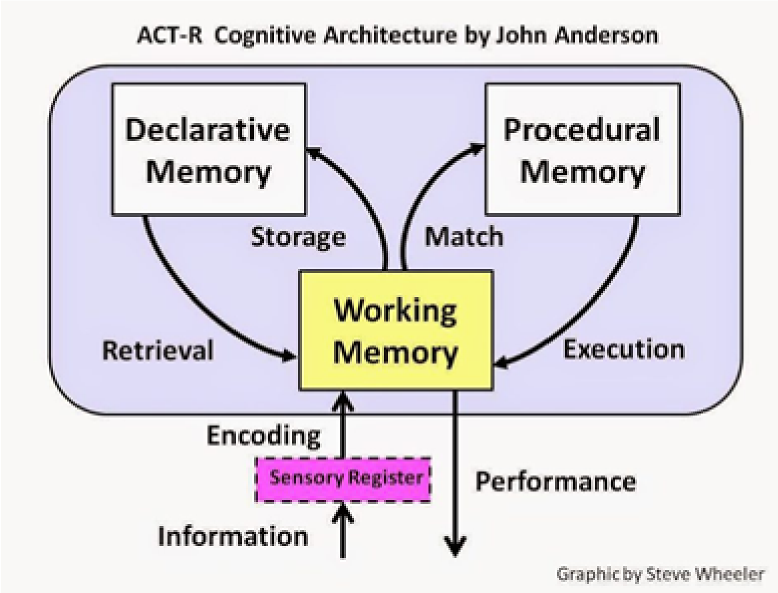


Figure 3 ACT-R Cognitive Architecture by Anderson

ACT-R divides human knowledge into two, declarative and procedural, which are different from each other but related. Declarative knowledge holds factual knowledge and any relevant association and context. Declarative knowledge also consists of many other knowledge chunks, a set of facts that are already known and active goals. An example can be “Schools are closed during holidays”. This chunk itself has many chunks like “school”, “holiday” and “closed”, which means chunks are connected. Procedural knowledge comprises sequences of actions based on matching the patterns like in computer systems “if/then” statements, which lead to achieving a certain goal when a specified pre-condition is satisfied – if this happens, then act this way. An example of procedural knowledge can be “If I am thirsty, I drink water”. Human memory is complex and can hold many declarative knowledge chunks and procedural knowledge (production rules). The procedural knowledge mechanism allows us to recognize appropriate patterns and take actions directly without further calculation. Unlike declarative knowledge, procedural memory is a slow-learning system in which new capacities materialize gradually over time. Shortly, declarative knowledge is all about facts and things, where procedural knowledge is about how to perform actions. With accumulating declarative knowledge, people start to build procedural knowledge. Human cognition has the ability to store, process, and project the past learned experiences into future cases, as well as the ability to override these experiences and adapt to new situations (Ohlsson, 2012). When people read a given text, listen to one another, thinking about a solution to different types of problems, in each of these types of situations people interconnect with the surrounding world by keeping themselves informed of the happenings. These are perception and learning at the same time; perceptual learning is learning how to learn from and about the surrounding world, the situations, problems and what we read and hear (Marton, 2006). Anderson (1993) also explains how working memory, which is the part that is currently working can serve as a buffer between long term memory as shown in Figure 3. ACT-R can explain a wide variety of procedures for higher-order thinking skills such as computer programming and language learning. Anderson's research studies looked at the neural process of the human brain during a person was performing mental activities, such as solving equations. The brain scan images showed increased activation in several brain parts and these activated areas became less active once a person acquired proficiency in solving these problems (Qin et al., 2007). The study concludes that it takes some time and energy to get proficiency in certain tasks, but once it is acquired the less brain energy and time is required to perform that task. In other words, over time, the more a person practices the better he gets and becomes an expert, who mastered particular ways to think and reason effectively. Bransford, Brown and Cocking (2000) proposed the idea of conditionalized knowledge when referred to experts and how their knowledge emerges. As reported by (Bransford et al., 2000):

- experts can see features and patterns of information that beginners cannot see;
- experts' knowledge is rich and well-organized, which reflects a deep understanding of the subject;
- experts' knowledge cannot be simplified to just facts and things but reflects contexts of applicability, their knowledge is conditionalized. Thus, experts can easily retrieve important information from their knowledge with a little attentional effort.

While experts show an ability to get the right information from memory fast and accurately, they also show a better transfer of knowledge to similar domains than beginners do (Willingham, 2012). For example, it is expected that the teaching of reading comprehension enhances students' reading and understanding of texts outside the classroom and in other areas and domains (De Corte, 2003), as the main goal of education is to enable students with cognitive tools so that they can apply them outside the initial learning context. As mentioned earlier, the experts make intuitive decisions depending on their experience in expert areas by simply recalling from experience (Athreya & Mouza, 2017). And the reason is experts identify the patterns, similarities and differences in new situations better than beginners, which is a key skill for positive transfer (J. A. C. Hattie & Donoghue, 2016). This type of expert thinking is based on big data, information, knowledge, skills and experience of the experts. Mechanics, chess masters and medical diagnosticians can be examples demonstrating expert thinking at work. Experts use some heuristic methods that they developed during their experience in solving tasks, challenges, doing mental activities and learning from them (Athreya & Mouza, 2017; Bransford et al., 2000). Human cognition, whether an expert or beginner, not only provides cognitive mechanisms for learning new knowledge and skills but the human mind also has the ability to override the learned

experiences and adapt to changing circumstances (Ohlsson, 2012). Perceptual learning (Marton, 2006) is a key factor that makes experts apply their skills and knowledge from one area to another, and this transfer of learned knowledge and skills is a fundamental goal of education (De Corte, 2003). Whether it is a far or near transfer, there are several dimensions of transfer of learning depending on its content and context (Barnett & Ceci, 2002) and it is better to focus on the dimensions of transfer than to debate over transferability of skills (McKeachie, 1987).
The thinking process is an important aspect of learning, as it is expected to transfer certain learnt knowledge and skills from one area to another. The role of higher-order thinking is more significant than the lower-order one's. To what extent computational thinking as a higher-order thinking is transferrable to other domains is a question to be investigated further.

2.2 Computational thinking

This section starts with the development of computational thinking term, then the definitions and descriptions of computational thinking by different organisations are explored. The role of computational thinking and computer science in the UK and Kazakhstan curriculum is discussed. Then, several programs delivering computational thinking are presented.

2.2.1 Development of the term

The idea of computational thinking has been present for a long time in academic publications. It has been expressed by various terms through several stages of technological progress. Since the 1950s the term "algorithmic thinking" was used which meant a mental orientation to formulate problems as conversions of inputs to outputs and looking for algorithms to make the conversion (Denning, 2009). Denning (2009) expresses the expanded version of this term as using mathematics to develop algorithms, thinking with multiple levels of abstraction and testing the solution. Alan Perlis advocated the idea that everyone should know programming as a part of liberal education and all college students of all disciplines need to learn to program (Greenberger, 1962). Perlis stated that programming courses could help students to construct complex processes out of simpler ones, better than any other course. Wilson in 1975, as cited by Denning (2009), argued that by utilizing computation and simulation, science could be more successful; and solve more problems, which were impossible to imagine or dare to solve, before the invention of computer machines. In the 1980s, he joined other scientists in various fields to support the idea that the major challenges of science could be tackled by computation; and he added that the government could help by supporting a network of supercomputing centres (Denning, 2009). Denning (2009) cites Wilson as a proponent of the idea that computation is the third leg of science, in addition to two legs, theory and experiment. Papert (1980) used terms like computational ideas, computational metaphor and computational models in his book "Mindstorms". He believed that by using LOGO programming and the physicality of turtles, teachers could ease education and turn it into a joyful process for children to learn procedural thinking. Papert (1980) argued that procedural thinking is a powerful intellectual tool. After Papert's joint work with MIT, computing became popular in K-12 education in the 1980s (M Guzdial, 2008). Sheil (1981) introduced the term procedural literacy in his paper The Psychological Study of Programming. In 2010, Moursund proposed that computational thinking was firmly related to the original ideas of procedural thinking created by Papert (1980), which contains representing, developing, testing and debugging procedures. An effective procedure is a detailed step-by-step set of directions that can be mechanically translated and carried out by a specified agent (Linn et al., 2010). Meanwhile, there were other scientists, who contributed to this long process of development of computing, programming and digital literacy. "Programming is a second literacy," claimed Soviet computer scientist Ershov (1981, p8). Ershov implied that the development and wide spreading of computers would lead to a universal ability to program, just like the development and distribution of typography led to universal literacy. ICILS (2020) results show that teachers who have higher computer and information literacy abilities are more likely to emphasise computational thinking and computer literacy in their teaching. Literacy and programming can be explained as an organic expression of a person (Ershov, 1981). Similarly, Papert (1980) argued that people need to be fluent with computation for learning. Even though many people were literate enough in that period, not all of them could program or code. As computers became more available and accessible to people with various backgrounds, the use of computers in everyday life helped millions of scientists, not only in engineering but also in social science settings. Khenner (2016) states that what shows up as a second literacy today is the ability to use modern information and communication technologies to solve emerging problems. diSessa (2000) in his book, Changing Minds: Computers, Learning, And Literacy presents the idea of computational literacy. He points out that the effort to teach computational literacy to everyone carries more social aspects. Besides, today, society demands literacy from all citizens and the development of computer literacy should be supported by all science classes not just by computer science (Beheshti et al., 2013). diSessa (2000) defines the term computational literacy with more demands and it carries a different meaning than terms like computer literacy, digital literacy and information literacy (Grover & Pea, 2013a). The notion of computer literacy is a two-way literacy where people do not just use computer artefacts but also produce them (Weintrop & Wilensky, 2013). Programming skills or coding skills is an important aspect of the development of computational thinking as they have underlying structure, they both use algorithmic thinking. Astrachan argues, as well, that computational literacy will allow civilisation to think and do things that will be new to us in the way that the modern literate society would be almost incomprehensible to preliterate cultures (Linn et al., 2010). People with computing skills carry more values and the importance of these skills is crucial both in education and in industry. Denning (2009) introduced the Great Principles framework and listed the four core practice areas of skills and abilities which computing people demonstrate as follows: 1-programming, 2-engineering of systems, 3-modelling and 4-applying. He claims that computational thinking can be considered as a fifth practice. Denning implies that computational thinking is the ability to translate the world as algorithmically controlled conversions of inputs to outputs. It has been reported that computational thinking has long historical roots and that it is innate in human beings, and our ancestors used it to model and represent reality through telling stories (Linn et al., 2010). It is debatable to state that computational thinking is innate, as others are claiming that computational thinking does not come naturally but instead it needs training and guidance (Sanford & Naidu, 2016). However, all people are engaged in computational thinking to some extent in their everyday life (Linn et al., 2010). Ham (2018) claims that computational thinking usually is perceived as something that needs to be developed and trained; but it is a semi-resident natural skill that everyone has. People use the same representation and modelling when accessing digital information using technology every day. One big factor that fostered spreading the idea of computational thinking was Wing's article "Computational Thinking" in 2006. Wing (2006) claims that computational thinking is a fundamental skill for everyone in the 21st century alongside reading, writing, and arithmetic as part of the core knowledge. She defines computational thinking as reformulating a hard to solve problem into a solvable one, using reduction, embedding, transformation, or simulation. Wing (2006) points out the need to reach schools, including teachers, parents, and pupils; and calls everybody to think like a computer scientist. How to get students to think like scientists so that they explore real-life problems and seek solutions? Cognition in the early stages is crucial as it differs from the one in late the stages of learning (Willingham, 2012). Wing (2011, p.1) reformulated the definition and clarified it as following 'Computational thinking is the thought processes involved in formulating problems and their solutions so that the solutions are represented in a form that can be effectively carried out by an information-processing agent.' After her clarification, Aho (2012, p.1) simplified the definition of computational thinking as 'the thought processes involved in formulating problems so their solutions can be represented as computational steps and algorithms.'

Wing (2011) argues that computational thinking is beneficial for everyone, including non-computer scientists and scientists, engineers and other professionals. As explained by Wing (2011), anyone with computational thinking skills can:

- Know which parts of a problem can be tackled by computation
- Assess the match between computational tools and techniques and a problem

- Know the limitations and power of computational tools
- Understand the limitations and power of computational tools and techniques,
- Use or adapt a computational tool or strategy to another utilisation
- Understand an opportunity to apply computation in a new way
- Use computational strategies such as the “divide and conquer” strategy in any field

Scientists, professionals and engineers who use computational thinking are also able to demonstrate the following abilities:

- Apply new computational techniques to their problems
- Reshape their problems to be accessible to computational strategies
- Discover new science through analysis of big data
- Ask new questions that were not thought of or dared to ask because of its big scale, but which are easily addressed computationally
- Express problems and their solutions in computational terms.

According to Malyn-Smith and Lee’s (2012, p.3) definition, the professional with computational thinking ability is the one who can collaborate “in a creative process to solve problems, design products, automate systems, or improve understanding by defining, modelling, qualifying and refining systems, processes or mechanisms generally through the use of computers”. Wing’s invitation to spread computational thinking across the whole education system served as a beginning point for many researchers, policymakers, leaders from the computing industry, education and science departments and academies (Tedre & Denning, 2016). The National Academy of Sciences explored the nature of computational thinking and its cognitive and educational implications; and the pedagogical aspects of computational thinking. Although the term computational thinking is relatively new for many, some types of thinking which are previously well-known are can help to form computational thinking. Logical thinking, algorithmic thinking, engineering thinking, and mathematical thinking and heuristic thinking are the types of thinking that can be found in the idea of computational thinking Selby (2014). The National Research Council (NRC) in the United States ran two workshops in 2010 with leaders from industry, higher education and K-12 education. These two workshops showed the lack of consensus on computational thinking (Bienkowski, Snow, Rutstein, & Grover, 2015), leaving some main questions with no answer. These are:

- How can computational thinking be recognised?
- What is the best pedagogy for promoting computational thinking among children?

- Can programming, computers, and computational thinking be legitimately separated?

(Linn et al., 2010)

Selby (2013) worked on developing the definition of computational thinking based on the consistency of usage and interpretation across the computational thinking related literature. The objectives of her studies on definition are

- to narrow the definition than of a broader one
- to improve the definition to facilitate the assessment
- to put the criteria into order
- to preserve the validity of previous studies (e.g. development of curriculums)
- to separate a definition from activities and artefacts that evidence the use of computational thinking skills

There are various terms associated and closely related to the term computational thinking. Selby and Woollard (2014) presented some aspects of computational thinking and described the terms more deeply and narrowly for teachers and educators in the UK, discussing the well-established definition and providing a clearer description for the concepts of the term computational thinking. Selby and Woollard (2014) listed the terms in four different areas as thinking, problem-solving, computer science, and imitation terms. These terms can be listed as follows: abstraction, decomposition, logical thinking, algorithmic thinking, engineering thinking, heuristic thinking, parallel thinking, mathematical thinking, problem-solving, analysis, generalisation, modelling, simulation, and visualisation. However, not all of these terms were concluded to take place in the final definition of computational thinking. Selby (2014, p.38) defines computational thinking as an often product-oriented activity, associated with, but not limited to, problem-solving; and a cognitive or thought process that reflects the abilities to think in abstractions, algorithmically and in terms of decomposition, generalisation and evaluation. The study by Hoskey and Zhang (2017) focuses on refining and categorising computational thinking, taking into consideration common aspects of computational thinking with other types of thinking, such as logical thinking, abstract thinking, algorithmic thinking, mathematical thinking, lateral thinking, vertical thinking, scientific thinking and critical thinking. Hoskey and Zhang (2017) categorise aspects of computational thinking into lower and higher. Absolute vs relative thinking, automation thinking, categorisation, compound thinking, information representation, interface thinking, iterative thinking, modularisation, reuse thinking, sequential thinking and unique identification are all categorised under lower level as an entry-level programming course. Higher-level is related to problem-solving, design and modelling as for upper-level data structure courses; and the higher-level computational thinking has the following components: abstraction, object-oriented thinking, parallel thinking, recursive thinking and scalability thinking. Based on these categorisations, it can be said that novice and experts have different level of computational thinking, therefore when measuring computational thinking skills, the age-appropriate assessment is crucial. As a conclusion, it can be remarked that computational thinking is a thought process where a person converts raw unsolved problems into easily solvable chunks, it is a systematic way people think about solving complex problems.

2.2.2 Practical definition

Although there is no single accepted definition for the term, there are common main strands of computational thinking (Fraillon, Ainley, Schulz, Friedman, & Duckworth, 2019). Logical thinking often acts as a gate for other concepts of computational thinking and helps see problems through analysing given facts. One of the fundamental thinking skills of programming is algorithmic thinking. The term algorithm is the most frequently used in the current models of computational thinking (Curzon, Dorling, Selby, & Woollard, 2014; Selby, 2013; Selby & Woollard, 2014). Algorithmic thinking is developing concrete steps that lead to the solution of any problem. In other words, it means building a well-written set of instructions to follow to solve existing problems or given tasks. A good program is one with an efficient and good algorithm. However, this does not mean that algorithmic thinking is used by only programmers and coders; algorithmic thinking skill is widely used and applicable in all areas of life. Generalisation is identifying a pattern in a previously built algorithm. The similarities help to solve problems easier by adapting these patterns to other circumstances, domains and problems. Abstraction is a key thinking skill that enables people to focus on what is important by hiding unnecessary details to make the problem-solving process less complex. Using abstraction often comes with generalisation and decomposition skills. Decomposition is a way to manage big and complex ideas, problems and systems by breaking them into manageable pieces to allow a group of people to be involved in the process of development at the same time. Finally, evaluation is a process where the effectiveness of a provided solution is judged and can be followed by questions: Is the solution process well enough for its purpose and how quickly does it proceed? Many researchers and scientists have been working on the enhancement of computational thinking skills, building curricula with integrated computational thinking skills to fit in K-12 education, and creating assessment systems for computational thinking skills. They claim that integrating computational thinking will improve children's cognitive functions and help them in problem-solving (I. Lee et al., 2011).
Computational thinking appears as a problem-solving technique for students and at the same time as a contemporary style of political thinking (Williamson, 2016). According to Williamson (2016, p2), "political computational thinking is a style of thought, then, that aims to translate

social phenomena into computational models that can then be solved by being formalised as step-by-step algorithmic procedures that can be computed as proxies for human judgment or action”. All the learning to code and hacking discourses are the result of this political computational thinking mode (Williamson, 2015). Williamson claims that different campaigns, groups and networks of like-minded organisations deliver computational skills by learning-to-code activities. He also concludes that the effort of the government to learn to code is evidence of how computational thinking and algorithmic solutions can be applied to solve social problems. Allsop (2019) suggested that computational thinking is a cognitive process that can be regulated by metacognitive practices, which also involves the application of a series of computational concepts and the utilisation of learning behaviours, and also aims to design solutions to problems that are susceptible to automation. Although it is expected that computational thinking skills can be applied in many different fields, some researchers state that there is no solid evidence showing that computational thinking enhances other general cognitive skills (Tedre & Denning, 2016) and the question “Is computational thinking good for everybody?” remains unanswered (Denning, 2017). In general, the topic of transfer of knowledge and skills has been studied for over 100 years with both theoretical and practical perspectives. While some findings show the transfer is possible, some show it does not, claiming that the transfer can only be seen when there are specific identical elements between the initial domain and the applied area, however from a cognitive psychology perspective, the transfer of problems solving skills is possible (De Corte, 2003). Tedre and Denning (2016) claim that computational thinking skills do not transfer to other environments and highlight that computational thinking itself has enough power and no need for such exaggerated claims. Other researchers warn that such big claims about computational thinking might have a negative effect (Blackwell et al. 2008). Blackwell et al. (2008) list possible side effects of computational thinking that might occur as a result of the constrained application of abstraction, saying that abstraction is an enemy when we proceed without due caution. Denning (2009) also warns us about possible misleading because of the wrong perception of computational thinking, pointing to the fact that computational thinking is one of the key practices but not unique to computer science. People should look back at history and learn from the past and not ignore the work done before to avoid failures in promoting computational thinking ideas (Tedre & Denning, 2016). Grover and Pea (2013) suggest answering the following two questions before beginning any integration of computational thinking into curricula: What to anticipate from children to become better in after introducing and delivering the curriculum fulfilled with computational thinking? And how to measure this computational thinking? Although without a clear definition of computational thinking, it still might seem possible to see examples of the practice of computational thinking skills, however, this issue itself can be an obstacle on the way to measure computational thinking (Selby 2014, p 24).

Computational thinking is introduced and described in detail by several organisations such as Google, ISTE, CSTA, Royal Society, CAS, the University of Canterbury, BBC and the College Board. Although they all discuss the same term “computational thinking”, the definition and the aspects related to it are introduced differently. It is useful to list these complete and detailed versions to see the similarities and differences of computational thinking by these organisations. Google for Education (2015) introduced a computational thinking course for educators. They define computational thinking as a problem-solving process that includes several characteristics and dispositions. Being essential to the development of computer applications, computational thinking can also be used to support problem-solving across all disciplines, including the humanities, mathematics, and science. Students who learn computational thinking across the curriculum can begin to see the relationship between academic subjects, as well as everyday life both inside and outside the classroom. The mentioned course provides an opportunity to experience some of the concepts of computational thinking, such as 1-Decomposition: Breaking down data, processes or problems into smaller and manageable parts; 2-Pattern Recognition: Observing patterns, trends and regularities in data across the processes and systems; 3-Abstraction: Identifying the general principles that generate these patterns; reducing the complexity by removing unnecessary (irrelevant) details; representing the artefact, process or system in a diagrammatic, visual and/or computable format; 4-Algorithm Design: Developing step-by-step instructions for solving problems or the rules by which a system or process works. (Google for Education, 2015)

The International Society for Technology in Education (ISTE) and the Computer Science Teachers Association (CSTA) collaborated with leaders from higher education, industry and K-12 education, and collected feedback from nearly 700 science teachers, researchers and practitioners to develop an operational definition of computational thinking that provides framework and vocabulary for better understanding and integrating computational thinking. ISTE and CSTA (2011) introduce computational thinking as a problem-solving process that includes 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 analysing data; representing data through abstractions such as models and simulations; automating solutions through algorithmic thinking (a series of ordered steps); identifying, analysing and implementing possible solutions to achieve the most efficient and effective combination of steps and resources; generalising and transferring this problem-solving process to a wide variety of problems.

These skills are supported and enhanced by several dispositions that are essential dimensions of computational thinking. Being confident in dealing with complex and open-ended problems, maintaining persistence in working with any problems, showing tolerance for ambiguity and collaborating with others to reach the common goal are the dispositions that are essential to computational thinking skills. The Royal Society (2012, p29) defines computational thinking as a "process of recognising aspects of computation in the world that surrounds us, and applying tools and techniques from Computer Science to understand and reason about both natural and artificial systems and processes". It can be used to make useful and usable computational systems that enable us to understand artificial and natural systems (The Royal Society, 2017). The main idea is that any solution to a problem lies in the algorithms leading to the answer, which can be followed by people or computers. The power of computational thinking can be applied to many areas using algorithmic techniques that can offer solutions to even bigger and complex problems. Computing at School (CAS) is a UK-based community that promotes computer science at schools. CAS has almost 30,000 registered users, over 90,000 discussion posts, over 250 local hubs and nearly 4500 teaching resources. Teachers and educators share their ideas, experiences and resources through this organisation. The resources such as lesson plans, classroom activities, and guidelines for teachers are available for different levels starting with primary teachers and secondary teachers. CAS lists the concepts of computational thinking as:

- Logical reasoning: predicting, analysing and explaining
- Algorithms: making steps and rules
- Decomposition: breaking problems or systems down into parts
- Abstraction: managing complexity, sometimes through removing unnecessary detail
- Generalisation: spotting and using patterns and similarities
- Evaluation: making judgements

The approaches of computational thinking are introduced as:

- Tinkering: experimenting and playing
- Creating: designing and making
- Debugging: finding and fixing errors
- Persevering: keeping going
- Collaborating: working together

(CAS 2014)

Barefoot Computing is for primary school teachers, while QuickStart Computing is for both primary and secondary school teachers with an emphasis on the transition from primary to secondary. Barefoot introduces computational thinking as a two-step process, where people first think about the steps needed to solve a problem, then use their technical skills to get the computer to work on that problem. Barefoot's (2014, p1) definition of computational thinking is as follows: "Computational thinking is about looking at a problem in a way that a computer can help us to solve it". The "CS Unplugged" project by the Computer Science Education Research Group at the University of Canterbury, New Zealand introduces computational thinking with practical activities. Through the activities and lessons promoted by CS Unplugged, students learn how to describe a problem, find the important details of the problem, break this problem into smaller steps, go through these steps to solve the problem and finally evaluate the process. CS Unplugged adopted six concepts of computational thinking such as algorithmic thinking, abstraction, decomposition, generalising and patterns, evaluation and logic (CS Unplugged, 2016). CS Unplugged (2016) states that these computational thinking skills are transferable to other areas, but more relevant to problem-solving by using the power of computers and digital systems. The BBC (2018) explains computational thinking as a technique that allows people to deal with complex problems; first, by understanding the problems, then developing possible solutions to those problems. The solutions can be presented in such a way that a computer, a human, or both of them can understand them. According to the BBC (2018) computational thinking has four key elements, just like the four legs of a table, which cannot stand alone with missing legs: 1-Decomposition: breaking down a complex system or problem into more manageable parts smaller parts; 2-Pattern recognition: identifying similarities among and within problems; 3-Abstraction: ignoring irrelevant details by focusing on only the important ones; 4-Algorithms: making a step-by-step solution or set of rules that leads to the solution of a problem.

A non-profit organisation The College Board (2017) introduces six computational thinking practices. These practices are 1- Connecting computing: explaining the connection between people and computing as well as computing concepts; 2-Creating computational artefacts: creating artefacts by writing codes and programs, remixing digital music, making animations, creating websites, and applying computing methods to solve problems creatively; 3-Abstracting: developing models and simulations of natural and artificial phenomena; using abstraction to predict and analyse the world; 4-Analysing the problems and artefacts: Evaluating the solution of the problem, finding the errors, and correcting them; 5-Communicating: providing written or oral descriptions supported by charts, graphs, visualisations, and other computational analysis; 6-Collaborating: Collaborating with others in solving computational problems in creating artefacts, sharing the work by contributing individually to the final product of teamwork, exchanging experience and skills, giving/receiving feedback to/from other members of a team.

As can be seen from these organisational views that computational thinking is not a simple idea with plain characteristics but an idea with diverse features and approaches. Some name pattern recognition while some use the term generalisation, some call automation while others say algorithmic design, in fact, perhaps classifications may vary although the underlying structures are similar. Computational thinking can be defined in its simplified form as the “stuff” that people have to learn to tackle their tasks using computers (Mark Guzdial, 2019). Although there is no single widely accepted definition, the listed concepts and approaches are common strands of computational thinking.

2.2.3 Computational thinking concepts

As discussed in the previous section, the computational thinking concepts in the Kazakhstani national curriculum are similar to the UK system introduced by Barefoot. These concepts of computational thinking are logical reasoning, algorithmic thinking, decomposition, generalisation and evaluation. This section briefly describes each of these concepts.

Logical reasoning

Logical reasoning is the process of consistently using a rational and systematic series of steps based on given statements to conclude. By logical reasoning, people can make sense of problems by analysing given statements and existing facts and thinking of the relationships between these facts clearly and logically. People use logical reasoning in everyday life to make sense of various types of daily affairs, from everyday problems at home to specific professional tasks. Computer programmers use logical reasoning when they test and debug codes of their own or others. Using logical reasoning can lead to other computational thinking skills such as algorithmic thinking and abstraction (Livingstone & Saeed, 2017).

Algorithmic thinking

Algorithmic thinking is getting to a solution to a certain problem through concrete steps to implement it. An algorithm can simply be described as a series of instructions (steps) to solve a particular problem. Algorithms are used not only in informatics or the computer world but in other subjects and domains, even in our everyday lives. Preparing a meal in the kitchen, learning a new dance in a dance club, doing multiplication operations at mathematics class or delivering a parcel from point A to point B all can be considered as examples of algorithmic thinking as all cases require to follow a certain set of steps (instructions) or rules. There can be more than one solution (set of steps) in each case, a person who performs the task can come up with various solutions, some easier and shorter, some longer and harder. Good solutions mean they are clear to understand and follow a logical order, which makes them applicable to the problems by the person who performs the task. Algorithmic thinking is one of the crucial skills in coding and programming and good programmers are good algorithmic thinkers (Curzon & McOwan, 2017). The algorithmic thinking skill is applied to many areas not only in programming. All programmers are algorithmic thinkers but not all algorithmic thinkers are programmers.

Abstraction

Abstraction is defined as the ability to decide what details of a problem are important and what details can be ignored (Livingstone & Saeed, 2017). By removing unnecessary details without losing anything important, it is easy to deal with problems, especially with complex algorithms. Abstraction makes problems easier to think about it by using appropriate representations and models. One example is a handbook for students. Consider a handbook for freshmen and a handbook for PhD students. These two handbooks provide different information as the focus and the target audience were different in each handbook, thus different unnecessary details of students' life at the university have been hidden and not provided at all. In handbooks for freshmen, some advice for novices can be found and no information about the processes at the university such as choosing a dissertation topic and supervisors are present. Likewise, handbooks for a PhD student would contain some sort of information that is related to their research, thesis and supervisors. An ability to use abstraction is a key computational thinking skill that often follows on from generalisation.

Decomposition

Decomposition is the ability to see the problems, systems, algorithms, ideas and solutions in smaller parts. Each smaller part can be separately solved, understood, developed, designed and evaluated; and this fact makes large complex problems and systems easier to solve design also by allowing a group of people to work on the same task as a team. The popular software people use nowadays can be an example of how the program is developed over many years by a group of people, each of whom is responsible for a certain level of the software.

Generalisation

Generalisation means identifying the patterns by looking at the previous solutions and algorithms used to solve new problems. An algorithm that can solve a particular task can be adapted for a set of similar tasks and problems. It is important to ask questions like "Is the new problems/task/system has any similarity to what have already been solved/understood/designed?" and "What are the differences between the new problem/task/system and the ones that have already been solved/understood/designed?". The answers to these questions give us a way to see a certain pattern in algorithms and data processing (Livingstone & Saeed, 2017). That is how a single task algorithm can be applied for a wider range of problems, which makes the solutions more effective. Thus, whenever a similar new problem arises a general solution is applied. 'Lazy programmers' are expected to avoid writing monotonous and repetitive code, thus eliminating redundancy in the code and a dumb mindset in programmers allows to be open for critics and search for the better solution (Lenssen, 2005). This 'laziness' requires skills of computational thinking such as identifying the pattern (generalisation) and a dumb mindset uses the evaluation skill.

Evaluation

Evaluation is a thinking skill that questions how effective the solution/algorithms are and asks if the algorithm does its job as designed, how fast it is, how economic it is in the use of available resources and checks if it is correct (Livingstone & Saeed, 2017). The answers to these questions help people improve and adapt the solution/algorithms with an extreme focus on attention to detail as each case might require different approaches. For example, if the task is to deal with a certain type of contagious virus infection widely spread worldwide there are several phases to stop this infection. Firstly, everyone is expected to be isolated. Then, if patients have some common symptoms that type of virus infection have, they need to be tested. If the test is positive, self-isolate those patients and everyone they contacted recently. And in case patients get worse, then get them hospitalised and make sure all of them are provided with a ventilator to deliver oxygen in case patients are unable to breathe themselves. Here, in this example, it is important to close all facilities and shut down the whole area (town, city and the country), but more importantly to shut down on time, because early or late actions might also bring negative consequences. Each phase requires different resources thus approaches vary. The judgement made about how fast and available the solution has to be made systematically and carefully, as any change to the solution decisions at any phase play a vital role.

Having described and explained each of the computational thinking concepts, the next section discusses computational thinking in the international and Kazakhstani curriculum.

2.2.4 Computational thinking in the curriculum

There is a trend in introducing computing, programming and digital competencies at early primary school levels in many developed countries (Heintz et al., 2016). An early introduction of computational thinking is important in the modern world (Sanford & Naidu, 2016), like any cognitive skills (Willingham, 2012). Computational thinking can be easily integrated into the computer science curriculum when students are well supported by educators (Flannery et al., 2013). With the help of computer science and informatics courses, students are provided with the power of computational thinking which contributes to a better understanding of other scientific disciplines (Caspersen, Gal-Ezer, McGettrick, & Nardelli, 2019). The UK and US organisations have introduced some definitions and practical applications of computational thinking (CSTA & ISTE 2011; Csizmadia et al. 2015; Peyton-Jones 2014). Kazakhstan has also integrated computational thinking into the national informatics curriculum. There are particular differences in the background and history of the UK and Kazakhstan in computer science education. The UK computing curriculum is the result of long-term planning and teamwork, and it is evidence of the need for a digital government, which requires a digital literacy of citizens and generation (Williamson, 2016). The Computing subject in the UK that replaced the previous ICT, does not focus only on programming as an integrated part but also embraces the larger concepts as a subject. When looked at as a subject, computing, ICT or informatics are all computer science-based subjects, which carry different names but have similar content, roles and values. Computer science is a set of knowledge, which carries new languages to express computational concepts and terminologies related to both software and hardware (Bienkowski et al., 2015). To compare the level of computer science/informatics education in the UK and Kazakhstan, there is a need for objectively and thoroughly analysed rich data from both countries. There are several limitations in getting access and analysis of those data in Kazakhstan. In Kazakhstan, mostly the authors and/or committee members of the related work contexts assess the level of training in computer science; thus, the results can be far more subjective. Unfortunately, there are no reliable recent statistics reports on informatics in the high schools in Kazakhstan. Such statistics and analytics are necessary for the further development and modification of the subject. Indeed, the history of informatics in Kazakhstan is different than in the UK because in Kazakhstan informatics has been a compulsory course for all students since the very beginning period of informatics in the classroom. In general, as it was reported by Gander et al. (2013), there was a lack of computer science/informatics teachers in European countries. Nevertheless, this does not imply that the situation with informatics education is better in Kazakhstan. The fact that informatics as an independent subject is included in the compulsory part of general education in Kazakhstan, is one of the main differences from the UK education system. This factor affects many decisions such as equipping schools technically, training computer science teachers, creating digital educational resources and delivering the curricula both online and offline. However, it does not mean that Kazakhstani school children get informatics classes better than their peers in other countries. The fact that informatics has not existed as an optional subject in old format of UNT until 2017 might have a negative effect on its priorities at schools among other subjects that are compulsory or optional in UNT. When informatics courses are presented as elective-only courses, this electiveness will result in low participation of females and will not attract many other participants (Wilensky et al., 2014). The lack of reliable statistics on school education does not provide the required opportunity for comparisons. In addition, the fact that specialists, educators and researchers in Kazakhstan are less active in international cooperation in the field of computer science education than their peers from Europe or US less contributes to the development of domestic education and its recognition by the international community. Wing (2006b) claims that computational thinking is fundamental and for everyone and also emphasizes that computational thinking is not unique only for computer science. The concepts of computational thinking can be developed by various activities in informatics lessons as well as across the other disciplines. In science classes, students seek solutions by observing to find the key elements of solutions. They can focus on particular data and apply solutions and check the results, if successful then generalise their solution to other problems in other areas.

In the UK, computational thinking lies at the heart of the computing curriculum (Department for Education, 2013b). This is where the new enriched computing discipline came instead of the old “ICT” (Peyton-Jones, 2014). The national computing curriculum in England states that “a high-quality computing education equips pupils to use computational thinking and creativity to understand and change the world” (Department of Education, 2013, p1). The computing subject consists of three parts, computer science, information technology and digital literacy. The main aims of computing, as stated in the UK curriculum (DfE, 2013a), are to ensure that all students: can understand and apply the fundamental principles and concepts of computer science, including abstraction, logic, algorithms and data representation; can analyse problems in computational terms and have repeated practical experience of writing computer programs to solve such problems; can evaluate and apply information technology, including new or unfamiliar technologies, analytically to solve problems; are responsible, competent, confident and creative users of information and communication technology. In the UK curriculum, computing is the main subject that delivers computational thinking for all. Computing as a discipline consists of three strands that have been identified by the Royal Society (Furber, 2012). Each of these strands in Figure 4 is complementary to one another; they are computer science (CS), information technology (IT) and digital literacy (DL), each of which carries significant importance in preparing successful young learners for developing the digital world.

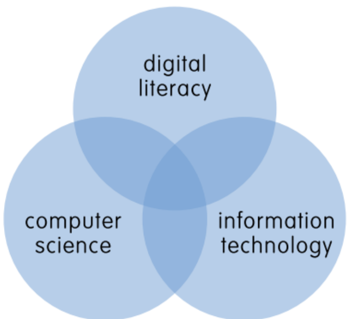


Figure 4 Three strands of computing (Furber, 2012)

The National Academy of Education is one of the leading research and educational institutions in Kazakhstan. Following the Development Strategy of the Education, its main objectives include the following features: scientific and methodological support of the education system, development of compulsory state education standards (curricula and teaching aids); scientific and methodological support for the transition of secondary education for 12 years of study; accumulate, transform and replicate best national and international teaching practice; and conducting fundamental and applied scientific and educational research. The National Academy of Education in Kazakhstan approved the updated curricula in 2016. Kazakhstani secondary schools with the updated informatics curriculum started new academic plans with integrated computational thinking in 2017. Unlike in the UK, there has not been any rebranding of any discipline. The new informatics curriculum for 5th-9th graders includes computational thinking as shown in Table 2.

Table 2 Computational thinking in Kazakhstani new curriculum

	Sections	Subsections
1	Computer systems	1.1 Hardware
		1.2 Software
		1.3 Networks
2	Data processing	2.1. The representation and measurement of data
		2.2. Creation and modification of information objects
3	**Computational thinking**	**3.1 Modelling**
		3.2 Algorithms
		3.3 Programming
4	Health and safety	4.1 Ergonomics
		4.2 Information and online safety

The main objectives of this curriculum are to help students to gain the following abilities (National Academy of Education, 2016):

1. An understanding of the role of information processes in society, the use of information technology in various fields of human activity and its technical opportunities and prospects;
2. An understanding of the usage of information technology in everyday life, in education and applying them more effectively in their future life;
3. An analysis of school systems, solutions, and software applications to create, develop their improvement, as well as allowing them to evaluate their products, their understanding of the basic principles of working with a computer;

4. Teaching students to tackle a variety of tasks through analysis, abstraction, modelling, and programming.
5. Enabling students to gain abilities, such as thinking logically and algorithmically, finding patterns, and thinking in terms of generalisation, decomposition and evaluation.
6. The formation of students' information culture, to follow the general rules and to provide for the interests of an individual and the whole Kazakhstani society;
7. Providing learning the scientific language and enriching the assistance of the conceptual apparatus on subjects.

Comparing items 4 and 5 from the objectives list above with the UK programme (Csizmadia et al., 2015), it is possible to see the similarities between the UK and Kazakhstani computational thinking concepts.

2.3 Programs and tools for computational thinking

This section consists of two parts. The first part discusses some programs delivering computational thinking. The second part discusses several studies and their instruments used for the evaluation of computational thinking. These studies are all relevant as this research aims to measure the computational thinking performance of secondary school students in Kazakhstan. The table in Appendix E shows several research studies by different researchers and organisations.

2.3.1 Programs for delivering computational thinking

As technology is progressing, the number of programs and technologies supporting the development of computational thinking is increasing. As computational thinking is a multidimensional complex construct (Brennan & Resnick, 2012; Grover, 2011), for better integration into education, several recommendations are provided by researchers (Csizmadia et al., 2015; Ozcinar, Wong, & Ozturk, 2017). There is a range of tools from application programs (Gouws, Bradshaw, & Wentworth, 2013a; Grover & Pea, 2013b; Sherman & Martin, 2015) to educational robotics (Ambrosio, Almeida, Franco, & Macedo, 2014; Atmatzidou & Demetriadis, 2015; Berland & Wilensky, 2015; Catlin & Woollard, 2014), various classroom activities for different age ranges (Bell, Witten, & Fellows, 2015), board games (Berland & Wilensky, 2015) that can positively affect the integration and enhance the computational thinking skills of learners. Lee et al. (2011) studied what computational thinking looks like in practice among youth from different cultural and socioeconomic backgrounds, both in and out of school. They focused on identifying strategies for integrating computational thinking into K-12. Modelling and simulation, robotics and game design development are three main domains across which Lee et al. (2011) found common attributes of youths' computational thinking skills. They suggest a rich computational environment can enhance computational thinking skills and one can become not just a user, but also a creator of tools. To enable this transformation from a user to a creator, they offer a three-stage progression "Use-Modify-Create" (Lee et al., 2011 p.35). The "Use" stage is the beginning point, where children are consuming others' products. By using and practising, children start to "modify" the pre-existing model; this is where the tinkering takes place (Csizmadia et al., 2015). Successful tinkering requires concepts such as abstraction and automation. After a series of modifications to others' pre-existing models or products, ownership of those models or products is transferred. Finally, in the "create" stage, children are confident and have enough computational skills to build their own models or products, which requires the concepts of abstraction, automation and analysis (I. Lee et al., 2011). Selby (2012) investigated how programming can be implemented as a tool to teach computational thinking and problem solving, and explored the relationship with a learning taxonomy. Programming and literacy complement each other by forming a new understanding of the harmony of the human mind (Ershov, 1981). Selby (2012) reported the relationship between computational thinking, programming activities and Bloom's taxonomy as shown in Figure 5.

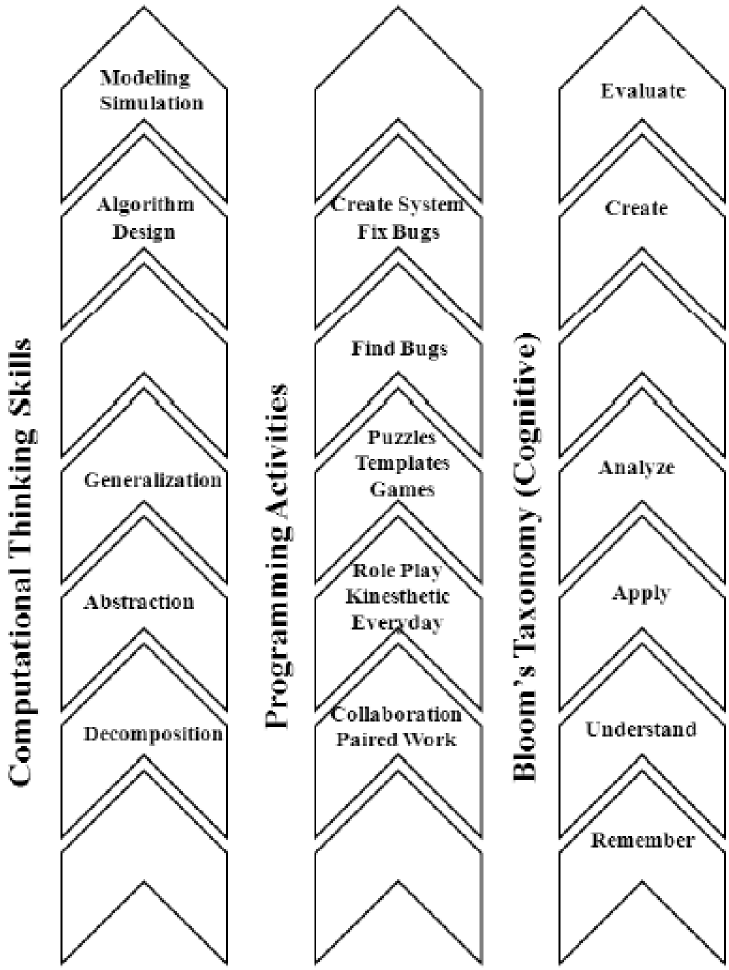


Figure 5 An illustration of computational thinking, programming activities and Bloom's taxonomy by Selby (2012)

New learners perceive computational thinking as computer science when they are first introduced. Moreover, computational thinking is not unique to the computer science discipline (ACM, Code.org, CSTA, Cyber Innovation Center, & National Math and Science Initiative, 2016). Other disciplines are explicitly linked with computational thinking concepts (Weintrop et al., 2016). Students must understand these computational thinking terms and make a distinction (Denning, 2009). It is also worth stating that computer science is not programming and computational thinking is not computer science. However, there is a strong impact of programming on developing the computational thinking skills of students (Brennan & Resnick, 2012; Selby, 2014). Although in some literature, the terms programming and coding are used interchangeably, the term coding is relatively new in computer science. Coding or writing codes is an important part of computer science as it forms computer science's rules. It also reflects computer science's system of computational thinking into the world. Coding converts real-world assumptions into formal models of computers, which can be algorithmically computed (Williamson, 2015). Chuang et al. (2015) investigated teaching computational thinking in K-12 and concluded that there are content differences in each grade. After analysing the questionnaires about skills that need to be taught at each level, Chuang et al. (2015) reported the following statements about the topics to be taught for each level: Before grade 3, children learn problem-solving and decomposition. From grade 5, they start to learn algorithms, data analysis, modelling and simulation, and automation. As the generalisation concept is grasped by children, they will be able to see the connection between what they learn and what is in other fields. Then children learn logical thinking and develop their problem-solving skills. In upper grades 7-9, children begin coding activities; data representation and abstraction are taught during this period. Further, in grades 10-12, as children grow, they start to deal with problems that are more complex and develop their skills in abstraction and generalisation. Data representation, modelling and simulation, and the algorithms are the important topics in which students should be well taught in the final years. At this final level, students should be able to use high-level skills to apply the concepts of abstraction, modelling and decomposition to solve problems in various fields (Chuang et al., 2015), which can be labelled as transfer of the learned skills to other areas (J. A. C. Hattie & Donoghue, 2016; Marton, 2006; Perkins & Solomon, 1992). The Table 3 by Chuang et al.(2015) demonstrates the grades and the corresponding topics.

Table 3 Computational thinking topics for grades by Chuang et al. (2015)

Topics	Essential				optional	
	3	5	7-9	10-12	7-9	10-12
Problem Solving(PS)	1	1	0	3	0	1
Problem Decomposition (PD)	2	2	0	0	0	0
Algorithms (AL)	0	2	2	5	0	3
Data Representation (DR)	0	0	2	6	0	1
Data Analysis (DA)	0	1	1	2	0	0
Modeling and Simulation (MS)	0	1	0	6	0	0
Abstraction (AB)	0	0	2	3	0	2
Automation (AU)	0	2	0	2	0	0
Connections to Other Field (CO)	0	1	2	0	1	0
Total (%)	3 (5)	10 (17)	9 (15)	27 (45)	1 (2)	7 (11)

Hoskey and Zhang (2017) study delivering it to the K-16 computer science community in the United States by refining the distinction of computational thinking from other types of thinking. To successfully deliver and teach computational thinking there is a need for specific exercises related to each skill for teachers and educators (CS Unplugged, 2016; Hoskey & Zhang, 2017). Reaching the aim of delivering computational thinking to everybody requires not only the high-quality content carefully prepared according to the levels of students but also the training and professional development in subject areas of those who teach and deliver computational thinking. Computer science teachers and technology educators need to come together and develop activities that enhance computational thinking in line with the subject area concepts (Yadav, Hong, & Stephenson, 2016b). The Network of Excellence, established in 2012, is a part of the Computing At School group (CAS) in the UK, and it provides networking opportunities by bringing together teachers and professionals. It is actively contributing to the training and supporting teachers as well as maintaining the active community of practice. The Network of Excellence functions as a professional development programme and network of professionals, providing, organising and maintaining local gatherings, meeting face-to-face, peer-to-peer and professional relationships. The quality and impact of the CAS community are increasing through local meetings and networking. The role of the Network of Excellence is very important in delivering the digital skills required in the 21st century, particularly computational thinking skills, to all school pupils. There are 10 CAS regional Centres based in universities across the UK that support the training and recruitment of CAS Master Teachers (BCS - The Chartered Institute for IT, 2017). Professional Learning and Networking in Computing (Computing At School Scotland, 2018) has been established in 2013 in collaboration with Computing At School Scotland and the British Computer Society with similar initiatives by the Network of Excellence. Professional Learning and Networking in Computing focuses on developing pedagogical knowledge and computational thinking skills, rather than training in specific technologies (Computing At School Scotland, 2018). The US K–12 Computer Science Framework by ACM, Code.org, CSTA, Cyber Innovation Centre, and National Math and Science Initiative is a result of such teamwork of different individuals and organizations that contributed to the development of the framework. Beginning from primary level up to final grades, students learn and use new approaches to problem-solving by applying the power of computational thinking skills. There are seven core practices presented in this K-12 Computer Science Framework (ACM et al., 2016): Fostering an inclusive computing culture, collaborating around computing, recognizing and defining computational problems, developing and using abstractions, creating computational artefacts, testing and refining computational artefacts, and finally communicating about computing. These are the behaviours, which show that computationally literate students are engaged with the core concepts of computer science. The lower part practices in the cycle scheme, shown in Figure 6, such as, recognizing and defining computational problems, developing and using abstractions, creating computational artefacts, and testing and refining computational artefacts are computational thinking elements, where the upper ones are general practices. In this core-practices cycle, the level of required skills gradually increases from primary to upper grades.

Figure 6 The seven core practices in K-12 Computer Science Framework by ACM et al. (2016)

The Vienna University of Technology (TU Wien) introduces a special mandatory course “Ways of Thinking in Informatics” for first-year informatics students at TU Wien. The course was designed by Chris Frauenberger and Peter Purgathofer in 2015 to enable informatics students to reflect on Computer Science problems from rich perspectives and different ways and started in 2017 with over 800 registered students at the university. The Ways of Thinking in Informatics course includes cross-sectional topics that interlink with each other and consists of ten different ways of thinking, such as, pre-scientific thinking, scientific thinking, mathematical thinking, computational thinking, design thinking, critical thinking, creative thinking, responsible thinking, criminal thinking and economical thinking (Frauenberger & Purgathofer, 2019). The authors of the course appended the concept of “Analysis” as third A to Wing’s (2006) original Abstraction and Automation concepts in the computational thinking subsection of the course. Thus, expanding the initial two-step into a three-step process by having an iterative loop, which represents how people solve real-world problems more correctly (Purgathofer, 2019a). The course discusses the weakness of rational approaches of computational thinking when used to approach wicked problems, which are difficult to explain and solve due to unstable, incomplete, contradictory and changing requirements. They propose a solution to these types of problems in the design thinking section by introducing the concept of problem settings as the equivalent part of problem-solving (Purgathofer, 2019b). Frauenberger and Purgathofer (2019) claim that the course can serve as a way to ask questions and approach problems from a wider perspective. The integration of teaching and practising computational thinking in schools by classroom activities, educational robotics, programming and coding and also covering broader thinking types show how effectively it can be enhanced. All the studies carried out by various researchers and the recommendations carry important messages that computational thinking skills are an important part of education and they embrace broader areas, not just computer science. Delivering computational thinking skills opens the gate for innovations in various fields and areas, such as engineering, science and social science, humanities, arts, healthcare and medicine, business and education (Gretter & Yadav, 2016).

2.3.2 Evaluation of computational thinking

Without considering any assessment, computational thinking is unlikely to successfully advance in any curricula without proper evaluation and measurement of computational thinking. Also, to judge the effectiveness of any curricula integrating computational thinking, the measures and assessment tools should be validated to allow teachers to see what children learn (Grover & Pea, 2013a). Jan Cuny points to the importance of standardised tests that can assess whether students can think computationally; and also claims that in the absence of a consensus on the scope and the nature of computational thinking it will be hard to create an assessment tool to measure computational thinking (Linn et al., 2010). The evaluation of computational thinking is as important as its integration into curricula. When focusing on the evaluation of computational thinking, it is essential to consider various approaches that support the development of computational thinking in terms of context, time and structure (Brennan & Resnick, 2012). There are a limited number of studies comparing the development of computational thinking skills between males and females (Atmatzidou & Demetriadis, 2015), and some studies found gender differences and suggest that females should be encouraged to engage with computational thinking skills (Grover & Pea, 2013a). However, other studies found no gender difference in terms of attitude towards computational thinking (Lau & Yuen, 2009; Weintrop et al., 2015) There are different types of assessments and evaluations of computational thinking, such as project portfolio analysis, artefact-based interviews, design scenarios (Brennan & Resnick, 2012), self-reported instruments (Doleck, Bazelais, Lemay, Saxena, & Basnet, 2017; Korkmaz, Çakir, & Özden, 2017) and multiple-choice items (Mindetbay, Bokhove, & Woollard, 2019a), which are developed using various programs, applications and projects, beginning from elementary school and higher up to the university level. Several studies and recommendations were presented on the measurement of computational thinking skills designed for broad age groups, from young children and up to adults. Various software and technologies were developed and used to carry out these assessment tasks.

Yadav et al. (2011) studied the implementation and assessment of a computational thinking module in a compulsory course for elementary and secondary education levels. They used sixteen multiple-choice questions and open-ended questions, before and after the special course, to assess 100 participants' attitudes and understanding of computational thinking. Yadav et al. (2011) found that after providing relevant information about computational thinking, students' attitudes positively changed toward computer science, and students gained a better understanding of how they can benefit from computational thinking in their future careers by enhancing problem-solving and critical thinking skills. With the introduction of computational thinking in education, its integration into game programming has been discussed and studied by many researchers. Several specific design guidelines were proposed that help computational thinking skills to effectively integrate into the learning environment, which in turn can foster students' computational problem-solving skills (Pellas & Vosinakis, 2017). Programs, applications, environments and platforms such as ScratchJr, Scratch, Dr. Scratch, Alice 3D, Stagecast Creator, Fairy, LightBot and App Inventor we used by several researchers to evaluate the computational thinking skills of children. Grover (2011) carried out a study with ten students involved in a 5-day (8 hours/day) long Robotics and Engineering workshop. He used a prior knowledge survey and a pre-and post-interview to evaluate the elements and dimensions of computational thinking of the participants. In his study, taxonomic categories such as broad concepts, principles, vocabulary and procedural/operational ideas of computational thinking were developed to refine dimensions of computational thinking. Grover's results show students' computational thinking language increased both quantitatively and qualitatively. Grover (2011) consider the transfer of computational thinking skills to other areas is worth investigation. Werner et al. (2012) studied the performance assessment of a game-programming course and reported the factors affecting the performance, then discussed the aspects of computational thinking. Their research included 311 students' work results from seven public schools on the central California coast, USA. The participants were involved in Fairy assessments, in which Werner et al. (2012) were investigating how game creation and pair programming can enhance computational thinking. Their results suggest that the Fairy assessment successfully demonstrates different aspects of computational thinking across the assignments. Brennan and Resnick (2012) investigated the ways design-based learning activities, particularly Scratch programming, support the development of computational thinking in young learners. They set a computational thinking framework that consists of three dimensions: computational concepts, computational practices and computational perspectives. Brennan and Resnick (2012) examined the assessment of these dimensions by project portfolio analysis, artefact-based interviews and design scenarios. They also listed the strength and limitations of these assessment types. One limitation of the project portfolio analysis is that it cannot reveal how children develop their skills while working on projects, as it is product-oriented. To reveal how children can shape the process of development of their computational thinking skills is discussed by Grover and Pea (2013b) that offers a discourse analysis in computer science curriculum for children. Portelance and Bers (2015) conducted a research where children worked in pairs to create their projects on ScratchJr (DevTech Research Group & Playful Invention Company, 2018), a mobile version of Scratch for children who are 5-7 years old. As an assessment method, they followed a peer video interviewing technique with sixty-two 2nd grade primary school children as recommended by Brennan and Resnick (2012). Portelance and Bers (2015) state that children can demonstrate a wide range of computational thinking concepts when they present and explain their ScratchJr projects.

Werner et al. (2014) introduced a new concept of Game Computational Sophistication, by which it is possible to see children's understanding of how to use their thinking skills to design and build complex and dynamic systems in an engaging game environment. Brennan and Resnick (2012) who studied Scratch programming, listed six suggestions when assessing computational thinking through programming and claimed that learning takes place when children are engaged in programming activities. Children demonstrated that they can think like computer programmers when they are engaged in programming (Werner et al., 2015). However, Denner et al.'s (2012) research findings did not confirm their expectation that the more children engaged in programming games the more complex programmes they can produce. The results of Denner et al.'s (2012) study to evaluate computational thinking skills via game programming demonstrated moderate levels of complexity, where they studied 108 games created by fifty-nine middle school girls who worked on the Stagecast Creator environment to create games. Denner et al. (2012) concluded that game designing by programming activities can support the learning of computer science concepts. Gouws et al. (2013a) also stated that an educational game (Light Bot), which is focused on programming, is very helpful in learning computer science concepts. Werner et al. (2012) were also interested in measuring children's computational thinking by game programming, particularly in the Alice environment. Three hundred and twenty-five students from the USA participated in their study. Participants' performances were analysed based on three given tasks on the Alice platform. Werner et al. (2012) concluded that Alice is an interesting platform to assess children's computational thinking skills and their ability to apply those gained skills. Grover and Pea's (2013b) research results show that computational discourse plays an important role in a computationally rich learning environment, and enhances the development of computational thinking skills. App Inventor for Android was utilised as a tool for programming in Grover and Pea's (2013b) study that focused on the discourse role on learning. They describe "computational discourse" as a process, in which students are introduced to computer science techniques through gaining computational thinking competencies in a computationally-rich environment by productive talks. Gouws et al. (2013b) studied the assessment of the computational thinking abilities of students at the Rhodes University in South Africa by using a test they have developed. They compared eighty-three students' computational thinking test results with their class marks as an initial checkpoint, before starting the computer science course CSC 101. They constructed twenty-five test questions based on six classifications of computational thinking for the two-phase (pre- and post) testing. Gouws et al. (2013b) were interested in how students' initial level of computational thinking will change after the computer science course, and if there is any relationship between students' computational thinking abilities and their class marks. Their findings show that students with higher results in the computational thinking test are also well on their class tests.

A different approach has been suggested by Babbitt (2013) who studied how using ethno-computing can confront the myths of cultural-genetic determinism. He defines ethno-computing as a simulation of indigenous and vernacular cultural practices in computer science. According to Babbitt, an important goal of ethno-computing is enhancing the development of the computational thinking skills of students. His research was carried out by using Culturally Situated Design Tools and the mental model approach, which helped to formulate the assessment. The Progression of Early Computational Thinking Model (PECT) was proposed by Seiter and Foreman (2013) to assess the computational thinking level of primary school students of grades 1-6 in the USA. The PECT model is based on the design patterns concept that categorises computational thinking through the level of skill used in design patterns in Scratch programming. Children's level of knowledge in the design patterns is mapped into computational thinking concepts, such as procedures and algorithms, problem decomposition, parallelisation, abstraction and data representation. After piloting their framework to measure its effectiveness in identifying differences in computational thinking level among children of different grades, Seiter and Foreman (2013) reported that the PECT model was effective in monitoring students' progress in computational thinking.

The Real Time Evaluation and Assessment of Computational Thinking (REACT) project was developed by Koh et al. (2014) to help teachers to see in real-time which concepts students have understood and which ones they have not. The REACT is a formative, real-time graphical assessment tool that provides teachers with insights into students' mastery of computational thinking constructs as they are coding and designing games and simulations. Koh et al. (2014) claim that the REACT supports and facilitates the teaching the computational thinking skills by providing faster and more detailed information than other similar assessment systems. Teachers can identify the weakness and strength of students' understandings by looking at the patterns in their computational thinking abilities, even before students start to implement their programs.

One hundred and thirty-four 6th grade students participated in the study, which lasted more than four weeks. Koh et al. (2014) stated that there was very positive feedback from teachers for this formative assessment tool.

Ambrosio et al. (2014) conducted a study with twelve freshmen students from the University of Minho in Portugal. Their research interest was exploring how psychological assessment tests can be used to evaluate the cognitive processes of computational thinking related to programming activities. After identifying four central cognitive processes, Ambrosio et al. (2014) selected four validated cognitive tests, such as a general intelligence test, a spatial reasoning test, a mathematical reasoning test, and an attention to details test. The result of their research shows a high correlation between problem-solving and mathematics and students' academic success in a related field and suggest that spatial reasoning and general intelligence are important aspects of introductory programming. However, participants' performance in solving mathematical problems, which was limited to simple calculations, shows a low correlation with performance in problem-solving. Later, in 2015, Ambrosio et al. (2015) studied the application of a software tool that was developed to test computational thinking performance. This software is a digital ink "InkML", which computerises the assessment procedures. They suggest that the InkML tool may be utilised in any discipline not only in programming or computer science. The Scratch programming performance of students was studied by Seiter (2015), who evaluated students' responses based on the SOLO (Structure of the Observed Learning Outcome) Taxonomy designed by Biggs and Collins, to scaffold higher-order thinking for students. Seventy-two 4th grade students from two primary schools in the USA participated in their study. Seiter (2015) conducted three assignments at different periods during the fourteen-week long Scratch programming instruction course. The results indicated that students can learn to synchronise multiple concerns within a single script, along with a single concern across multiple scripts, however, synchronization of multiple concerns across multiple scripts was difficult. Seiter (2015) claims that the SOLO taxonomy can be used both to identify students' levels of understanding of a problem's structure and to assess students' programming skills.

Grover (2015) studied multiple approaches to measuring algorithmic thinking skills and their transfer to other domains along with non-cognitive aspects of computer science, such as communication of computational thinking ideas and perceptions of computing. His systems of assessment were conducted through the Foundations for Advancing Computational Thinking (FACT) computer science course. The six-week course that was created on the OpenEdX online platform used a blended-learning approach for twenty-eight 7th-grade and 8th-grade students in North Carolina, USA. Using the FACT students' understanding of computer science concepts was tested and they were provided with immediate feedback on their understandings. Grover's (2015) findings showed that pre-test results are the predictors of post-test results, and students' learning increases after taking the course. Students showed an average score on an understanding of the algorithmic flow of control in a text-based programming language, although those students who scored low were highly engaged in the final projects and also showed good performance on an understanding of algorithmic constructs. Grover (2015) claims that none of the assessments that have been conducted would be successful in the assessment of both cognitive and transfer aspects on their own without a holistic approach to assessment. Zhong et al. (2015) developed a three-dimensional integrated assessment of computational thinking at Carnegie Mellon University. In their three-week-long experiment, four classes were randomly selected at a primary school in Changshu City, China with one hundred forty-four participants. The three-dimensional integrated assessment is designed to integrate three dimensions, such as, directionality, openness and process into assessment tasks, then to assess the three dimensions of computational thinking, which are computational concepts, practices and perspectives. The participants of the study were engaged in an easy-to-learn environment and a 3D programming language, Alice2.4, where they built virtual worlds. The results of Zhong et al. (2015) experiment shows that the reverse tasks and the forward tasks score the same; the semi-open and the open tasks were easier to carry out and reflects more dimensions of computational thinking, and the self-reports are useful for guidance and learning diagnosis, and there is no score difference between males and females. Weese (2016) studied the development of a visual programming curriculum for teaching computational thinking at K-12 to examine how effective the curriculum is in delivering computational thinking. Weese's study aimed to develop mixed methods to assess computational thinking. The qualitative data from a survey on self-efficacy and quantitative data from students' programming data and problem-solving traces were analysed. Weese et al. (2017) studied the measurement of self-efficacy in computational thinking claim that computational thinking is not simply problem-solving. Weese et al. (2017) concluded that self-efficacy in computational thinking can be a good predictor of students' learning outcomes.

Moreno-Leon et al. (2016) used Dr. Scratch, a free open source software assessment tool for Scratch projects, and compared the results with two validated and globally accepted classic software engineering metrics, Halstead's and McCabe's Cyclomatic Complexity. The result of their research indicates the strong correlation between those metrics and Dr. Scratch's outputs, showing a validation of the complexity assessment process of Dr. Scratch. Dr. Scratch enables students to measure their projects with computational thinking scores. Dr. Scratch is widely used in education, as it is a useful tool both for students to develop their skills and for teachers to automate the assessment of Scratch projects. Developing and exploring automated assessment systems and tools are gaining high interest among researchers. Oluk and Korkmaz (2016) also used Dr.Scratch and compared students' Scratch scores with their computational thinking skills. The questionnaire of computational thinking scale level that covers five factors, such as, creativity, problem-solving, algorithmic thinking, collaboration and critical thinking was administered to measure participants' perception of computational thinking. Thirty-one 5th grade students were introduced to Scratch programming techniques and their works were evaluated via Dr. Scratch. Oluk and Korkmaz's (2016) research findings showed no difference in gender and period of computer use. However, the study found a strong positive relationship between students programming skills with Scratch and their computational thinking skills; students' with better programming skills in Scratch tend to be better computational thinkers or vice versa. The paper by Román-González, Pérez-González, Moreno-Leon, and Robles (2017) studied the non-cognitive sides of computational thinking abilities of children. The authors looked at the correlations between computational thinking skills and five dimensions such as, conscientiousness, openness to experience, extraversion, agreeableness and neuroticism, from the child version of the Big Five questionnaire for 8-15 year old children, which is adapted from the original Big Five questionnaire. As a result, the computational thinking test for computational thinking skills was developed, which consists of twenty-eight multiple-choice items that cover the following concepts: sequences, loops and repetitions, if and if/else conditionals and simple functions was developed. Ninety-nine 5th-10th grade school students took the computational thinking test and the Big Five questionnaire. Román-González et al. (2017) found a statistically significant correlation between computational thinking and perception of their self-efficacy, openness to experience, extraversion, and conscientiousness, as they expected, and no correlation with agreeableness and neuroticism factors. Their research results support the idea that computational thinking is mainly a cognitively psychological construct close to problem-solving skills and also state that there is a non-cognitive side of computational thinking (Román-González, Pérez-González, et al., 2017). Roman-Gonzalez et al. (2016) investigated the reliability and criterion validity of the computational thinking test they designed with respect to the psychological tests such as the Primary Mental Abilities battery and the RP30 problem-solving test. The number of participants was 1251 students from 5th-10th grade in Spain. Roman-Gonzalez et al. (2016) examined verbal, spatial, reasoning and numerical factors as well as the speed and flexibility of children when they were dealing with logical operations. They concluded that computational thinking correlates with spatial, reasoning and problem-solving abilities.

Jamil (2017a) proposed a fully automated online assessment system, MindReader, which understands different programming languages, such as C++, Java and Python. The MindReader can evaluate students' written codes and give them feedback to improve their programmes. The idea behind this approach in assessing the programming task is based on the matching of concept dependence graphs and program dependence graphs. This automated system could be used by MOOCs as an online assessment solution. It is expected that MindReader can help self-learning students by providing accurate feedback once it has been thoroughly tested.

Korkmaz et al. (2017) developed an instrument, computational thinking scales, to evaluate the computational thinking skills of the students. The instrument is a five-point Likert type scale with twenty-nine items, which covers five factors of computational thinking. The factors are creativity, algorithmic thinking, cooperation, critical thinking and problem-solving. The participant group comprised 1306 students at Amasya University in Turkey. Korkmaz et al. (2017) have conducted exploratory factor analysis, confirmatory factor analysis, item distinctiveness analysis, internal consistency coefficients and constancy analyses to check for the validity and reliability of the developed instrument. They concluded that the instrument is valid and it is a reliable measurement tool to assess the computational thinking skills of students. This scale was adapted to school students' level by Korkmaz et al. (2015). The number of items was reduced to 22 with the same five factors of computational thinking. The participants of this study were two hundred and forty-one 7th and 8th grade students at secondary school in Turkey. The study conducted by Durak and Saritepeci (2017) was aimed to find out how many various variables (gender, age, computer usage period, internet usage period, daily internet usage period, attitude against mathematics, point average in mathematics, attitude against science class, point average in science class, attitude against information technologies class, point average in information technologies class) explain students' computational thinking skills. They developed a model, which can predict and explain the relationship between the computational thinking skills of students and the variables listed above. One hundred and fifty-six students from grades 5-12 participated, and they have provided answers to the Personal Information Forms and Computational Thinking Skills Scale questionnaires. The findings by Durak and Saritepeci (2017) show a strong relationship between computational thinking skills, thinking styles, academic success in mathematics and attitude toward mathematics. Their results also support the assumption that computational thinking skills are closely related to problem-solving skills. However, computational thinking skills are negatively predicted by education levels; and the reason for this negative relationship might be region-specific (Durak & Saritepeci, 2017).

There are several contests and competitions on informatics and programming, as well as problem-solving and computational thinking. The “Bebras” International Challenge on Informatics and Computational Thinking started in 2004 in Lithuania (Bebras Community, 2015). The main goal of the “Bebras” is to engage children in informatics topics and to promote algorithmic, logical, operational and computational thinking. From the primary level, the Bebras tasks and activities enhance children's interest in computing and help them master the technology easier. The Bebras challenge has been designed for all primary and secondary school students. The tasks are computer-based and performed at schools. There are different problem sets for each age group: Little Beavers (grades 3–4), Benjamin (grades 5–6), Cadet (grades 7–8), Junior (grades 9–10) and Senior (grades 11–12). The Bebras contest is not aimed at checking the knowledge of students, but over a long period, serve as an assessment system, supporting the concept-based learning of informatics and computational thinking (Dagiene & Stupuriene, 2016). The problems and exercises are based on the operational definition of computational thinking by ISTE and CSTA (2011). Since 2004, the number of Bebras participants from all around the world has been significantly growing. Kazakhstan joined Bebras in 2016 with 2995 participants. The quality of the challenge is increasing by constructing more complex tasks and activities related to informatics educations (Dagiene & Stupuriene, 2016). The aim is to provide enriched tasks and activities that support the development of understanding the informatics concepts and their practice. Dagiene & Stupuriene (2016) claim that a Bebras challenge is an effective tool that promotes computational thinking and problem-solving skills. Román-González (2017) developed the Computational Thinking Test (CTt), a multiple-choice instrument that consists of 28 items with a maximum duration of 45 minutes. Each item of the CTt is constructed according to the following three dimensions: 1. Computational concept: Sequences, loops, if-else, complex conditionals and functions. 2. Answers were either visual blocks or visual arrows. 3. Students were required to use commands in an order by sequence, complete the unfinished set of commands and debugging the incorrect commands in given tasks.

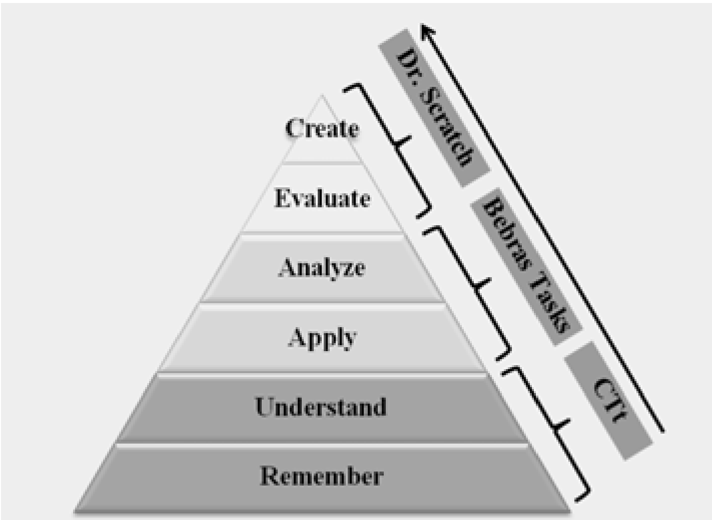


Figure 7 Bloom's revised taxonomy and computational thinking assessment tools by Román-González (2017) - Complementary Tools for Computational Thinking Assessment

The CTt, Bebras Tasks and Dr. Scratch are complementary to courses in middle school settings and they form a system of assessments, as shown in Figure 7. International Computer and Information Literacy Study (ICILS) is an international comparative study, which assesses the extent to which students know, understand, and use computers and information technology, and participate actively in the digital age (IAEEA 2015). The main goal of ICILS-2018 is to find out how well students are ready for work and life in the digital age, and how their overall performance compares with other students from other countries. The assessment measures and analyses differences in the computer and information literacy levels, computer use and computational thinking skills of 8th-grade students. This large-scale study also gathers information about the contexts, where students develop their computer and information literacy, both in and outside the school. The collected rich data by ICILS from different countries enable to investigate factors, which affect students' computer and information literacy, computer use and computational thinking skills. The gathered information is valuable for education systems, stakeholders, policymakers and computer literacy-based education programs. The computational thinking domain is included in the test in 2018 as the process of utilising computers in problem-solving, with the focus on structuring, manipulating data sets and programming (IAEEA 2018). Principled assessment of computational thinking (PACT) was developed by Stanford Research Institute (SRI) Education, which focuses on a principled approach to assessment tasks that can provide valid evidence of students' abilities to think computationally. To assess computational thinking, they focus on a special course - Exploring Computer Science. The principled assessment aims to evaluate important knowledge and practices by setting chains of evidence, which provide claims about what students know and what they do in fact. The design patterns obtained from this evidence guides assessment specialists in constructing tasks to assess both knowledge and skills in specific learning experiences. Bienkowski, the deputy director of SRI International's Centre for Technology in Learning, claims that people are still searching for the unanswered questions, 'what is coding vs. what is computational thinking' (Bienkowski et al., 2015). Dorling and Walker (2014) indicated the concepts of computational thinking on the Computing Progression Pathway, which was prepared by a team of authors and reviewers. It is a useful guide for teachers to identify concepts to teach in the UK computing national curriculum, which was introduced in 2014. This framework consists of five categories, each of which is denoted as algorithms, programming and development, data and data representation, hardware and processing, communication and networks. There are several levels shown in rows, where each row represents a certain level of students' progression. Selby et al. (2015) explored how the Computing Progression Pathways can be utilised as direct evidence for the assessment of computational thinking. This assessment framework by Dorling and Walker (2014) enables teachers with minimum experience to assess the computational thinking skills of learners. The studies on the efficiency of automated assessment systems of computational thinking and programming are gaining interest (The Royal Society, 2017). An automated assessment system of computational thinking could ease the process, but it is very complex and difficult to build. Instead, students themselves, peers, parents and teachers can be involved in the assessment process. Students' explanations of their solutions and problem-solving techniques may suggest some improvement to the assessment of computational thinking (Brennan & Resnick, 2012). What should be taught, how to deliver and assess the computational thinking skills, along with other pedagogical and educational roles of computational thinking still need close attention (Tedre & Denning, 2016). Apart from various assessments of programming abilities and computational thinking skills of students, there is a need for large-scale studies to measure students' computational thinking skills and their conceptual understanding (Kallia, 2017). Allsop (2019) explored multiple approaches of evaluation of computational thinking process in a classroom environment based on a longitudinal study. The study took 8 months of working with 10 and 11-year-old primary school students using Scratch and Alice 2.4 in a game-making project. Allsop (2019) claims that there is a need for multiple evaluation approaches to show the whole learning scope of the computational thinking process. To assess the elements of the computational thinking process in a computer game design rich data from observations, informal conversations, problem-solving sheets, semi-structured interviews and students' completed games were used. The multiple approach evaluation model was introduced by Allsop (2019) with the following elements of the computational thinking process: computational concepts, metacognitive practices and learning behaviours. A Beginners Computational Thinking Test (BCTt) is an example of a validated and stand-alone instrument designed to measure computational thinking at the Primary School level which does not depend on any specific programming language (Zapata-Caceres, Martin-Barroso, & Roman-Gonzalez, 2020). The BCTt is a multiple-choice test with three answer options where several elements and representations of computational thinking have been used. Assessment of computational thinking is an important part of all educational levels and phases. Since computational thinking is multi-dimensional, it is hard to measure. With regards to the assessment of computational thinking, Tedre & Denning (2016) suggest assessing the competencies rather than factual knowledge. The

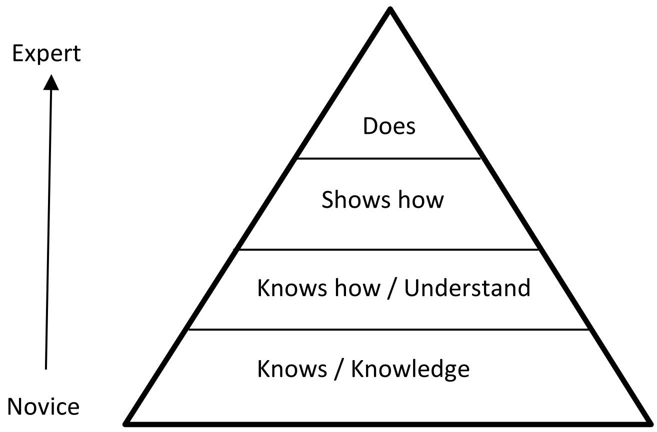


Figure 8 Miller's pyramid

assessment of competence is an important indicator in any setting. As an example, for the assessment of medical competence, Miller (1990) introduced a pyramidal structure of the cognitive domain. Miller's Pyramid of cognitive domain consists of four levels "Knows," "Knows How," "Shows How," and "Does", each of which requires specific methods of assessment.

From the lowest level “Knows” to the highest level “Does”, the pyramid represents the level of competency and proficiency from a novice up to an expert as shown in Figure 8, which can also be used to assist in the assessment of professionalism. And well-constructed multiple-choice questions can be utilized for the assessment of the first three to four levels of Miller’s Pyramid (Tarrant, Knierim, Hayes, & Ware, 2006). The studies listed in Appendix E illustrate some shortcomings like small-size samples, tool-specific assessment and self-reporting evaluations. It was reported that there is a need for large-scale studies of measuring computational thinking that could be linked with other skills or domains (Kallia, 2017; Román-González, Pérez-González, Moreno-Leon, et al., 2016). Also, there are only a few studies found on the transferability of computational thinking skills to other fields (Basawapatna, Han Koh, Repenning, Webb, & Sekeres Marshall, 2011). To get a larger picture of the relationship between computational thinking, large-scale studies that focus on computational thinking skills should be carried out (Kallia, 2017).

2.4 Relationship between computational thinking and other domains

As computational thinking become more popular, the number of studies on this topic increased correspondingly. Several research studies claim that computational thinking can be transferred, while some claim that the transfer is limited. Although the transfer is expected from a cognitive psychology perspective (De Corte, 2003), some claim that computational thinking skills do not transfer to other environments (Tedre & Denning, 2016), thus this section explores relationship studies that are closely linked to transfer (Grover, 2011).

This section lists and summarises several types of research that study the relationship between computational thinking skills and other domains using numerous approaches and from various perspectives. Selby (2015) introduces a model that demonstrates the relationship between computational thinking skills, programming activities and the cognitive domain of Bloom’s taxonomy as shown in Figure 9.

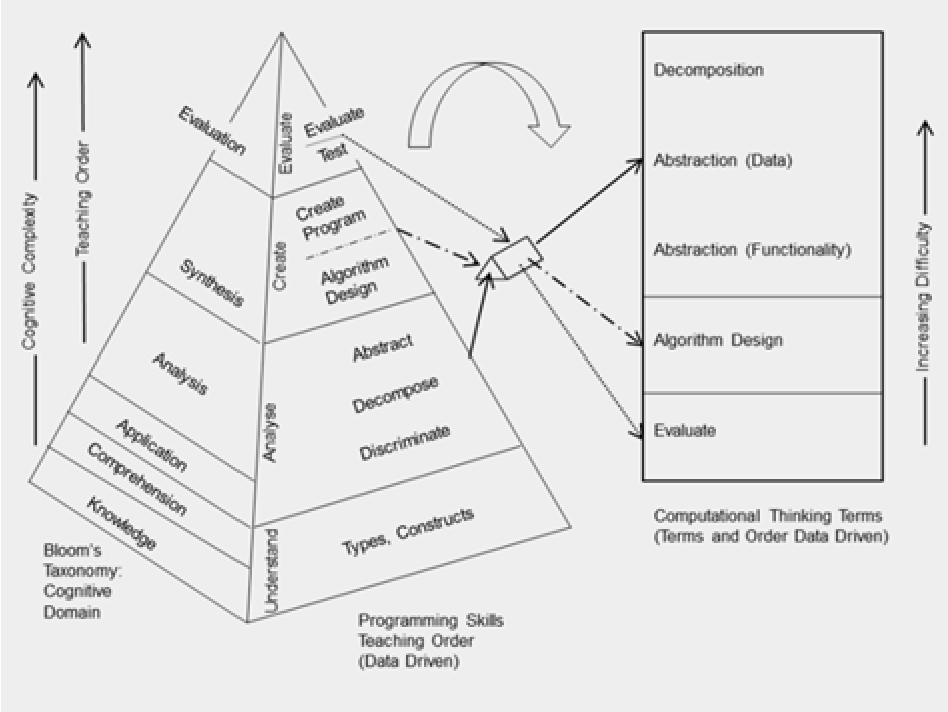


Figure 9 A relationship model by Selby (2012)

A study by Werner et al. (2012) investigates the performance assessment of a game-programming course and the results show that the students who use computers more frequently and with better academic achievement demonstrate a higher level of computational thinking. The results of Denner et al. (2012) support these findings and claim that students who are engaged in game programming are more likely to be in the state of active thinking, which prepares them for the next level of computing. Hu (2011) claims that computing is more mathematical, like recognizing and manipulating patterns; and computer programming, a significant form of computing is mathematical activity. Hu (2011) adds that the more people do in computation, the better computational thinkers they get; and learning computing while solving STEM problems foster computational thinking abilities. Werner et al. (2015) investigate the relationship between computer game programming and computational learning. They offered a new concept, the Game Computer Sophistication, to define and measure computational learning in children and analysed 231 games that are programmed by 325 children. Students who used programming skills to create computer games using Alice showed that they can think like computer programmers (Werner et al., 2015). Scratch programming offers an easy step into programming for beginners with its block programming features, this is why it is very popular in primary schools. A study by Brennan and Resnick (2012) proposed six suggestions for assessing computational thinking by programming. According to Brennan and Resnick's findings, those children who programmed games by using Scratch demonstrated that they can think algorithmically, use abstractions and see patterns. Oluk and Korkmaz (2016) analysed 5th-grade students' Scratch projects by Dr. Scratch. Their findings show a strong relationship between students' programming skills with

Scratch and their computational thinking skills. The study by Pellas and Vosinakis (2017) examines how to design a simulation game to foster the development of computational problem-solving techniques by learning the fundamental computer science concepts and introduces a theoretical game playing framework. This theoretical game playing framework with several game design guidelines, which includes concepts of computational thinking such as abstraction, decomposition, algorithmic thinking and evaluation, and this framework can be used to foster students' thinking about a real-life problem (Pellas & Vosinakis, 2017). Not only computer games but also various types of technology toys and board games are used in the learning process of computational thinking (Lin, 2015). Findings of the collaborative board game Pandemic is tested by Berland and Duncan (2016) if board games can help develop players' computational thinking skills. Some minor modifications were made to see the differences in the character and the level of computational thinking skills applied by the players during discourse exchange. The slight modifications in the Pandemic game allow players to express more complex logic in their decision-making discussions, which show that social elements of board games might have interactions with their computational systems (Berland & Duncan, 2016). Tseng, Doll, and Varma (2019) investigate a way to deliver the computational thinking skill abstraction by means of CAD games Ghost Blitz vs. Sushi Go. Their study explores the strategies used by students in gameplay, which requires multiple computational thinking skills. Three hundred and sixty-five students' gameplay strategies are examined whether players applied abstraction skills spontaneously in planning to reach their goals during the game. The results of the work show that playing board games increase the chances for players to learn pattern recognition (Tseng et al., 2019), which corresponds to the generalisation concept of computational thinking. The assessment methods of computational thinking and the findings of the relationship studies between computational thinking aspects and any other domain heavily depends on how computational thinking is defined, as domain-specific or domain-general, and what approach is used in terms of assessment. Some studies that carry domain-general approaches to computational thinking investigate the relationship between multiple intelligence such as general intelligence (Boom, Bower, Arguel, Siemon, & Scholkmann, 2018), spatial intelligence or spatial thinking (Ambrosio et al., 2014; Città et al., 2019; Ham, 2018), cognitive and non-cognitive psychological constructs (Román-González, Pérez-González, Moreno-Leon, et al., 2016), self-efficacy (Román-González, Pérez-González, et al., 2017; J. Weese & Feldhausen, 2017) and thinking styles (Durak & Saritepeci, 2017). Whereas, the studies with domain-specific approaches reveal how computational thinking skills are linked to non-digital board games (Berland & Duncan, 2016; Tseng et al., 2019), digital computer games (Denner et al., 2012; Werner et al., 2015, 2012), programming (Selby, 2012, 2014), logical reasoning, algorithmic thinking, mathematical thinking, abstraction, generalisation, academic achievement, problem-solving, specific computational creativity (Hershkovitz et al., 2019; Mindetbay, Bokhove, & Woollard, 2019b) and digital competences (Juškevičiene & Dagiene, 2018). Román-González, Pérez-González, Moreno-Leon et al. (2016) investigated the correlations between computational thinking skills and human personality within five dimensions such as conscientiousness, openness to experience, extraversion, agreeableness, and neuroticism. They found statistically significant correlations between computational thinking and openness to experience, extraversion, and conscientiousness. Their results support the idea that there is a non-cognitive side of computational thinking skills. Self-efficacy is yet another aspect that plays role in computational thinking studies (Román-González, Pérez-González, et al., 2017) if to consider that it can be a good predictor of students' learning outcomes (J. Weese & Feldhausen, 2017). Spatial thinking is considered one of the main components in the development of thinking (Città et al., 2019; Ham, 2018), as it is closely related to the ability to see the problem and mentally work it out (Ambrosio et al., 2014). Spatial intelligence introduced by Gardner (2011) is one of the components of different intelligence building human thoughts reasoning. Several studies found general academic achievement to correlate with computational thinking (Durak & Saritepeci, 2017; Gouws et al., 2013b; Rodrigues, Andrade, & Campos, 2016; Weintrop et al., 2016). Specifically, mathematics is a subject that has a long root in computational thinking (Weintrop et al., 2016), which makes it inseparable from computational thinking skills, and computing has many mathematical aspects that are similar to computational thinking (Hu, 2011). Hershkovitz et al. (2019)studied the associations between creativity and computational thinking skills among fifty-seven school students. Two types of creativity were considered: general creative thinking, and specific computational creativity. As for the computational thinking skills of the students, a game-based learning environment, the Kodetu app was used, which teaches basic programming skills. As for creativity, a paper-based creativity task by Torrance Test for Creative Thinking (TTCT) was used. Hershkovitz et al. (2019) found some association between Computational Creativity and Computational Thinking, but no associations were found between Creative Thinking and Computational Thinking. Durak and Saritepeci (2017) investigated the relationship between computational thinking and various variables and created a model of these relationships with a participant group of 156 students who were studying in 5-12 grades. Their study found a strong relationship between computational thinking skills, thinking styles, academic success in mathematics, and attitude toward mathematics. Their results also support the assumption that computational thinking skills are closely related to problem-solving skills. However, computational thinking skills were found negatively predicted by education levels.

The IEA's (International Association for the Evaluation of Educational Achievement) assessment framework, the International Computer and Information Literacy Study 2018 (ICILS) is a high profile internationally respected assessment. ICILS 2018 is designed to measure international differences in students' computer and information literacy and reveals the larger picture of how well students are prepared for study, work and life in today's digital age. The results of the ICILS-

2018 show that the computational thinking scores of the students were significantly different depending on various factors (ICILS, 2020). Students with 26 or more books at home score higher in computational thinking than those who have less than that. Students with parents who completed a Bachelor's or higher degree score higher than the ones with no degree parents. Likewise, students of a parent with higher occupational status have higher computational thinking scores than those with medium or lower. On average, girls perform better than boys do in computer and information literacy but boys perform better than girls do in computational thinking. ICILS 2018 reports that students from higher socioeconomic backgrounds have significantly higher computational thinking scores. They also state that students do not develop advanced digital skills just by growing up using digital devices, which means supporting students with Information and Communications Technology (ICT) equipment alone is not enough for them to master the digital skills. Students should be taught how to use computers effectively and teachers need support in their use of ICT in the classroom. ICILS 2018 also reports that teachers are more likely to promote Computer and Informational Literacy and Computational Thinking in a classroom environment when they are confident users of ICT (ICILS, 2020). Kazakhstani results in the ICILS-2018, in which Kazakhstan participated for the first time, report the following: In the ICILS-2018 fourteen countries participated and Kazakhstan showed the lowest results. The Kazakhstani participants were from all regions of Kazakhstan and there were statistically significant differences depending on the type of school, the language of test and location. 33% of participants could not reach the minimum level of computer and information literacy with an average score of 395, which is 101 points lower than the international ICILS-2018 average. 1% of Kazakhstani 8th grade participants attained the highest level of computer and information literacy with scores above 661 points ("Kazakhstani results in the ICILS-2018 International Computer and Information Literacy Study," 2020). Juškevičiene and Dagiene (2018) investigated the relationship between digital competence and the computational thinking of schoolchildren. They describe digital competence as a skill that involves the confidence, critical and responsible use of technologies and engagement with digital technologies for learning, at work and in social life, which includes the followings: information and data literacy, communication and collaboration, programming, digital content creation, cybersecurity and safety and problem-solving. The analysis of possible interconnections demonstrates that there are commonalities between the digital competencies (DC) and computational thinking (CT), where many abilities and competencies are overlapping as shown in Figure 10.

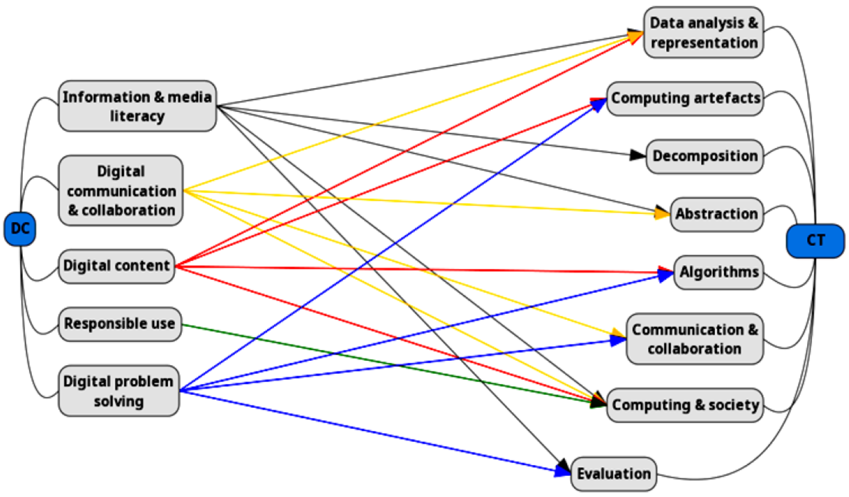


Figure 10 The interconnections between digital competence and computational thinking by Juškevičiene and Dagiene (2018)

Gouws et al. (2013b) carried out a study on an assessment of the computational skills of computer science students at the South African University. Their findings show that students with higher results in the computational thinking test did well on their class tests. Another study results by Ambrosio et al. (2014) suggest that spatial reasoning and general intelligence are important aspects for introductory programming courses and show high correlations between problem-solving, mathematics and students' academic success in a related field. Weese and Feldhausen (2017) studied the measurement of the role of self-efficacy in computational thinking and visual programming curricula for teaching computational thinking and reported that self-efficacy in computational thinking can be a good predictor of students' learning outcomes. Another decent study by Román-González, Pérez-González, & Jiménez-Fernández (2016) that investigated verbal, spatial, reasoning and numerical factors as well as speed and flexibility of children during mental logical operations. They concluded that computational thinking correlates with spatial ability, reasoning ability, and problem-solving ability. There is also a positive correlation between grade levels and computational thinking skills. The level of cognitive development and level of maturity are important factors in terms of the development of CT skills as well as in problem-solving skills (Román-González, Pérez-González, & Jiménez-Fernández, 2016). Rodrigues et al. (2016) study included 149 Brazilian high school students, where the research focus was the relation of computational thinking skills to students' problem-solving skills and school performance. They found a moderate correlation between the school performance of students and their computational thinking skills developed by computer programming activities. They also reported that students who are proficient in computer programming have higher performance in the National Exam of High School. Rodrigues et al. (2016) also stated that they could not identify significant effects of computational thinking in Humanities, Languages and Essay Writing disciplines; and studies needed to be carried out in these mentioned areas. Boom et al. (2018)

research seeks was how computational thinking is related to intelligence. As a measurement of computational thinking skills, twenty Bebras tasks and general problem-solving ability, a non-verbal intelligence test (TONI-3) were answered by Seventy-one pre-service teacher students. A high and significant correlation shows that these two concepts are highly related, which means as computational thinking abilities increased, intelligence tended to increase as well. A possible reason of this large relationship is due to these both concepts are naturally related, in other words, computational thinking is nothing more than just a part of general intelligence. Clarifying and identifying this fact might have an impact on how we see computational thinking and its future development. A study by Città et al. (2019) extensively examined the relationship between spatial reasoning and computational thinking skills of ninety-two students from grades 1-5 in five primary schools by using unplugged approaches to coding. The focus of the research was the role and action of spatial reasoning and the effect of mental rotation on computational thinking skills. Città et al. (2019) found a positive correlation between computational thinking skills and mental rotation ability and concluded that the higher the mental rotation ability the higher the chances of responding correctly in the coding test. They did not find any significant gender difference in the relationship between coding ability and mental rotation skills.

Several studies investigate the relationship between computational thinking and language skills. Nesiba, Pontelli and Staley (2015) explored the relationship between English literature in K-12 and computational thinking within a DISSECT (DIScover SciEnce through Computational Thinking) project, where high school students were involved in STEM courses. DISSECT blends computational thinking practices with composition and literature to provide school students with the ability to write critical and comparative analyses. Nesiba et al. (2015) suggest that computational thinking should be explicitly taught to schoolchildren as a problem-solving method within the context of all school courses. They claim that students in the DISSECT classes utilize computational thinking skills such as, algorithmic thinking through spatial reasoning and abstraction through data analysis, which enhance their reading and writing abilities and help them succeed in the disciplines. Likewise, Sabitzer, Demarle-Meusel and Jarnig (2018) introduced computational thinking in language classes. They taught the basics of computer science and computational thinking tools like modelling, which support language learning in various ways and help in text comprehension, acquiring and elaborating new vocabulary and visualising grammar rules. The results of questionnaires and interviews by Sabitzer et al. (2018) show that teachers and students accept modelling as a useful and practicable computational thinking tool in learning. They conclude that the integration of modelling, as a computational thinking tool in language lessons is possible without any obstacles in primary and secondary education. In general, there are more research studies related to computational thinking in STEM-related areas than other humanities areas (Nesiba et al., 2015; Román-González, Pérez-González, Moreno-León, & Robles, 2018). If there is any relationship between intelligence and computational thinking, and computational thinking is a part of general intelligence (Boom et al., 2018), then it is worth investigating either the direct or indirect relationship of computational thinking skills to other domain areas and their transferability.

2.5 Assessment in education

This section broadly discusses assessment in general, and then focuses on selected-response items, specifically the multiple-choice questions along with their advantages and limitations. Recommendations on constructing good multiple-choice questions are followed and the Diagnostic Questions platform, which is used in data collection in this research study, is introduced as well. The Diagnostic Questions platform was used during the data collection in this study. Bespoke multiple-choice questions were constructed in line with the recommendations taking into consideration the Kazakhstani curriculum and the topics covered in informatics textbooks.

Learning is a basic, adaptive and natural feature of all people that lasts lifelong. Our mind enables us to be flexible and become active learners by gaining knowledge and skills. In education, the purpose of learning is to enable learners to be prepared for their future life and help them to reach their full potential as lifelong learners. In a classroom environment, the increasing number of taught courses requires several regulations and monitoring. The value of the knowledge taught in classes is tested for its applicability to new situations, not only in a classroom environment but also outside the school. The modern views of how effective learning proceeds practice focusing on learners' understanding and application of acquired knowledge and learnt skills. Studies emphasize the importance of the following questions in a learning environment: What is taught?, How is it taught?, and How is it assessed? (Bransford et al., 2000). The American Educational Research Association, a research society founded in 1916 that promotes the use of research to improve education and serve the public good by practical application of research results, defines assessment as any systematic procedure for collecting information that can be used to make inferences about the characteristics of people or objects (AERA 1999). In other words, it is a process where students' comprehension, knowledge, skills and abilities are measured (Peariso, 2006). Reynolds et al. (2010) define measurement as "a set of rules for assigning numbers to represent objects, traits, attributes, or behaviours" (p 3). In education a test is used as a measuring tool, therefore; there must be specific rules to represent the test takers' performances. The performances are represented in terms of numbers and these numbers can be interpreted as reflecting characteristics of individuals that taking the test. In this study, the correct number of answers to a given test might be interpreted as reflecting students' computational thinking performances. A test is a procedure of obtaining, evaluating and scoring an individual's behavioural characteristics (AERA 1999). It is not assessment; it is a tool for assessment, which is much simpler than the versatile process of assessment. Assessment is an essential part of the teaching process, which helps to strengthen teaching and to facilitate learning (Reynolds et al., 2009). It plays a vital role in any education system not only by indicating the students' achievements and learning needs but also by informing decision making and policy (Scott, Scott, & Webber, 2016). A standardised assessment allows us to see if the achieved results of students meet the objectives set for them. Likewise, it also provides a way to compare the performances of students between different schools and organisations, which helps to share the best practices of pedagogy among schools. Therefore, school administrators can benefit from achievement data by measuring and judging how well the resources and materials are being utilised. The data obtained from the assessment outcome play role in the design of education policies for a better learning environment and to shape schools and set educators, teachers and students up for success. In the modern education system, there is a wide variety of methods and tools of assessment such as essays, constructed-response (or open-ended questions) tests, selected-response tests, project-based tests, oral presentations and performance-based tests. They aim and are designed to measure and document the academic readiness, learning and comprehension progress, skills and abilities, and educational needs of learners. Many different methods of assessment have been developed, each with its advantages and drawbacks in terms of reliability, validity, educational impact, costs and acceptance. They may be divided into the following types based on their purposes: placement assessment, high-stake assessment, pre-assessment, summative and formative assessment. All assessments, regardless of what is being measured, the achievement or aptitude, should be utilised as a prediction and give directions for future learning (Wood, 1987). They are required to identify the learning needs, evaluate teachers, improve instructions and have accountability for schools. Among these types of assessments, certain types are more frequently utilised by educational organisations, such as selected-response items or multiple-choice questions (Gierl, Bulut, Guo, & Zhang, 2017; Xu, Kauer, & Tupy, 2016). It is expected that any kind of assessment system should be designed to support the school curriculum, rather than to expect the school curriculum to fit the assessment system (Wiliam, 2014).

2.5.1 Selected response items

When an item requires a responder to choose a response from available options, it is called a selected-response item. True/false, matching items and multiple-choice questions are examples of selected-response items. If it requires a student to create or construct a response, it is classified as a construct-response item. Essays, short-answer items, project-based tests, oral presentations are examples of construct-response items. Construct-response items are mostly used to test knowledge, achievement and abilities. However, they are not as efficient and reliable as selected-response ones, thus, it is recommended to construct-response format questions to measure skills that are impossible to measure by selected-response items (Downing & Haladyna, 2006). The most frequent used selected-response format questions are multiple-choice questions (Downing & Haladyna, 2006; Reynolds et al., 2009) and they are widely used in colleges as assessment tools (Ebel & Frisbie, 1991; Gierl et al., 2017; Xu et al., 2016). Generally, any multiple-choice item has a stem, answer options, and some supporting information. The stem may consist of context, content, or the question for the learner. The answer options have a set of alternative answers with only one correct option and some other incorrect options. The incorrect options are also called distractors, as their function is to distract the test takers by offering plausible answers for the given question (Gierl et al., 2017). Supporting information may be present in the body of the stem or the answer options and may contain text, images, charts, graphs, tables, audio and video. When students are given multiple-choice questions, there is a stem and at least two (or more) answer options, which all may look quite plausible and very similar to each other. Although the number of options is completely up to the test maker and can exceed the number five, some recommendations and guidelines suggest using three options for the answer choices (Haladyna & Downing, 1993; Rodriguez, 2005). According to the given question type and format, students are expected to utilise their prior knowledge about a given topic and show their ability to solve problems by recognising the relationship between the given content in the stem and the right answer (Gierl et al., 2017). Although multiple-choice questions are the most popular ones in educational organisations, they have some shortcomings, which are presented in the next section.

2.5.2 Drawbacks of multiple-choice questions

Despite the objectivity, accuracy and effectiveness of multiple-choice questions there are several disadvantages as well. There have been many debates about the effectiveness and validity of multiple-choice items. The limitations of multiple-choice questions can be grouped as follows:

1. An ambiguous meaning: A stem or question might be poorly worded and have ambiguous meaning that will lead to misunderstanding or not understanding at all, thus, measurements will not be reliable. (Downing & Haladyna, 2006; Paxton, 2000) This issue can be seen mostly in small-scale tests, rather than large-scale tests prepared by professional testing agencies (Downing & Haladyna, 2006).

2. Test-wiseness: If test-takers choose the correct answer without knowing the correct answer it means they are using their test-wiseness skills (Chaoui, 2011). There is evidence that multiple-choice questions may be vulnerable to “test-wiseness” which diminishes a pure random-guessing factor (Rogers & Harley, 1999).

3. Guessing: The major criticism of multiple-choice type questions is that the random guessing assumption is present (Chan & Kennedy, 2002; Fenna, 2004; Zimmerman & Williams, 2003). This issue of guessing with multiple-choice testing is that it offers an opportunity for test-takers with no preparation and knowledge to guess the correct answer among answer choices. The presence of random guessing in multiple-choice tests prevents teachers and educators from identifying if students fully understand a particular topic or not (Paxton, 2000). There are some bits of advice on preparing better quality multiple-choice questions (Haladyna, Downing, & Rodriguez, 2002; Xu et al., 2016) particularly to minimise guessing that may involve the process of elimination by using partial knowledge, but it is not a very frequent issue (Ebel & Frisbie, 1991).

4. Backward approach: Since multiple-choice questions have answer choices with distractors and one correct answer, a backward approach can be used simply by substituting the given options with the unknown being asked to track the correct answer. This case usually applies to mathematical equations or calculations and some programming problems (Bridgeman, 1992; Chan & Kennedy, 2002; Hamilton, 1994; Simkin & Kuechler, 2005).

5. Lack of diversity: The limitation of covering higher-order thinking skills (Popham, 1995) makes it challenging to prepare multiple-choice format questions that can test and measure above the first three (remember, understand, apply) cognitive levels of Bloom's taxonomy (Hancock, 1994; Simkin & Kuechler, 2005). This narrowness of range (Simkin & Kuechler, 2005) and lack of precise indication of complex performance make multiple-choice questions less preferable (Martinez, 1999). They are applicable only for assessing cognitive knowledge (Simkin & Kuechler, 2005), but not the psychomotor skills such as writing or playing a musical instrument, productions skills such as human traits or creativity skills (Downing & Haladyna, 2006). They measure the knowledge and understanding of test-takers but fail for measuring how this knowledge is used when facing real-life problems (Gayef, Oner, & Telatar, 2014). They cannot precisely indicate complex performance and divergent production (Martinez, 1999). Furthermore, they are useless to measure very complex abilities that cover high-level complex skills and abilities such as written communication (Downing & Haladyna, 2006).

6. Prevention of critical thinking: Carey (1997) states that multiple-choice questions prevent thinking critically as test takers do not produce an answer, instead only select among answer choices already provided.

7. Lack of developing communicative competence: Multiple-choice tests are more likely not to produce any active effect on networking and communication. Paxton (2000) implies

that using multiple-choice tests does not promote the development of communicative competence, which requires interaction with people in a certain linguistic medium.

8. False knowledge: When test-takers do not know the correct answer or have no clue, they might start to look for a closer answer or familiar terms among distractors. If the feedback is delayed or not provided, this case might lead to false knowledge and students will consequently learn wrongly (Roediger & Marsh, 2005).
9. More vulnerable for cheating: Unlike the other formats of assessment, multiple-choice questions are more vulnerable to cheating, especially if it is computer-based and/or online (Xu et al., 2016).
10. Need special construction: It is difficult to prepare multiple-choice questions (Reynolds et al., 2009) and it takes plenty of time (Chaoui, 2011). To obtain valid and reliable high-quality multiple-choice tests, test makers must very accurately construct questions intentionally considering the cognition level for each stem or question (Haladyna et al., 2002; Simkin & Kuechler, 2005).

Heim and Watts (1967) found that the multiple-choice technique might provide a significantly easier task than does open-ended answering. This interpretation contrasts with that of Schuwirth et al. (1996) who argued that no one could presume that all multiple-choice questions are easier than open-ended ones. Students might prefer multiple-choice questions to open-ended or response constructed questions because they simply do better on multiple-choice questions. In general, students' learning should not be judged based on only multiple-choice question scores (Funk & Dickson, 2011). At the same time, the multiple-choice format has many advantages, which are discussed in the next section.

2.5.3 Advantages of multiple-choice questions

Despite some disadvantages listed in the previous section, multiple-choice items have some significant advantages, which is why they are the most frequently used assessment type educational testing. And well-constructed multiple-choice questions can be utilised to further understanding of students (Glass & Sinha, 2013). The advantages of multiple-choice questions different sources are summarised as follows:

1. Higher-order thinking can be measured (Downing & Haladyna, 2006; Haladyna & Steven, 1989; Woodford & Bancroft, 2005).
2. Large audience: using multiple-choice tests is an efficient way to collect data and evaluate them for large numbers of test-takers (Dufresne, Leonard, & Gerace, 2002).

3. Accuracy: the computerised multiple-choice tests provide more accurate scores as machines grade them not humans (Holder & Mills, 2001).

4. Economical: it is very convenient for teachers and instructors to use multiple-choice questions as they take less effort to evaluate and analyse (Roberts, 2006). Computer-adaptive tests provide flexibility in levels and can be adapted for an individual's level during the testing without spending time on special designing the tests (Shaftel & Shaftel, 2007).

5. Large projects: the multiple-choice questions are particularly useful in evaluating large-sized projects or large amounts of materials. They enable instructors and teachers to ask, not only a large number of questions but also cover a wide range of subject materials on a particular course (Becker & Johnston, 1999).

6. No penalty: there is no score losing for test takers' grammar and/or spelling mistakes, or low writing abilities (Zeidner, 1987).

7. Objectivity: it is perceived that multiple-choice tests are objective (Becker & Johnston, 1999; Haladyna & Steven, 1989; Zeidner, 1987) therefore they avoid teacher bias (Simkin & Kuechler, 2005).

8. Feedback: an immediate feedback assessment technique is effective and practical in an assessment that supports broad options by educators (Epstein & Brosvic, 2002; Merrel, Cirillo, Schwartz, & Webb, 2015).

9. Less anxiety: multiple-choice questions reduce student anxiety (Snow, 1993).

10. Prevent cheating: cheating can be almost completely avoided by randomising seats and making multiple versions of tests (Wesolowsky, 2000).

11. Opportunities of digital systems: utilizing computerised multiple-choice question banks, answer keys, and test generators it is easy to store, use, re-use and edit questions (Haladyna & Steven, 1989).

12. When selected-response questions are designed with high accuracy by professionals, they may measure the same aspects of the cognitive domains as they do with constructed-response (Rodriguez, 2003).

13. Test theory: Test theories, such as classic test theories and item response theories can be applied to multiple-choice questions (Haladyna & Steven, 1989).

According to Downing and Haladyna (2006), multiple-choice items are the most suitable format for the assessment of higher-order cognitive skills and abilities, such as problem-solving, synthesis, and evaluation. When multiple-choice questions are carefully constructed they can measure the first three levels of cognitive domains of Miller's Pyramid (Tarrant et al., 2006). Haynie's (1994) findings suggest that multiple-choice tests provide a better quality of student learning, and this advantage of multiple-choice questions over open-ended short-answer questions might be due to the correct answers provided among the distracters, thus supports retention. The key to increasing later retention is by repeating the retrievals (Karpicke & Roediger, 2007). As per Hattie and Gregory (2018), during consolidating surface learning in the learning process, the retrieval, that is getting the stored information from the memory, is more emphasized. Smith and Karpicke (2014) studied and experimented with multiple-choice tests, short-answer questions and hybrid tests. They state that practising retrieval with multiple-choice questions might significantly improve learning and later performance. As explained in ACT-R theory, human learning, memory and mind consist of complex procedures (Weber, 2012); and Anderson's study results show that the more a person practices the better he gets in performing the given mental tasks (J. Anderson, 1993). Smith and Karpicke (2014) claim that both short-answer and multiple-choice tests positively affect long-term learning and multiple-choice questions are likely to produce more effect on improving learning. As developmental psychology, cognitive psychology, and neuroscience, to name but three, have contributed vast numbers of research studies, details about learning and development have converged to form a more complete picture of how intellectual development occurs. There is a positive relationship between the level of experience in a certain area and the amount of structural change in human brain. The studies concluded that practice increases learning and it changes the physical structure of the brain (Bransford et al., 2000).

2.5.4 Recommendations to construct good multiple-choice questions

This section discusses advice on writing multiple-choice items from different studies. Haladyna and Steven (1989) have developed a taxonomy of 43 multiple-choice item-writing rules, which derived from an analysis of 46 authoritative textbooks and other sources in the educational measurement literature. Haladyna et al. (2002) revised the taxonomy of 31 multiple-choice item-writing guidelines and validated it based on two factors: results of a review of 27 textbooks on educational testing and the results of 27 studies and reviews published since 1990. Although this taxonomy mainly suits the classroom assessment, it can also be used for developing test items for large-scale assessments. Frey et al. (2005) in their research analysed twenty classroom assessment textbooks to find a consensus list of rules for writing multiple-choice items and developed rules that address four different validity concerns: poor wording, guessing, test-taking efficiency, and test-wiseness. Reynolds et al. (2010) list seventeen guidelines for developing multiple-choice items, which can be used flexibly. Moreno et al. (2015) has offered nine guidelines for preparing multiple-choice items, from the result of their study, which was based on validity criteria. Gierl et al. (2017) has focused on distractors rather than stem or correct answer aspects and listed several recommendations regarding the distractors while writing multiple-choice items. These researchers' and authors' main recommendations and guidelines regarding the content of multiple-choice items can be summarised as follows:

- Avoid the complex multiple-choice format (Frey et al., 2005; Haladyna et al., 2002; Haladyna & Steven, 1989; Reynolds et al., 2009).
- Avoid tricky items which might result in answering incorrectly (Haladyna et al., 2002; Haladyna & Steven, 1989).
- Use multiple-choice items and novel materials to measure higher-level thinking (Haladyna and Steven, 1989; Haladyna et al. 2002).
- Keep the central idea in the stem not on the choices (Haladyna et al., 2002; Moreno et al., 2015; Reynolds et al., 2009).
- Avoid states such as "all of the above" (Haladyna et al., 2002; Reynolds et al., 2009).
- Word the stem positively and avoid negative states such as "not", "none of the above" (Haladyna et al., 2002; Reynolds et al., 2009).
- There must be only one correct answer (Haladyna et al., 2002; Reynolds et al., 2009).
- Answer options should be plausible (Frey et al., 2005; Gierl et al., 2017; Haladyna et al., 2002; Reynolds et al., 2009).
- Answer options should be between three and five (Frey et al., 2005; Gierl et al., 2017; Haladyna et al., 2002; Reynolds et al., 2009).

Some recommendations on writing multiple-choice items are case-specific and unique. However, a certain pattern in the recommendations is commonly addressed. When writing and developing multiple-choice items for measuring computational thinking performance in this research study, these guidelines and recommendations were followed. The recommendations by the Diagnostic Questions (2016) platform used are in line with the recommendations listed above.

2.5.5 Project Quantum

Project Quantum is a crowd-sourced project with a bank of high-quality multiple-choice questions used for the assessment of computing in schools. developed jointly by Computing At School (Diagnostic Questions, 2020). The Centre for Evaluation and Monitoring (CEM), Cambridge Assessment and the Diagnostic Questions team. The project provides a free online assessment system developed by Computing At School (CAS), The Centre for Evaluation and Monitoring (CEM), Cambridge Assessment and the Diagnostic Questions' team. It supports teaching computing by providing opportunities to test students' understanding and help students with their progress in their learning journey. It is an assessment system, which can evaluate and give marks immediately with many quizzes on various topics. Quantum helps to evaluate students' progress and find misconceptions on topics. Quantum is a platform, which utilises the multiple-choice format that is assessable by computers and gathers data from students' responses, which is analysed to enhance the quality of the assessment system. One aspect of the project Quantum is openness to research. The anonymised data is available and accessible to researchers to work on the development and improvement of the assessment system of Quantum. Cambridge Assessment, Durham Centre for Evaluation and Monitoring (CEM), Computing at School (CAS) and Diagnostic Questions are the partners of the project Quantum (Oates, Coe, Jones, Scratcherd, & Woodhead, 2016). The CEM monitors the quality of the crowd-sourced question bank. It provides quantitative methods to produce evidence-based feedback on the effectiveness and quality of the assessment items. In this research study, the Diagnostic Questions platform is used for measuring computational thinking performance by fifty multiple-choice questions presented in five quizzes.

2.6 Conceptual framework

This research is based on the conceptual understanding of computational thinking and its concepts as introduced by Computing at School (CAS), a UK-based community that promotes computing in schools, which is sponsored and supported by BCS, Ensoft, Microsoft, Google, and the UK Committee of Heads and Professors of Computer Science. Application of computational thinking can include the following artefacts: systems, processes, objects, algorithms, problems, solutions, abstractions, and collections of data or information. The concepts of computational thinking described by CAS are logical reasoning, algorithmic thinking, decomposition, abstraction, generalisation, and evaluation. Similarly, computational thinking has been integrated into the updated informatics curriculum in Kazakhstan. The objectives of this curriculum are to teach students to tackle a variety of tasks through analysis, abstraction, modelling, and programming; and to enable students to gain the abilities, such as thinking logically, thinking algorithmically, identifying patterns, and thinking in terms of generalisation, decomposition and evaluation (National Academy of Education, 2016). An operational definition: computational thinking is a thought process involved in a logical approach to problem-solving that is applied in different fields using algorithmic thinking, analysis, abstraction and generalisation, with or without using the power of computers. The assessment of a computational thinking performance by multiple-choice questions is a concept-based approach. This study investigates to what extent students' computational thinking performance is related to their school achievement, as illustrated in Figure 11. The study will help to see the link between students' achievement, perception of computational thinking skills and computational thinking performance. It seeks to identify the extent of the interaction between their computational thinking skills and their school achievement.

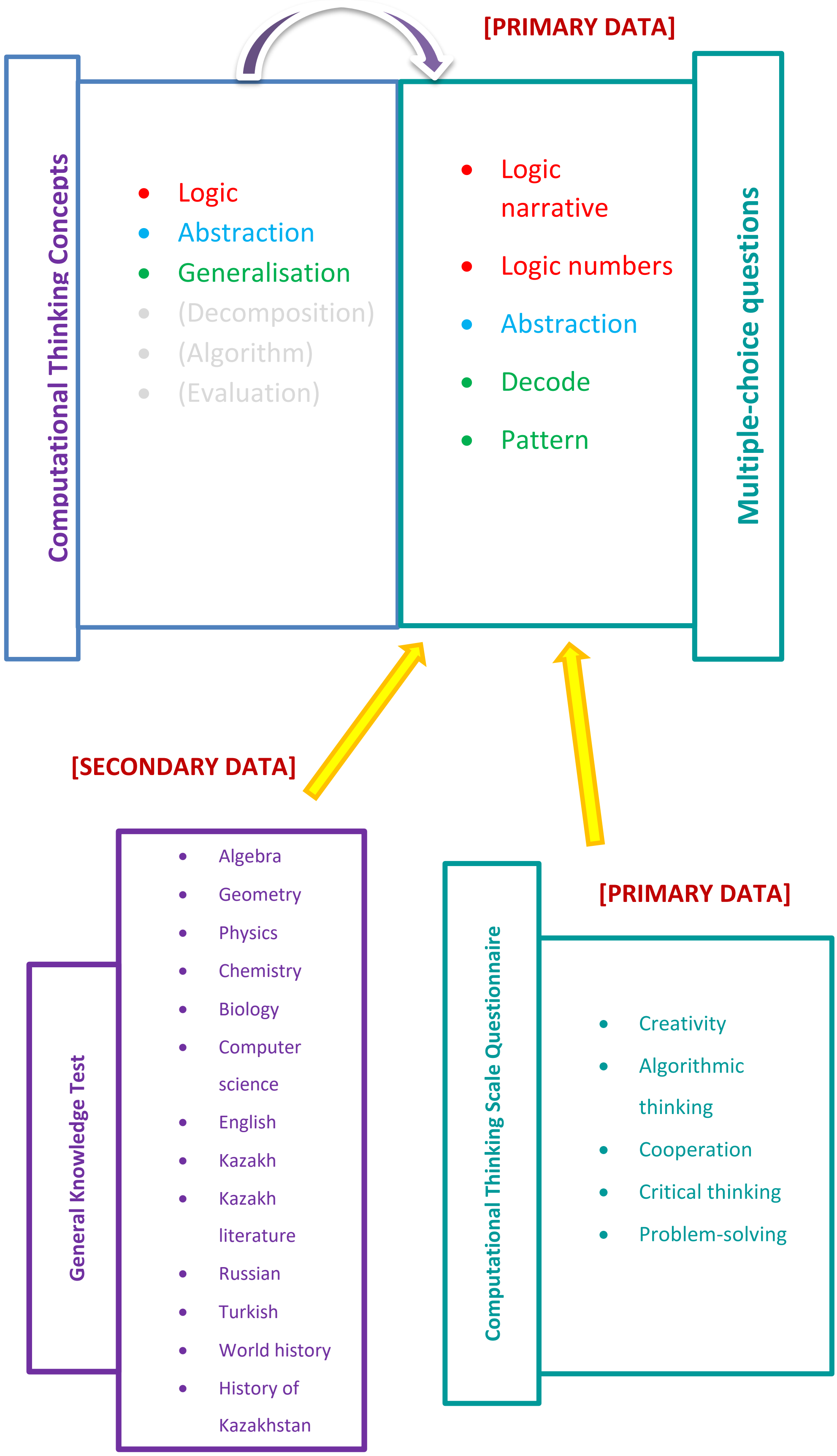


Figure 11 Framework: Computational thinking performance measured by multiple-choice questions and school achievement

2.7 Summary

Thinking is a mental process and Dewey, Bartlett and Baron each provided their versions of the definition of thinking. Athreya & Mouza (2017) define it as a high-order cognitive function used in the process of making choices and judgments. The thinking process can be separated into lower-order and higher-order sub-processes. Higher-order thinking is related to problem-solving, critical thinking, creative thinking, and decision-making. As thinking elements, the function of the human brain is studied by scientists. Although there is still much unknown about the brain and its function, there are several assumptions. Moreover, there is a controversial claim that lower-order cognitive skills are good for factual knowledge and the higher-order learning activities mostly enhances the higher-order learning skills (Willingham, 2012). The human brain has an active state when a person focuses on something and a resting state when not focused; they equate to a focused mode and a diffuse mode, respectively. Five unique ways of thinking are introduced by Ham (2018) sensory thinking, instinctive thinking, spatial thinking, computational thinking, and language thinking. Computational thinking is not a novel idea but a very popular trend in education. Back in the 1960s, Alan Perlis supported the idea that programming should be taught for all disciplines. Later, Papert introduced the term computational ideas and was a firm believer in the usefulness of programming in education, emphasizing the power of procedural thinking. Algorithmic thinking, logical reasoning, decomposition and abstraction are if taken separately, individually established and have a long history. However, when combined to construct computational thinking, it is not yet agreed on what to expect from this set of skills. Besides, there is a lack of supporting evidence on the transfer of these skills in solving real-life problems. The multi-definition of computational thinking, the lack of evidence supporting the transfer of computational thinking skills to other fields and the complexity of measuring computational thinking level bring difficulty in description and teaching (Denning, 2017; Mark Guzdial, 2015). Even though there is no single accepted definition of computational thinking, some of the main strands of computational thinking are commonly defined. The lack of a single accepted definition and its clarity is the biggest obstacle in promoting the idea of computational thinking among educators and the community (Grover & Pea, 2013a). Computational thinking is also described as critical thinking and problem-solving with an added integral computational component (Ater-Kranov, Bryant, Orr, Wallace, & Zhang, 2010). Studies are stating that computational thinking can be used by everyone to some extent. Some research findings suggest several ways to teach and develop computational thinking skills (Brennan & Resnick, 2012; Hoskey & Zhang, 2017; Selby, 2014). However, the idea of computational thinking and ways of delivery and teaching is still in the early stage of development.

Instead of debating whether computational thinking is an innate ability of a human being or not, it can be concluded that computational thinking as a mental activity can be found in a so-called "semi-resident" state, which is activated only if it is trained and developed, if not it is not there. This means everyone can think computationally to some extent but can be effectively used only when it is taught and developed. In the focus of this research study, it is important to identify what aspects of computational thinking is taught to students at each grade at school and what level of computational thinking skills can be expected. As mentioned previously in Chapter 2, generally 8th-grade students are expected to be familiar with algorithms, generalisation, logical thinking, and abstraction (Chuang et al., 2015). Having analysed the updated curriculum (National Academy of Education, 2016), annual plan and the new informatics textbook (Shaniyev et al., 2017), the following computational thinking concepts have been identified: logic, algorithms, generalisation and abstraction. For this reason, the study narrowed its focus to these concepts.

Many studies on computational thinking have a small sample size between 7-30 participants, which cannot be generalised or self-reported studies that are not validated and do not measure performance. The self-reported evaluations are found to be more prone to biases (R. Anderson, Thier, & Pitts, 2017). Multiple-choice items are considered the most suitable format for the assessment of higher-order cognitive skills and abilities, such as problem-solving, synthesis, and evaluation (Downing & Haladyna, 2006). Problem-solving skills in one context may not transfer to another without a deeper sense of what to use and when to use, knowing the underlying principles to be applied in a new context (Bransford et al., 2000). Several factors are affecting the assessment of transfer and its complexity, such as learner characteristics, learning tasks and transfer contexts (De Corte, 2003). Also, the transfer is more likely seen on experts rather than beginners or novices in a particular area (Willingham, 2012). There are still disputes on the transferability of computational thinking skills to other domains and areas. Whether it is proven with evidence or not that the computational thinking skills are transferrable, it is certain that the focus on teaching should be on skills like computational thinking rather than exclusively on hard coding (Naveh, 2006). Easy-to-use tools and instruments to measure the computational thinking skills of students should be developed to support the teachers and educators (Moreno-leon et al., 2018). There is a need for large-scale research studies (Kallia, 2017) that reveals the relationship between computational thinking and other disciplines (Rodrigues et al., 2016). The bespoke multiple-choice test was constructed specifically for 8th-grade BIL students in Kazakhstan to measure their computational thinking performance, and the Quantum project (DiagnosticQuestions.com) is used as a platform in this study.

Chapter 3 Methodology

3.1 Purpose and paradigm of research

The lack of large-scale studies that address the question 'how computational thinking skills are related to other disciplines and academic achievement' shapes the design of this research study. This research aims to examine the relationship between computational thinking perception/performance and the school achievement of secondary school students in Kazakhstan. In this research, variables such as language instruction, gender, questionnaire responses, scores from multiple-choice questions measuring computational thinking performance, responses to these multiple-choice items, and school achievement scores from fourteen subjects of 8th-grade students in Kazakhstan will be used to investigate the relationship between them. While measuring their computational thinking performance using an online test that consists of multiple-choice items, this study investigates the validity and reliability of these items. It aims to add some empirical evidence to establish the validity and reliability of multiple-choice questions designed to measure computational thinking performance and the statistical results of interrelationship or correlations between computational thinking performance and school achievement of secondary school students in Kazakhstan. Quantitative research is more suitable for large-scale studies (Dawson, 2009) that is more generalisable than qualitative research, which studies in-depth that requires a much longer time for data collection and analysis. Generally, this type of quantitative research is associated with the relationship with the evidence obtained from the measurement of the variables, and focuses on the nature of causality (Brundrett & Rhoades, 2013) and conducted outside, thus objectivity is maintained. This research follows a positivist paradigm as it is based on certain assumptions, which fits the aim of the study and matches the characteristics set by Creswell (2014). According to the positivist paradigm, by observing the natural phenomena, the collected data and evidence shape the knowledge and can be expressed in terms of generalisation. Therefore, the positivist paradigm will help with examining the relationship between computational thinking performance and perception; and other variables such as school achievement, instruction language and gender.

3.2 Research Design

Research design can be of different types, fixed and flexible (Robson, 2002) with different ways of approaches in social research studies, such as quantitative and qualitative social research (Robson & McCartan, 2016). In qualitative design, the questions such as “how” and “why” can be pursued and reveal how students improve and why they struggle in developing computational thinking skills. However, it would be more difficult to use qualitative methods for numerical data as such the school achievement (General Knowledge Test results), admission scores, questionnaire and multiple-choice test results. For quantitative researchers, the world can be measured and expressed by numbers that represent certain phenomena. Quantitative research, as defined by Babbie (1998), is “the numerical representation and manipulation of observations to describe and explain the phenomena that those observations reflect” (p.366). Creswell (2012) states that quantitative research explains phenomena by collecting numerical data, which can be analysed by using statistical methods. Creswell (2014) also adds that, in quantitative research, objective theories are tested by examining the relationship among variables that can be measured by instruments, so that numbered data can be analysed using statistical methods. In quantitative research studies, researchers seek to answer the research questions in terms of describing a relationship between variables. Explaining these relationships between variables helps in identifying the degree to which two or more variables are related to each other (Punch & Oancea, 2014). In quantitative social research, reliability and validity are important, the studies are replicable, mostly statistical analysis is expected, generalisation of the findings from a sample to a wider population is sought, objectivity is maintained as well as the neutral position of the researcher (Robson & McCartan, 2016). As can be seen from the literature review in Chapter-2, most of the studies on computational thinking are either small-scale quantitative or descriptive qualitative with a small sample size between seven and 30 participants. There is a need for large-scale studies to generalise the findings (Czerkawski, 2018; Kallia, 2017). This research study’s focus is to find how one variable is influenced by other variables, and explain the phenomena by collecting numerical data that can be analysed and generalised. Researchers develop true statements about the research area they are interested in to explain the causal relationship among variables and turn them into hypotheses. Since this research investigates the relationship between computational thinking performance and perception with prior measures of general school achievements of secondary school students, the use of quantitative methods was considered more appropriate as it would be possible to measure the variables and test the relationship between those variables. Measurable variables and observable data need to be collected using research instruments to investigate the causality, prediction and generalisation of the findings. The correlational design is used to describe and measure the association or relationship between two or more variables or sets of scores (Creswell, 2012). This degree of relationship is expressed as a number that tells whether one can predict another or whether these two variables are related to each other (Creswell, 2012). However, it cannot answer questions like how, when and why the certain characteristics occurred; this design does not explain what caused a certain situation. To get answers to the research questions, the qualitative research approach cannot provide feasible solution features as mostly objectivity is less open to generalisation, deal with small scale in terms of participants; deal with non-numerical data and objectivity is not valued (Dolowits, Buckler, & Sweeney, 2008; Robson & McCartan, 2016). This study has a quantitative research design. The primary data and secondary data were used in this research. The researcher has taken a set of steps for collecting quantifiable information about computational thinking performance/perceptions and general achievement by 8th-grade secondary school students. For the primary data collection, the multiple-choice questions were designed to measure the computational thinking performance of secondary school students. In addition to multiple-choice questions, the computational thinking scale questionnaire (Korkmaz et al., 2015) was used to measure perceptions of computational thinking skills in the standardised online form. These two instruments provide numerical data that can be analysed. The online multiple-choice questions and the online questionnaire are used to get numerical data from the participants and to identify their level concerning the variables such as perception, performance, etc. The quantitative data is gathered by the online questionnaire and multiple-choice questions along with the secondary data, the Test of General Knowledge results are obtained from Bilim Innovations Lyceums. These data are analysed to see the relationship between the provided variables. The research seeks to identify if there exists any relationship and if exists, investigate these relationships and patterns to understand how these variables are interrelated. Thus, the multiple linear regression method is used to find the relationship between the dependent variable and independent variables.

3.3 Research Questions

As discussed in previous chapters, there is a need for large-scale studies that could objectively measure computational thinking and also show the relationship between computational thinking and other domains. The following research questions are developed to address the research problems. To measure the computational thinking performance of the students by multiple-choice questions a validated objective measurement is required, therefore it is addressed in the first research question.

1. How to measure the computational thinking performance of secondary school students by multiple-choice questions?

2. The following research questions (2a, 2b and 2c) address the relationship between the computational thinking performance of secondary school students in Kazakhstan and their general school achievement.
    a. To what extent is there a relationship between the computational thinking performance as measured by the multiple-choice questions and the science-subjects achievement of the secondary school students in Kazakhstan as measured by the General Knowledge Test?
    b. To what extent is there a relationship between the computational thinking performance as measured by the multiple-choice questions and the language-subjects achievement of the secondary school students in Kazakhstan as measured by the General Knowledge Test?
    c. To what extent is there a relationship between the computational thinking performance as measured by the multiple-choice questions and the humanities-subjects achievement of the secondary school students in Kazakhstan as measured by the General Knowledge Test?

The relationship between the perception of computational thinking skills and computational thinking performance is addressed in the third research question.

3. Is the perception of the computational thinking skills a predictor of the computational thinking performance of the secondary school students?

3.4 Instruments

The instruments used for the primary data collection are a test consisting of multiple-choice questions and a computational thinking scale questionnaire.

3.4.1 Multiple-choice questions

The first instrument is a multiple-choice test for measuring the computational thinking performances of the participants. The multiple-choice questions have been developed in line with the recommendations provided by various sources, which were discussed in Chapter-2. The test questions were assessed by two reviewers with experience in assessing computational thinking, later were reviewed by seven informatics teachers. This multiple-choice test consists of 5 quizzes, each quiz consists of 10 items, for a total of 50 questions. The concepts of computational thinking included in this test are abstraction, generalisation and pattern, and logic with 10, 20 and 20 questions, respectively. These test questions are designed to measure the computational thinking performance of 8th-grade secondary students in Kazakhstan taking into consideration the national curriculum, annual plans for computer science lessons and Informatics textbooks (Shaniyev et al. 2017) at Bilim Innovation Lyceums and students'. The validity and reliability of these multiple-choice questions are important as stated in research question 2. Each item in this multiple-choice test has four response options, with one correct answer and three distractors. This multiple-choice test is conducted online with a duration of 100 minutes.

3.4.2 Developing multiple-choice questions

The recommendations for constructing good multiple-choice items discussed in Chapter-2 page 64 have been taken into consideration during the writing of the multiple-choice items. To have good quality multiple-choice questions, it is recommended that the multiple-choice should have plausible distractors, not just questions or stems with four simple answer options. As a main instrument for the data collection, the multiple-choice questions for measuring the computational thinking of students were constructed according to the listed recommendations discussed in Chapter-2. As an online test-taking platform DiagnosticQuestions.com was used. The recommendations provided by the Diagnostic Questions for writing multiple-choice items are in line with those recommendations from the literature discussed in Chapter-2. Each item on Diagnostic Questions is marked with corresponding ID numbers of computing taxonomy and level of Bloom's taxonomy. Computing taxonomy consists of Computer Science, Information Technology, and Digital Literacy. Computational thinking taxonomy is under Computer Science and presented as shown in the hierarchy chart in Appendix F. As this research study mainly focuses on logic, abstraction and generalisation, other concepts such as algorithm, decomposition and evaluation are not presented extensively. 8^{th}-grade students have are already familiar with Boolean and logical operators (Shaniyev et al., 2017, p54), multiple-choice questions with narrative logic statements and illustrations were constructed. The chart in Appendix F also demonstrates that under Logic (1686) truth tables and logic statements can be used. Graphs, charts, diagrams and relationship illustrations were used to measure abstraction. Likewise, sequences, spotting commonalities and identifying patterns were used for the generalisation part.

3.4.3 A good multiple-choice question

Here are examples of good multiple-choice questions, which were constructed purposely for the target age group (Chuang et al., 2015; Kalelioglu, Gulbahar, & Kukul, 2016) and taking into consideration the topics covered in the curriculum. The questions followed the advice for writing multiple-choice questions by Diagnostic Questions and the recommendations for constructing good multiple-choice questions listed in Chapter 2. As explained in each given question, each answer option is based on misconceptions and common mistakes, thus they are plausible; questions are in the non-complex format (Frey et al., 2005; Gierl et al., 2017; Haladyna et al., 2002; Reynolds et al., 2009) and the main idea is in the stem (Haladyna et al., 2002; Moreno et al., 2015; Reynolds et al., 2009). All the fifty questions in the test are designed to meet the following criteria of a good multiple-choice question set by Diagnostic Questions: questions can be answered in a reasonable time between 30 seconds to 3 minutes, avoid multi-step solutions as it causes difficulty to identify where mistakes are made, can be learned from each incorrect answer and it is still impossible to see the correct answer by simply having the key misconceptions.

Example 1.

Figure 12 demonstrates an example of a multiple-choice question on abstraction.

Figure 12 A multiple-choice question on abstraction

Computing Taxonomy Tag: 1688

Computational Thinking: Abstraction

Bloom's Revised Taxonomy: Apply.

This question corresponds to the Apply level at Bloom's revised taxonomy. Students are tested whether they can use the given information and interpret it correctly and then demonstrate, link, operate, visualize to answer the question correctly as discussed in Chapter 2.

Students should be able to tackle the given problem shown in figure 6 by visualisation and sequences. First, they need to build the cogs according to the given criteria, and then identify the correct answer. When D turns clockwise, A must turn anticlockwise since D is opposite to A. Since C turns the same way as A, the only right answer is option **A:** "C turns anticlockwise and B turns opposite". To solve this problem, higher-order thinking is required. Students are not asked to remember some factual knowledge, but to use the given information and apply some mental skills, components of computational thinking to get to the correct answer.

Example 2.

Figure 13 demonstrates an example of a multiple-choice question on logic.

Computing Taxonomy Tag: 1686

Computational Thinking: Logic

Bloom's Revised Taxonomy: Analyse.

This question corresponds to the Analyse level at Bloom's revised taxonomy. Students are tested on their ability to analyse whether they can identify people according to given criteria, can compare, contrast and differentiate people within a family tree as expected according to the Analyse level as discussed in Chapter 2.

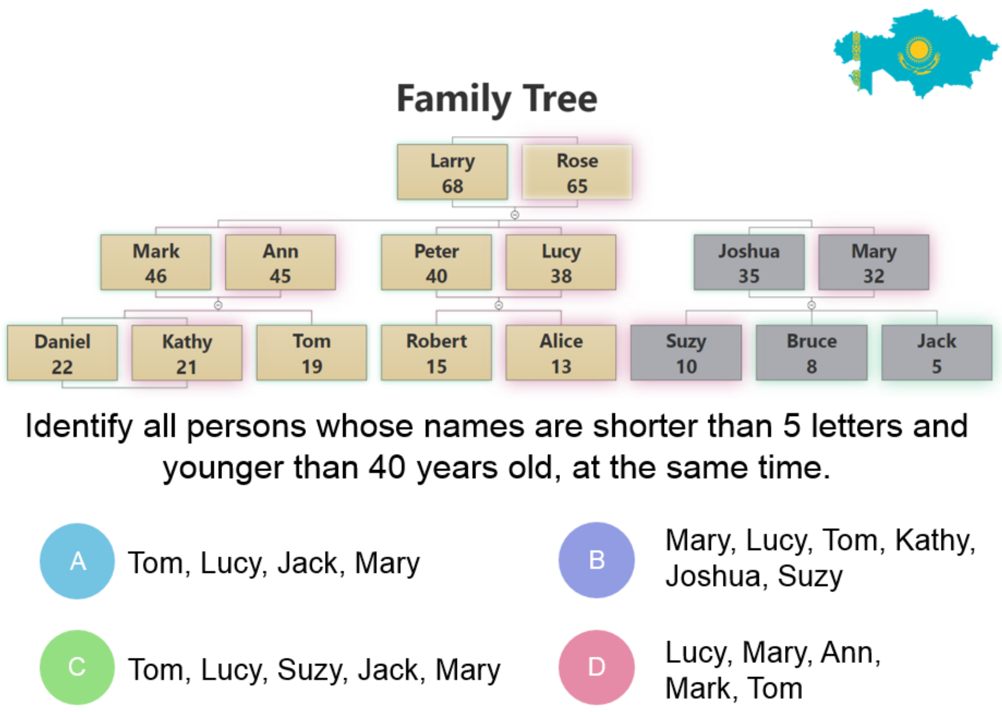


Figure 13 A multiple-choice question on logic

One important recommendation when constructing good multiple-choice items is to make plausible distractors so that each of the answer options reflects a different level of understanding. The answer options for the family tree example in Figure 13 are explained in detail as follows:

A: Although all four persons, Tom, Lucy, Jack, and Mary fit the criteria, there is still one person left, Suzy (her name is four letters long and she is ten years old), as the question asks to identify ALL persons who fit the given criteria (name shorter than five letters and younger than 40).

B: Mary, Lucy, Tom, Kathy, Joshua, and Suzy are all younger than 40 but Kathy and Joshua's names are longer than five letters; and Jack, who is younger than 40 and whose name is shorter than five letters, is missing in this answer option.

C: Correct answer. Tom, Lucy, Suzy, Jack, and Mary are all younger than 40 and all names are shorter than five letters. These are all persons in the given family tree and they all fit the given criteria.

D: Lucy, Mary, Ann, Mark, and Tom all have names that are shorter than five letters, but Mark and Ann are older than 40; Jack and Suzy are missing, whose names are shorter than five letters and younger than 40 years old, thus, this answer option does not fit the given criteria.

3.4.4 Computational thinking Scale Questionnaire

The second instrument used in this study is the Computational Thinking Scale questionnaire which was originally developed by Korkmaz et al. (2017) in 2015 to measure the levels of computational thinking skills of university students, covering five factors with a total of 29 items. Later, the Computational Thinking Scale questionnaire has been adapted to the secondary school level (Korkmaz et al., 2015) where the number of items was reduced to 22 with the same number of factors. These five factors in this computational thinking scale questionnaire are “Creativity” with 4 items, “Algorithmic thinking” with 4 items, “Cooperation” with 4 items, “Critical thinking” with 4 items and “Problem-solving” with 6 items. For the validity and reliability of the original scale the exploratory factor analysis, confirmatory factor analysis, item distinctiveness analyses, internal consistency coefficients and constancy analyses were conducted by the author and concluded that the CTS scale is a valid and reliable measurement tool to measure the computational thinking skills of the students. The validity and reliability of the CTS questionnaire in this study are discussed in more detail in the next section. The Computational Thinking Scale questionnaire was originally constructed in Turkish language, then to fit the Kazakhstani participants it was back-translated into English, Kazakh and Russian as the participant students in the research study are fluent in one (or more) of these three languages. The CTS questionnaire then was reviewed by seven informatics teachers who speak all four languages. By answering the questionnaire participants self-reported their perception of their computational thinking skills. This questionnaire was conducted online on iSurvey with a maximum duration of 20 minutes.

3.4.5 Instrument validity and reliability

Validity and reliability are two crucial components of any research study regardless of whether it is quantitative or qualitative. Testing the validity of a certain instrument used in the research is to show that it is measuring actually what it is supposed to measure. Although perfect validity is not possible to reach, in quantitative research, through appropriate sampling, instruments and certain statistical analysis, data validity can be improved (Cohen, Manion, & Morrison, 2011). The validity of a research instrument is closely related to its reliability because an instrument must be reliable to be valid. The numerical measures obtained must be tested or verified by the scientific method (Creswell, 2014). A positivist researcher has to ensure that the study follows such principles of positivism as replicability, predictability, randomization in sampling and controllability (Cohen et al., 2011). Reliability in quantitative research means consistency and replicability over time, over instruments and participants of the study, with accuracy and precision (Cohen et al., 2011). Replicability means the research instrument gives similar results when it is used with a similar population of participants. Poor reliability is the indication of the lack of a valid instrument.

Simply, validity can be defined as measuring the correct construct, and reliability as measuring the construct correctly (Thissen & Wainer, 2001, p 11). In multiple-choice questions, ample numbers of high-quality items can provide high reliability (Haladyna & Steven, 1989). Validity consists of several factors including content validity, face validity and construct validity, and the overall validity of multiple-choice questions is established by determining each of these factors. Content validity checks whether the multiple-choice questions are relevant, appropriate and representative of the construct being measured (Cohen et al., 2011). Although there is no absolute objective way of establishing content validity, it can be established by reviewing content, which should be undertaken by experts in the domain being measured, who also have some expertise in tool development. The first instrument of this research study is the multiple-choice test with fifty questions constructed to measure the computational thinking performances of 8th-grade secondary school students in Kazakhstan. And these multiple-choice questions were carefully assessed by two experts with experience in assessing computational thinking, then the test was also reviewed by seven informatics teachers. The reviews and pilot studies help to establish the face validity of multiple-choice questions and should check the clarity of content and readability of the test (Haladyna & Rodriguez, 2013). Construct validity means the extent to which a research instrument measures a theoretical attribute (Cohen et al., 2011). The construct validity of multiple-choice questions can be established using item response analyses such as item difficulty and item discrimination analysis (Haladyna & Rodriguez, 2013; Violato, 1991). Validity and reliability are important factors where quantitative researchers try to demonstrate that methods they use in their research studies succeed in measuring what they aimed to measure (Dawson, 2013). For the multiple-choice questions of the test, the Item Response Theory models were used. Item Response Theory (IRT) is a theory of testing based on the relationship between test takers' performances on a question of the test and their levels of performance on an overall measure of the ability that question was constructed to measure. The quality of these multiple-choice items is determined by both the 2-parameter IRT model and the 3-parameter IRT model. The item difficulty and discrimination coefficients are found, the item characteristic curves for each item and test information functions for each quiz are obtained. The IRT findings are extensively discussed in Chapter 4. The IRT models' results aim to demonstrate that the test items fit well and this bespoke multiple-choice test is a valid and reliable tool to measure the computational thinking performance of secondary school students.

The second instrument is an adapted Computational Thinking Levels Scale questionnaire. The original Computational Thinking Scale (CTS) questionnaire was developed by Korkmaz et al. (2017) to measure the computational thinking skills of university students. This original CTS questionnaire consists of 29 items, measuring creativity, cooperation, algorithmic thinking, critical thinking and problem-solving.

The reliability analysis of the original Computational Thinking Scale questionnaire by Korkmaz et al. (2017) according to each computational thinking factor and total values have been calculated with the use of the Cronbach Alpha reliability coefficient, the correlation value between two equal-half, Spearman-Brown formula and Guttmann split-half reliability formula as shown in Table 4.

Table 4 The reliability analysis of the original CTS questionnaire by Korkmaz et al. (2017)

Factors	N	Two Congruent Halves Correlation	Spearman-Brown	Guttmann Split Half	Cronbach's Alpha
Creativity	8	0,713	0,832	0,832	0,843
Algorithmic Thinking	6	0,756	0,861	0,860	0,869
Cooperativity	4	0,835	0,910	0,908	0,865
Critical Thinking	5	0,562	0,719	0,687	0,784
Problem-Solving	6	0,406	0,578	0,578	0,727
Total	29	0,344	0,512	0,498	0,822

This Computational Thinking Scale questionnaire (Korkmaz et al. 2017), which was initially developed for determining the levels of computational thinking skills of university students has been adapted for secondary school students. The adapted Computational Thinking Scale questionnaire (Korkmaz et al., 2015) is a five-point Likert type scale that has 22 questions that cover the same five factors as in the original version: creativity, cooperation, algorithmic thinking, critical thinking and problem-solving; tested on 7^{th} and 8^{th}-grade students at a secondary school, the gender and grade distributions are shown in Table 5.

Table 5 Distribution in adaptation study according to gender and grades by Korkmaz et al. (2015)

Grade	Boys	Girls	Total
7	67	80	147
8	42	52	94
Total	109	132	241

Confirmatory factor analysis was performed and the regression values of the items were calculated, as a result, seven items with very low values were excluded from the original scale (Korkmaz et al., 2015). The validity and reliability of the adapted instrument were tested by exploratory factor analysis, confirmatory factor analysis, item distinctiveness analyses, internal consistency coefficients and constancy analyses. Korkmaz et al. (2015) reported that the reliabilities for the five aforementioned factors ranged from 0.640 and 0.867 with total factors of 0.809, as shown in Table 6 and concluded that the scale is a valid and reliable measurement tool, which can be used to measure the computational thinking skills of the secondary school students.

Table 6 The reliability test results of the adapted CTS questionnaire by Korkmaz et al. (2015)

Factors	N	Cronbach's Alpha
Creativity	4	.640
Algorithmic thinking	4	.762
Cooperation	4	.811
Critical thinking	4	.714
Problem-solving	6	.867
Total	22	.809

This adapted version of the Computational Thinking Scale questionnaire with 22 items was used to measure the perception of computational thinking skills of secondary school students in Kazakhstan in this research study.

3.5 The General Knowledge Test

By the end of each term, all subjects are graded by a final mark, except for some optional and non-graded subjects. But these final quarter marks cannot serve as a good indicator of school achievements of students because in Kazakhstan 1-to-5 scale grading system is used (OECD & World Bank, 2014). If we consider the fact that 1-2 is rarely used, the rest 3-4-5 do not provide a proper measurement of pupils' learnings and achievements. For that reason, there is a need for other reliable evidence that could serve as objective school achievements of students. Providentially, all Bilim Innovation Lyceums have a unique test, the General Knowledge Test, which is taken at the end of each term, on the same agreed day all around the country. It is based on the curriculum frameworks of the current quarter (or term). The General Knowledge Test is common to all Bilim Innovation Lyceums and it is designed to measure the students' knowledge, in other words, their understanding of the content according to the annual academic plan. It is a standardised test that measures how well a student has learned a specific body of knowledge and skills (Popham, 1995). The General Knowledge Test is a multiple-choice test that is more systematic and objective than officially given term marks, which use a 1-to-5 scale grading system. The General Knowledge Test results are used as secondary data in this research. The survey was used to provide explanations of what is described in the form of correlations, analysing the pattern of correlations to see where the relationships are strong and where they are weak or non-existent. The purpose is to gather substantial information based on structured questions that allow precision and comparability of responses. As we need quantifiable data to be collected from several schools, the online questionnaire and multiple-choice questions are appropriate instruments for this study (Creswell, 2012; Robson, 2002). The General Knowledge Test can also be considered as an external assessment, because the local teachers and tutors are not involved in the process of preparing, administering, or evaluating the test, thus they can serve as indicators that are more objective than the term marks given by the local school teachers. Students from 7th grade up to 10th grade are obliged to take these multiple-choice tests from the following subjects: algebra, geometry, physics, chemistry, biology, computer science, English language, Kazakh language, Kazakh literature, Russian language, Turkish language, world history, history of Kazakhstan and geography, which result in a score out of 160. The General Knowledge Test results are grouped under three subscales: science, language and humanity subjects. As mentioned in Chapter 1, the two main routes offered for students in upper secondary education in Kazakhstan are science and humanities. The classes in groups with science orientation are more focused on mathematics and science subjects, whereas classes in groups with humanitarian orientation are focused more on language and social sciences. It has been also discussed in Chapter 2 that computational thinking skills are most delivered by computer science, informatics and Science, Technology, Engineering, and Mathematics (STEM) subjects; and that STEM education plays important role in enhancing computational thinking skills for students. As discussed in Chapter 1, the classes in Kazakhstani schools follow two main directions: natural-mathematical that is more science oriented and social-humanitarian that is more humanities oriented. Therefore, the General Knowledge Test (GKT) results are grouped into three subscales: Science (SC), Language (LL) and Humanities (HUM).

3.6 Test-taking procedure

The General Knowledge Test is composed of multiple-choice questions. These multiple-test questions are not prepared by local teachers, they are prepared and designed by the Bilim Innovation Foundation, in line with the national curriculum and annual academic plans of each subject at local schools. Normally, the test is taken over two consecutive days, beginning from 10 am till midday. The following subjects are taken on the first day: Physics, Chemistry, Biology, English language, Kazakh language, Kazakh literature and Russian language. On the second day, students take tests in the following subjects: Algebra, Geometry, Computer Science, Turkish language, General history, History of Kazakhstan and Geography. All grade students (from 7 to 10) are obliged to take this test. Only the final year (11th graders) students have a different test, a preparation for Unified National Test, which is specific to 11th graders only. All students, including 11th graders, are randomly mixed and allocated to different classrooms. The list of students with their test classrooms are announced on the test day usually 30 minutes before the test and the lists of students' allocations are hung up at the schools' mainboards and outside of each classroom. Usually, for each classroom, there are two teachers are allocated to invigilate the tests. Also, on the test day all the students are mixed within the classroom, so that no classmates or parallel grade students could sit close to each other. Therefore, by shuffling all of the students, there are only 4-5 students from each grade in the same classroom. Considering that there are four variants of the test, the chance of cheating is down to a minimum. Students are not allowed to talk, use cell phones, smartphones, books or other materials (other than allowed materials, such as the periodic table, engineering calculators, etc.). Students are not allowed to leave the classroom during the first hour of the test. The duration of the General Knowledge Test is 120-150 minutes. After the test, the answer sheets are scanned locally. Then, the collected data are sent to the Bilim Innovation Foundation's test centre. Usually, the results are published and announced the next day. All these precautions before, during and after the test and well-organised invigilation are to avoid cheatings, maintain fairness and openness and provide a clear assessment.

3.7 Variables

Due to the type of research study, it is useful to tabulate the variables. As shown in figure 4 on page 68, the computational thinking performance is measured by logic narrative, logic numbers, abstraction, pattern numbers and decoding tests, based on three computational thinking concepts: logic, generalisation and abstraction.

Variables are numbered as follows:

1. **CTP** (Computational Thinking Performance) is a construct measured by 50 items of multiple-choice questions covering three concepts of computational thinking: Logic, generalisation, and abstraction. CTP is a dependent variable with three subscales: CTP_L=Logic subscale, CTP_G=Generalisation subscale and CTP_A=Abstraction subscale.
    - CTP_L: A 20-item subscale that measures logical reasoning.

   - CTP_G: A 20-item subscale that measures generalisation, the ability to see the similarities and connections, and identifying patterns.
   - CTP_A: A 10-item subscale that measures abstraction, the process of making a complex system easy to understand by keeping important details and removing the unnecessary ones.

   CTP_L, CTP_G and CTP_A subscales are scored a maximum score of 10 for each subscale, therefore the maximum score for **CTP** is 30.

   A sample copy of multiple-choice questions (Generalisation/Pattern recognition) can be found in Appendix F.

2. **CTS** (Computational Thinking Scale) is a construct measured by 22 items questionnaire covering five areas of computational thinking: Creativity, algorithmic thinking, cooperation, critical thinking, and problem-solving. CTS is an independent variable with five subscales: CTS_CR=Creativity, CTS_AT=Algorithmic thinking, CTS_CO=Cooperation, CTS_CT=Critical thinking, CTS_PS=Problem-solving.
   - CTS_CR: A 4-item subscale that measures creativity level by self-reporting.
   - CTS_AT: A 4-item subscale that measures algorithmic thinking level by self-reporting.
   - CTS_CO: A 4-item subscale that measures cooperation skills by self-reporting.
   - CTS_CT: A 4-item subscale that measures critical thinking skills by self-reporting.
   - CTS_PS: A 6-item subscale that measures problem-solving skills by self-reporting.
3. **GKT** (General Knowledge Test) is a score of the test that consists of 160 items on the following subjects: Physics, Chemistry, Biology, English language, Kazakh language, Kazakh literature, Russian language, Algebra, Geometry, Computer Science, Turkish language, General history, History of Kazakhstan and Geography. GKT is an independent variable with the following subscales: PHYS, CHEM, BIO, ENG, KAZ, KAZ_LIT, RUS, ALG, GEOM, CS, TUR, GH, KZH, GEOG.
   - PHYS: A 10-item subscale that measures physics performance
   - CHEM: A 10-item subscale that measures chemistry performance
   - BIO: A 10-item subscale that measures biology performance
   - ENG: A 20-item subscale that measures English language performance
   - KAZ: A 10-item subscale that measures Kazakh language performance
   - KAZ_LIT: A 10-item subscale that measures Kazakh literature performance
   - RUS: A 10-item subscale that measures Russian language performance
   - ALG: A 10-item subscale that measures algebra performance
   - GEOM: A 10-item subscale that measures geometry performance
   - CS: A 10-item subscale that measures computer science performance

- TUR: A 20-item subscale that measures Turkish language performance
- GH: A 10-item subscale that measures general history performance
- KZH: A 10-item subscale that measures the history of Kazakhstan performance
- GEOG: A 10-item subscale that measures geography performance

4. **SC** (Science) is the average of science subjects scores (Physics, Chemistry, Algebra, Geometry and Computer Science) is taken as a science variable.
5. **LL** (Language Level) is the average of language subjects scores (Kazakh, English, Russian and Turkish) is taken as a language level variable.
6. **HUM** (Humanities) is the average of humanities subjects scores (Kazakh Literature, general history, history of Kazakhstan and geography) is taken as a humanities variable
7. **ST** (School Types) are classified by gender: boys, girls and mixed schools
8. **G** (Gender) is an independent variable: male, female.

LI (Language of Instruction) is an independent variable that shows in which language (Kazakh or Russian) the lessons are conducted (except the science subjects, as all science subjects are conducted in English at Bilim Innovation Lyceums): Kazakh, Russian.

3.8 Data collection

Having described the population of the research study as described in Chapter 1, the process of sampling allowed to focus on the data collection. In this study, the purposive sampling method is used, as the Bilim Innovation Lyceums and their students were purposely selected as participants (Cohen et al., 2011). Bilim Innovation Lyceums are found suitable for data collection particularly on computational thinking skills, as the schools have a technology-enriched practical curriculum, highly qualified teachers, well-established education and monitoring system and practice of sharing the experience, knowledge and skills of teachers for 21st-century demands. Data collection procedures are illustrated as shown in Figure 14. Admission Test results (AT), which has initially been planned to be included in data collection and data analysis, has been omitted due to inconsistency of the obtained data.

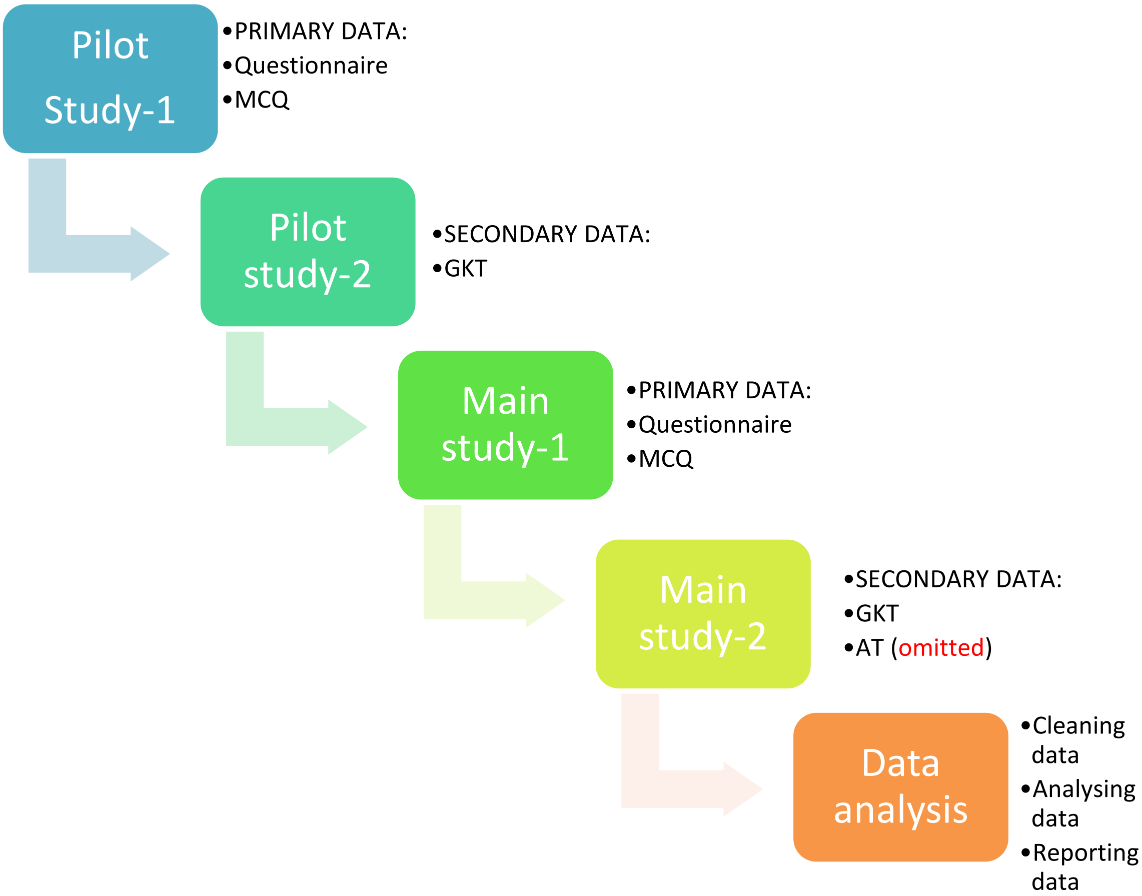


Figure 14 Data collection flowchart

3.8.1 Participants

The Bilim Innovations schools are considered the leading schools among high schools in Kazakhstan (Koshegulova & Mindetbay, 2020). There is a big number of applicants each year, approximately 10 applicants for one place at school but this number differs from region and schools. According to the report by Telyukanov (2021), the chief specialist at Bilim Innovation Foundation, the achievements from subject-specific Olympiads by the students of Bilim Innovation Lyceums, since their establishment in 1992, are as follows: 13110 regional medals, 2801 national medals (36.5 % of all participant school nationwide), 1950 international medals, and 169 world medals; regional 2463 medals, national 387 medals, and international 1920 medals in project competitions**.** At Bilim Innovation Lyceums five subjects, mathematics, physics, biology, chemistry and computer science are conducted in English. Bilim Innovation Lyceums' experiences are shared during the transformation period where three language education is being integrated. English speaking teachers at Bilim Innovation Lyceums run series of English language summer courses for 5th-grade students as a complementary supporting step towards three language education. There were a series of seminars and workshops for high school teachers organised by teachers of English language at Bilim Innovation Lyceums, not only in language education but also in technology usage. Robotics workshops and training sessions were organised by informatics teachers at Bilim Innovation Lyceums for other high school informatics teachers. Some textbooks for informatics, physics, biology and chemistry were prepared by using the educational experiences of Bilim Innovation Lyceums. As STEM education becomes more popular, integrated and easily available, students are more comfortable solving STEM problems regardless of explicitly mentioning any “thinking” methods under any umbrella, which are super popular as 21st-century skills. STEM education delivered without any foundational computing skills would be considered improper way. Learning STEM and solving STEM problems with the help of computer science would benefit to develop students’ computational thinking skills. As computational thinking is better delivered and taught in a technology-rich environment, whether it is stated specifically or how it is defined does not play a major role in delivering computational thinking to students. Once the essence of computational thinking is there in the classroom, rather than just formally in the curriculum, it ends up fostering the students' computational thinking skills. All of the above-mentioned facts demonstrate that Bilim Innovation Lyceums are suitable for data collection, as they already have a technology-enriched practical curriculum. There are several reasons for choosing 8th-grade students of Bilim Innovation Lyceums. The first year at Bilim Innovations Lyceums starts from 7th grade, as mentioned in Chapter-1. The first year at school is spent more as an adaptation period for students in a new classroom, new classmates, new environment with new school regulations; thus assessment at this point may not reflect their actual performance and achievements. Therefore, 7th-grade students are not the best sample for data collection. All 9th graders have state examinations by the end of the academic year and particular selected schools are tested by the External Assessment of Academic Achievement (OECD & World Bank, 2014), which means 9th-grade students have various tests and state examinations and school administrations do not usually allow any intervention studies, external assessments and data collections, while 8th graders have no such examinations and assessments that could serve as indicators and measurements of their achievement. Also, beginning in 2018, 8th-grade students in Kazakhstan started to take part in the International Computer and Information Literacy Study (ICILS) test, which includes measurement of computational thinking, in which Kazakhstan participated for the first time. Considering all the facts listed, 8th-grade students are better suited as participants for data collection.

3.8.2 Pilot study

A pilot study is one of the essential stages in any research study. It can be defined as a mini version of a full-scale study to test data collection instruments, sample participation and other methods and techniques planned for the main data collection. In other words, a pilot study is an important stage to check logistics procedures in researching to spot potential problem areas before starting the main data collection. It also helps researchers to get familiar with the data collection procedures of the study. Although a pilot study cannot be the guarantee successful data collection at the main phase, it is a crucial part of a good research design.

The pilot study was conducted with 9th-grade students as shown in Figure 14, where seven well established Bilim Innovation Lyceums were selected based on school types (boys/girls/mixed). The selected schools are four boys, two girls and one mixed school with a total of 309 9th-grade students. These schools are located in different regions of Kazakhstan that are, North, North East, Central and South. Among these 309 students, 211 participated in the pilot study. During the pilot study-1 phase, the primary data collection, which is the Computational Thinking Levels Scale questionnaire and the multiple-choice test to measure the computational thinking performance were taken. Those participants who did not successfully submit the online Computational Thinking Levels Scale questionnaire and/or just partially answered the multiple-choice test were omitted from the participants' list. As a result, 151 participants' fully responded data were gathered. At the pilot study-2 phase, General Knowledge Test results of the first and second terms were obtained from the databases of the schools as the secondary data.

Table 7 Computational Thinking Scale reliability test of the pilot study

Factors	Number of items	Score	Cronbach's alpha
CTS_CR (Creativity)	4	77.3	.724
CTS_AT (Algorithmic thinking)	4	72.0	.848
CTS_CO (Cooperation)	4	78.3	.853
CTS_CR (Critical thinking)	4	74.0	.763
CTS_PS (Problem-solving)	6	47.0	.786
CTS (Total)	22	69.7	.821

The Computational Thinking Scale is a five-point Likert type scale and consists of 22 items that could be collected under five factors. Each one of the items taking place in the factors has been scaled as never (1), rarely (2), sometimes (3), generally (4), always (5).

Testing the reliability of the CTS questionnaire. The result of the reliability test of the CTS questionnaire Cronbach's alpha for every five factors is greater than 0.7, creativity 0.724, algorithmic thinking 0.848, cooperation 0.853, critical thinking 0.763, problem-solving 0.786 and the total coefficient is 0.821, which shows that the instrument is reliable as shown in Table 7. Students' perception of computational thinking differs between 20 and 100, and the mean is 69.7. All factors' mean values are very close to each other except the problem-solving factor, which is 47.0, which might indicate that the students are less confident in their problem-solving skills.

The means for multiple-choice tests on logic, abstraction and generalisation range between the scores of 52 and 62 with the total mean of 57.6 as shown in Table 8.

Table 8 Descriptive Statistics CTP for the pilot study

	N	Min	Max	Mean	SD
CTP_L (1) (Narrative)	151	0	100	52.3	25.4
CTP_L (2) (Numbers)	151	10	100	61.6	24.9
CTP_A (Abstraction)	151	0	100	60.3	24.9
CTP_G (1) (Pattern numbers)	151	10	100	55.2	20.0
CTP_G (2) (Decode)	151	0	100	58.5	29.5
CTP	151	20	100	57.6	18.5

Among 151 participants, 114 (75.5%) are boys and 37 (24.5%) are girls as shown in table 8. The mean of computational thinking performance for boys is 56.5 and 61.3 for girls. SD for boys is 19.9, for girls 12.5. To see the difference between them in computational thinking performance (MCQ) the t-test is used. On average, there was no significant difference in CTP score between boys (M=56.5, SD=19.9) and girls (M=61.3, SD=12.5), $t(98.7)=-1.770$, $p=.08$ as shown in Table 9.

Table 9 Gender difference in CTP for the pilot study

Boys	Girls	Total
114 (75.5%)	37 (24.5%)	151
t-test significance level = 0.08 (>0.05).		

The Computational Thinking Scale (CTS) 0.776 is not a statistically significant predictor. But Science performance (SC=Algebra, Geometry, Physics, Chemistry and Informatics) 0.031 (<0.05), as shown in Table 10, and R square is 0.032, which suggest that the average score of science subjects (SC) is a moderate predictor of Computational Thinking Performance (CTP).

Table 10 Coefficients CTS and SC for dependent variable CTP for the pilot study

Coefficients		
		Sig.
	(Constant)	.001
	CTS	.776
	SC	.031
a. Dependent Variable: CTP		

The pilot study helped to identify several concerns, such as technical problems of online delivery of the questionnaire and the test, language barrier for participants and difficulty level of the test. During the pilot study period, two schools experienced problems with accessing computer classrooms. The number of computers was not enough for one class of students to complete the test and the questionnaire online at the same time. As a solution, mobile technology was used to access the tests. The questionnaire was presented in Kazakh, Russian, English and Turkish, participants could select any version that is more convenient for them. The multiple-choice questions were prepared in English only. There was a concern if the language level of the test suits the participants' English proficiency, though it was taken into consideration when writing questions to use simple English and avoid complex sentences. The feedback from participants and assisting teachers showed that there was no problem with understanding the questions. The most common feedback from informatics teachers who helped to conduct the study can be expressed as follows: ***"The questions seem easy and simple, however, they make you think"***, which confirms that the test suits students' level.

3.8.3 Main data collection

The main data collection period was between April and May 2018. Thirty-five Bilim Innovation Lyceums were invited for the research study. The questionnaires and tests were offered to over 2000 8th grade students across Kazakhstan. Local teachers helped in test-taking procedures organizing the classrooms with PCs and Internet access. The consent forms by the participants and their parents were administered by the school administrations. The local informatics teachers were instructed in advance, step-by-step with each test, the platform, the students' usernames and IDs, etc. There was a WhatsApp group for the informatics teachers, where all the questions regarding the procedures of this research study tests and questionnaire were discussed, answered and monitored at the same time. As for the main phases of the study, the same instruments that are used in the pilot study were used. The main study-1 phase of the data collection included the primary data collection, which are the Computational Thinking Levels Scale questionnaire, the self-reported scale as a perception of computational thinking and the multiple-choice questions to measure the computational thinking performance of the students. As the participation was voluntary, among over 1800 students, only 775 students' responses could be collected as primary data. At the main study-2, the General Knowledge Test results for all four terms were obtained from the schools' database. As planned initially, the Admission Test results were obtained, however, due to inconsistency of the obtained data the Admission Test results were omitted from the data analysis removed from the secondary data.

3.9 Ethical considerations

In educational research, ethics should be respected whether the research involves cognitive, diagnostic, aptitude or achievement assessments, or using secondary data (British Educational Research Association [BERA], 2018). Important ethical principles considered in this study are the informed consent of the participants before taking the questionnaire and test. They were informed about the purpose of the study, using secondary data (Admission Test and General Knowledge Test) and what the researcher will do with the results. Also, consent and permission from the parents of the participants are obtained. All the obtained primary and secondary data are kept confidential on a password-protected computer database at the University of Southampton. The study followed the University of Southampton's ethical protocols and guidance. The consent forms for participants and their parents, information sheets for participants and their parents were prepared. In addition to these forms, the main ethics form, risk assessment sheets, secondary data analysis form and the instruments for the data collection were submitted and successfully gained approval from the Ethics and Research Governance Online system, which is the online system designed to facilitate the process of gaining ethical, governance and insurance approval for research studies of the University of Southampton. To begin the pilot study, the researcher addressed some requests to the head of the computer science department at the Bilim Innovation Foundation. After the discussion with the headteacher, seven schools were chosen based on their types and locations. The biggest ethical issues were to get consent from school administrations. To begin with, school administrations were contacted and introduced to the research study. Then, computer science teachers at these schools were contacted to begin the pilot study. The research topic, the aim of the research, and the data collection procedures were explained to the local computer science teachers. The local computer science teachers delivered the participants' and parents' consent forms, information sheets, and testing information to the $8^{th}/9^{th}$-grade participants. In these forms, it was clearly stated that the admission test results and the general knowledge test results of participants would be used in this research as secondary data, along with the primary data: the online questionnaire and multiple-choice test results. In Appendix A and B, the researcher provided the ERGO approval and the copy of the participant information sheet, along with their right to participate or withdraw, using secondary data for the research, and the maintenance of confidentiality. These consent forms are important when conducting a research study. This study does not reveal the names of the participants and the schools to protect their identities, as research should respect the rights, needs, values, and desires of the participants (Creswell, 2014). Instead of names, there are ID numbers given to schools and participants. An important ethical issue considered in this research was the principle of voluntary participation. In Appendix C and D samples of the multiple-choice questions can be found.

3.10 Data analysis

For data analysis, collected data are cleaned and formatted as they initially had different formats. The data cleaning is done before the main data analysis, to make sure there is no missing data and that the data have a correct and appropriate format. Then, the data analysis is done by using the RStudio, an open-source integrated development environment (IDE) for statistical programming language R, and the SPSS Statistics software package program (RStudio, 2021). In R Studio the “ltm”, “mirt”, “TAM”, “mvtnorm” and “msm” packages are used for the Item Response Theory models, which are discussed in the next section.

2067 students were registered on the Diagnostic Questions platform from 35 secondary schools (BILs) across Kazakhstan. The number of students who voluntarily started the multiple-choice test is 1881, but not all completed, so this number later dropped due to missing data criteria. After omitting the blank data and missing lines from five quizzes (10 questions each) measuring the computational thinking performance on the Diagnostic Questions platform, the number of participants for each quiz was found as follows: Logic narrative - 905, Logic numbers - 876, Abstraction - 979, Decode - 876, and Pattern - 919.

These quizzes were grouped under three concepts such as logic, abstraction and generalisation, as they were previously constructed to measure these concepts of computational thinking. The logic category consists of the logic narrative and logic numbers quizzes, the abstraction category consists of the quiz abstraction, and the generalisation category consists of the decode and pattern quizzes. To keep the integrity and avoid relying on partial computational thinking concepts, only those participants' results who completed at least one quiz from each category were selected for further analysis. That is, the results of the participants are accepted if they completed the tests as one of the following combinations:

1. Logic narrative, logic numbers, abstraction, decode, and pattern.
2. Logic numbers, abstraction, decode, and pattern.
3. Logic narrative, abstraction, decode, and pattern.
4. Logic narrative, logic numbers, abstraction, and decode.
5. Logic narrative, logic numbers, abstraction, and pattern.
6. Logic narrative, abstraction, and decode.
7. Logic numbers, abstraction, and decode.

Those participants who completed the quizzes but did not cover all three areas were eliminated as they were not considered as enough evidence of measurement of computational thinking. As a result, the number of participants who completed all three category quizzes was 895. Then, the mean values of quizzes were calculated under each category correspondingly. From the school database that contains the secondary data, the General Knowledge Test results, 830 students were identified among 895 participants who completed all three category quizzes of the computational thinking performance test.

Next, the computational thinking scale questionnaire responses were collected from the iSurvey platform where 1018 participants fully answered. 63 participants who did not give their names and ID numbers and who were not found on the school database were eliminated, so the questionnaire was answered by 995 students from the schools addressed. The General Knowledge Test results were optimised for data analysis. The mean value of all General Knowledge Test results was calculated as follows: The missing data criteria (e.g. when a student did not participate in a particular test or day) were ignored and the mean values were calculated accordingly. Finally, the test and the questionnaire responses were merged with the secondary data table (General Knowledge Test results), and as a result, the number of participants who completed the computational thinking test, the computational thinking scale questionnaire and the General Knowledge Test was reduced to 775. Then, these data are summarized in tables using the statistical program SPSS. Descriptive statistics, the means, standard deviations, Cronbach's alpha, t-test, and Regression analysis are obtained. Descriptive statistics give us a general picture of the data collected. Cronbach's alpha is used to test the reliability of the instrument, the questionnaire (CTS) as a whole and each factor, subscales (CTS_CR, CTS_AT, CTS_CO, CTS_CT, and CTS_PS).

Item Response Theory is used for the validation of the instrument that measures the computational thinking performance of students, which is discussed more extensively in the next section. The measurement properties of the items (multiple-choice questions) and the participants, which are estimated for the scales based on IRT models are as follows: item difficulty coefficients, item discriminations coefficients for each item, guessing parameter (for 3PL models) and item characteristic curves, test information functions for the quizzes (logic narrative, logic numbers, abstraction, decode and pattern). In this study, both correlation and regression analyses are used. Shortly, correlation analysis is mainly used to quickly summarize the direction and strength of the relationships between two or more numeric variables; and regression analysis is primarily used to get models and equations for prediction purposes (Field, 2009). None of these relationships can be interpreted as describing the causal effect. In this study, correlation and regression analysis are used to see the relationship between the variables, such as CTP, CTS, GKT, SC, LL, HUM, G, ST and LI. T-tests and one-way ANOVA are applied to see if there is any difference between the groups. The Admission Test results, which consist of mathematics and logical test results of sample groups when the participants of this study were in 6th grade that is 2 years before the data collection time, were expected to be included as secondary data along with the General Knowledge Test results. However, there were too much missing (over 60%) and incomplete data such as ID number, name and surname. As a result, due to the inconsistency of the obtained data, the Admission Test results were not included in this study.

3.10.1 Item Response Theory

Item Response Theory (IRT) is a set of mathematical models for the design analysis and scoring of test instruments that measure attitudes, abilities and other variables (Magno, 2009). IRT as a theory has its origins from back in the 1920s by Louis Thurstone's work continued by Lawley, Mosier, and Richardson studies in the 1940s, and further by Alan Birnbaum, Frederic Lord, and George Rasch in the 1950s (van der Linden, 2015). Sometimes, the IRT is also called the strong true score theory or modern mental test theory. The IRT approach considers the chance of answering the given items correctly or incorrectly, where each item has its item characteristic curve that describes the probability of answering each item correctly or incorrectly given the ability of the test takers (Denmars, 2010). IRT can offer solutions to various measurement challenges that need to be addressed when constructing a test, therefore it is one of the most important psychometric methods of validating scales (Lord, 1980). IRT presumes a single underlying “ability” and uses logistic regression to describe the tendency of correct answers to items, where the estimated item measure and the test taker’s abilities are independent of each other. For example, a student with a high level of computational thinking skills will get a high probability of responding correctly to the items and score higher, where a student with a lower level of computational thinking skills will be more likely to score lower. In IRT a test taker’s ability theta, the item discrimination parameter ***a***, the item difficulty parameter ***b*** and the guessing parameter are denoted by (***ϴ***), a, b and ***c*** respectively.

The probability of a correct response is determined by the item’s difficulty and the subject’s ability. This probability can be illustrated by the curve in Figure 1, which is called the item characteristic curve (ICC) in the field of IRT. Item Characteristic Curves are graphical representations of the relationship between the measurement properties of the test taker and of the items (multiple-choice questions) that makes it easy to visually inspect and interpret the items in the scale. From this curve, you can observe that the probability is a monotonically increasing function of ability. This means that as the subject’s ability increases, the probability of a correct response increases; this is what you would expect in practice. As the name suggests, the item difficulty parameter is used to show how difficult it is to get a 0.5 probability of a correct answer for the item given the test taker’s level of the latent variable (ability). This means that the more difficult it is for a test taker to get a 50% chance of correct response to the question, the higher the ability level needed to achieve it. (2PL) The discrimination parameter is a measure of the differential capability of an item. A high discrimination parameter value suggests an item that has a high ability to differentiate subjects. In practice, a high discrimination parameter value means that the probability of a correct response increases more rapidly as the ability (latent trait) increases. The item discrimination parameter allows for determining how well the item identify test-takers at different levels of the latent trait. The higher the item discrimination the better differentiates between test-takers. The item discrimination parameter is simply the slope of the model shown in Item Characteristic Curve figures. when the item discrimination is high the response function can be seen by a steeper and sharper curve in Item Characteristic Curve figures. When the item discrimination is low, the Item Characteristic Curve is flatter, spreading through the scale, which means that the item can be answered by more participants along the scale continuous. Below, there are presented three examples of items with different item discrimination and difficulty parameters in Figure 15, Figure 16 and Figure 17.

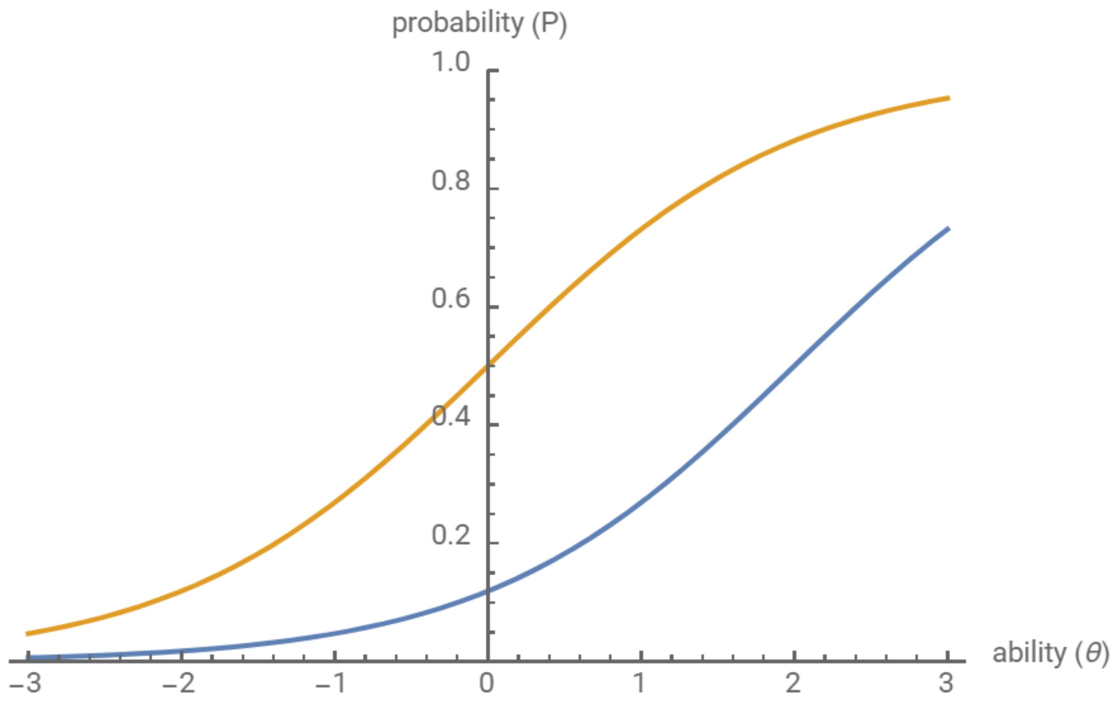


Figure 15 Two item characteristic curves with the same item discrimination and guessing but different levels of item difficulty.

Blue curve: Discrimination (a) = 1, difficulty (b)=2, guessing (c)=0

Yellow curve: Discrimination (a) = 1, difficulty (b)=0, guessing (c)=0

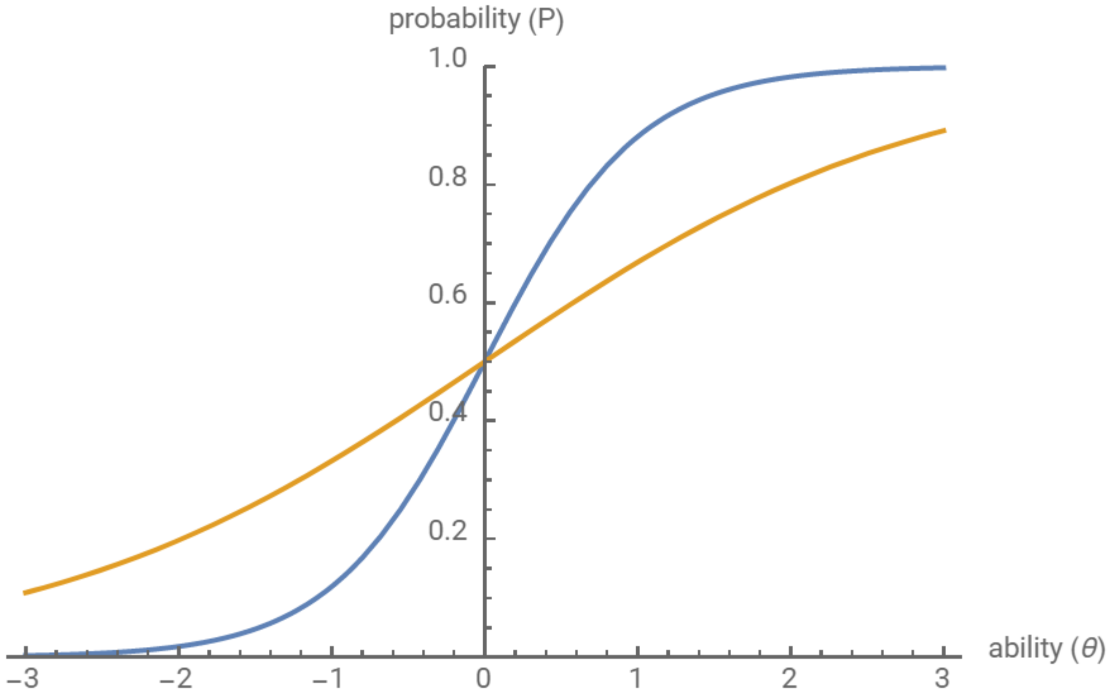


Figure 16 Two item characteristic curves with the same item difficulty but different levels of item discrimination

Blue curve: Discrimination (a) = 2, difficulty (b)=0, guessing (c)=0

Yellow curve: Discrimination (a) = 0.7, difficulty (b)=0, guessing (c)=0

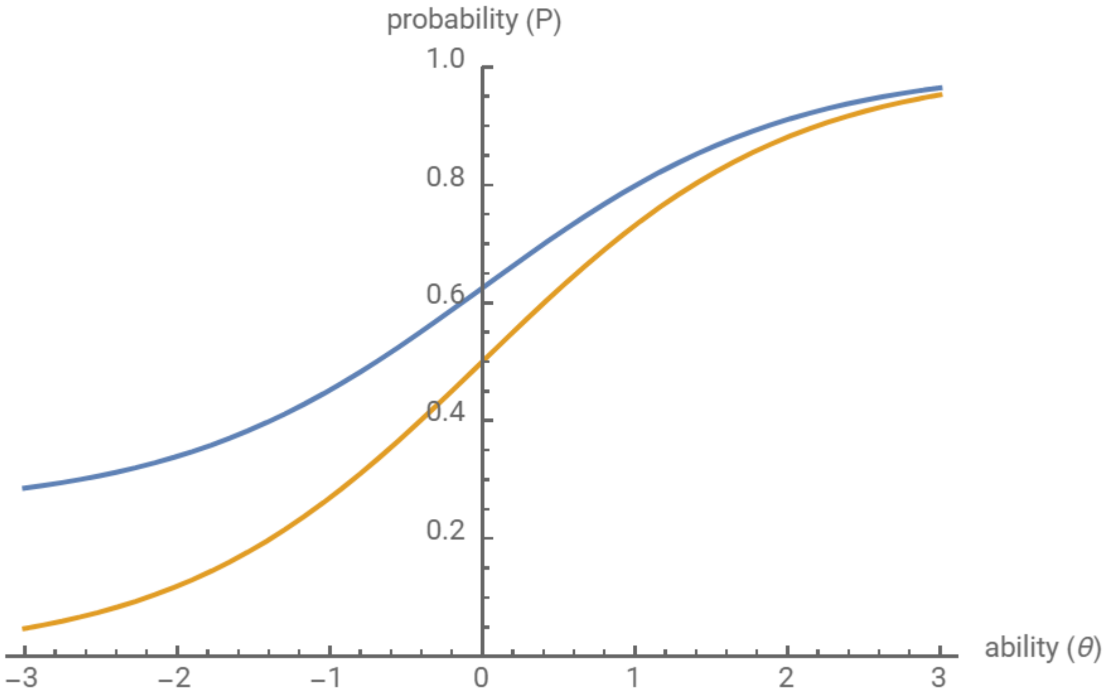


Figure 17 Two item characteristic curves with the same item difficulty and discrimination but different guessing parameter

Blue curve: Discrimination (a) = 1, difficulty (b)=0, guessing (c)=0.25

Yellow curve: Discrimination (a) = 1, difficulty (b)=0, guessing (c)=0

The item information function graphs illustrate how much information each item (multiple-choice question) produces for measuring the latent trait. Generally, a higher item information curve is determined by higher item discrimination and greater item difficulty at a specific value of theta (ability) relative to other items in the scale. The test information function is the summation of the item information functions at each value of theta (ability) for all items in the scale. Item analysis is the process that can confirm if the items on a test are functioning in the expected way (De Ayala, 2009). Each item in the multiple-choice test has to be constructed in such a way that it can measure the higher-order cognitive skills of test-takers while evaluating their understanding and application process appropriately (Palmer & Devitt, 2007). Item analysis using IRT models, help establish the difficulty, discrimination and guessing parameters for each item in the test. There are different approaches and models used in the construction of tests using item response theory. The one-parameter model (1PL) is the more restrictive model among other IRT models. All items are given the same weight in identifying the latent trait of a test taker. 1PL models use only the item difficulty parameter such as the Rasch Model. The Rasch model is based on the assumption that the guessing parameter and item discrimination parameter are constant (equal to 1 or negligible) (Embretson & Reise, 2000). Simply, according to the Rasch model the easier the item the more likely the test taker can answer it correctly and the higher the ability of the test taker, the more likely he/she can get the item right compared to a less able test taker (van der Linden, 2015). Some approaches use the two-parameter model (2PL) with item discriminations and item difficulties. Other approaches use a three-parameter model (3PL) for the probability of test-takers with very low levels of ability to answer correctly. IRT uses logistic regression to formulate observed data. A graphical representation of this logistic regression also known as the item characteristic curve (ICC) as shown in Figure 15, Figure 16 and Figure 17. In the item characteristic curve plots, the horizontal axis shows the ability and the vertical axis represents the probability of answering an item right. There are two important points in describing the shape of the curve. The first one is the curve's middle point location, the second is the slope of the curve at the middle point. The midpoint is at $P(\boldsymbol{\Theta})=0.5$ (midpoint between 0.4 and 0.6), and its corresponding value along the ability scale Θ is defined as item difficulty. That is the ability value at a 50% probability of answering right. Item characteristic curve plots show that the higher the item difficulty, the greater level of ability is required to have a 50% probability of correct answer. If the midpoint slope is big, the curve is steeper, which means respondents with high abilities have a more probability to answer correctly than those who have low abilities. If the midpoint slope is small, the curve is flatter that is respondents with high abilities have nearly the same probability to answer correctly as those with low abilities. Thus, the slope at the midpoint shows the measure of an item's discrimination.

IRT is also known as an itemized theory, where each item of the test measures the underlying latent trait. As a result, the amount of information obtained from a single item can be calculated at any ability level. Item Information Function is a mathematical way to calculate how much information each Item Characteristic Curves (ICC) can provide. The item information function can provide the information and precision of a particular item parameter, whereas the test information function can provide us with the same information at the test level (Frank B. Baker, 2001). A test is a set of items, therefore, the test information at a given ability level is the sum of the item information at that level. In this study, the test information function is calculated for each quiz with 10 items each. The test information function is one of the useful features of the IRT, which shows how well the test is doing in estimating ability over the whole range of ability scores. IRT is a modern test theory based on the relationship between individuals' performance on a test item and the test takers' level of performance on an overall measure of the ability the item is designed to measure. In this research study, an integrated development environment for the R language, RStudio was used for the IRT models. R is a programming language for statistical computing and graphics widely used by researchers. Being open-source and free makes RStudio accessible. R language is highly used in many academic and applied settings, which allows many scientists and researchers to develop good and reliable programs. Psychometrics and particularly the item response theory is no exception. In the RStudio environment, for IRT models with one or more item parameters, several packages can be used. For instance "ltm" package (Rizopoulos, 2018) allows the fitting of the one-parameter 1PL model (Rasch) and helps in estimating two-parameter 2PL and three-parameter 3PL models to dichotomous items (Linden, 2018). This study applies both two-parameter (2PL) and three-parameter (3PL) models in the multiple-choice test for measuring computational thinking performance and discusses findings of the obtained results for the two models. The more complex model in IRT does not always mean it is better than the simpler model, it is better to choose the model which has a positive impact in terms of increases in the predictive validity of the test scores (Lacourly, Martin, Silva, & Uribe, 2018).

Chapter 4 Results

4.1 Introduction

The computational thinking performance of secondary school students in Kazakhstan is measured by using bespoke multiple-choice questions that cover the following concepts of computational thinking: logical thinking, abstraction and generalisation. Students' perception of computational thinking skills is measured by the computational thinking scale questionnaire that covers the following five factors of computational thinking: creativity, algorithmic thinking, cooperation, critical thinking and problem-solving. The multiple-choice test and the questionnaire results are used as primary data. For the secondary data, the General Knowledge Test results that include scores for several school subjects are used, which are grouped into science, language and subjects, as indicators of students' general school achievement. Descriptive statistics, reliability analysis, correlations, regression and the Item Response Theory (IRT) results are discussed in this chapter. For the data analysis, the RStudio, an open-source integrated development environment (IDE) for R language and the SPSS Statistics software package programs were used. Although this study does not focus on the gender of the students, school type or language of classroom instruction, the cross-tabulations and test results are provided for each mentioned subgroup, as was mentioned in Chapter 1.

In the first section, the descriptive statistics are presented, showing an overview of the level of computational thinking performance (CTP) of secondary school students and the subsets of the computational thinking performance – logical reasoning (CTP_L), abstraction (CTP_A) and generalisation (CTP_G). Secondly, descriptive results on students' school achievement (GKT) and its subsets – science subjects (SS), language subjects (LL) and humanities subjects (HUM) are included as secondary data. Then, students' perception of their computational thinking skills (CTS) with the subsets - creativity (CTS_CR), algorithmic thinking (CTS_AT), cooperation (CTS_CO), critical thinking (CTS_CT) and problem-solving (CTS_PS) are presented. At last, Pearson's correlation with scatterplots for the variables - CTP, CTS, GKT, SC, LL and HUM are discussed. The next section discusses the reliability of the computational thinking scale questionnaire using Cronbach's alpha. Then, the multiple regression analysis is presented. The final section of this chapter discusses the Item Response Theory that is about validating the computational thinking performance test.

4.2 Descriptive results

This section describes the statistical results for the general school achievement of the students. After that, the perception of computational thinking skills statistics is presented. Finally, the correlation table for the used variables is given.

There are 192 girls from girls' schools out of 226, the rest 34 are from mixed schools; 518 boys from boys' schools out of 549, the rest 31 are from mixed schools, as shown in Table 11.

Table 11 School type and gender cross-tabulation

		Gender		Total
		Female	Male	
School Type	Boys	0	518	518
	Girls	192	0	192
	Mixed	34	31	65
Total		226	549	775

The Kazakh instruction language group consists of 461 participants from boys' schools, 187 participants from girls' schools, and 65 participants from mixed schools. The Russian instruction language group consists of 57 participants from boys' schools, 187 participants from girls' schools, and no participant from mixed schools, as demonstrated in Table 12.

Table 12 School type and instruction language cross-tabulation

		Instruction language		Total
		Kazakh	Russian	
School Type	Boys	461	57	518
	Girls	187	5	192
	Mixed	65	0	65
Total		713	62	775

Appendix G contains a table where the average scores of students for each school are presented to make a comparison among all participant schools to see how each school differs from one another. The average scores for computational thinking performance (CTP), perception of computational thinking skills (CTS) and the general school achievement (GKT) were found as 14.8, 74.8 and 11.5 respectively which are more precisely presented in the next section.

4.2.1 Computational thinking performance

The descriptive statistics are shown in Table 13-Table 16. The computational thinking performance is measured by the multiple-choice test which was specially constructed for 8th-grade secondary school students as described in Chapter 3. The test focuses on logical reasoning, abstraction and generalisation concepts of computational thinking skills. The computational thinking performance (CTP) consists of the following subscales: Logical reasoning (CTP_L), abstraction (CTP_A) and generalisation (CTP_G). CTP_L, CTP_A and CTP_G subscales are scored a maximum score of 10 for each subscale, therefore the maximum score for CTP is 30.

The mean of computational thinking performance (CTP) for all 775 participants is 14.8, a maximum score of 29.5 and a minimum of 4.5 as shown in Table 13. The computational thinking performance (CTP), the perception of the computational thinking skills (CTS), general knowledge test results (GKT) and its subscales: the science subject scores (SC), the language subject scores (LL) and the humanities subject scores (HUM) are continuous variables. Therefore, to see the difference between two groups such as boys and girls or Kazakh group and Russian group, the t-test is used. When there are more than two groups to compare, the ANOVA method is used. In this study, the ANOVA is used to compare the variables CTP, CTS, GKT, SC, LL, HUM between the schools grouped according to their gender-segregated type as boys, girls and mixed schools.

Table 13 Computational thinking performance (CTP) including subscales

	N	Min	Max	Mean	SD
CTP	775	4.5	29.5	14.8	5.2
CTP_L	775	1	10	5.0	2.3
CTP_A	775	1	10	5.0	2.2
CTP_G	775	1	10	4.7	2.3

Table 14 Computational thinking performance (CTP) by gender

Gender	N	Mean	SD
Male	549	14.8	5.5
Female	226	14.8	4.3

The independent samples t-test found no difference in the computational thinking performance between boys and girls, $t(526)=-0.18$, $p=0.85$.

Table 15 Computational thinking performance (CTP) by instruction language

Instruction language	N	Mean	SD
Kazakh	713	14.8	5.2
Russian	62	15.3	5.1

An independent samples t-test found no difference in the computational thinking performance between groups of which the instruction language is Kazakh and Russian $t(72)=-0.70$, $p=0.48$.

Table 16 Computational thinking performance (CTP) by school type

School type	N	Mean	SD

Boys	518	14.8	5.5
Girls	192	14.6	4.3
Mixed	65	15.6	5.1
Total	775	14.8	5.2

A one-way ANOVA test was conducted to find out if the computational thinking performance of secondary school students varies by school type variables (boys, girls, mixed). There are 20 boys' schools, 5 girls' schools and three mixed schools that participated in this study. 518 students from the boys' school, 192 students from girls' schools and 65 students from the mixed school participated in this study as shown in Table 16. The one-way ANOVA found no significant difference in computational thinking performance between school types $F(2,772)=0.86$, $p=0.425$.

4.2.2 School achievement

As a general school achievement of the secondary school students, the results of the General Knowledge Test (GKT) are used. The GKT-the score of the General Knowledge Test that consists of 14 subjects, which are grouped into the following three blocks: Science (SC), Language (LL) and Humanities (HUM) subjects' achievement. Science (SC) is the mean value of algebra, geometry, physics, chemistry, biology and informatics subjects'; Language (LL) is the mean value of Kazakh, Russian, Turkish and English languages; Humanities (HUM) is the mean value of Kazakh literature, geography, general history and Kazakh history subjects. Table 17 presents descriptive statistics for GKT and its subscales: Science (SC), Language (LL) and Humanities (HUM) subjects' achievement. Table 18 shows the statistics for each subject in the General Knowledge Test.

Table 17 School achievement descriptive statistics

	N	Min	Max	Mean	SD
GKT [0-20]	775	6.1	18.5	11.5	2.3
SC [0-20]	775	5.1	17.6	10.5	2.3
HUM [0-20]	775	4.4	15.8	10.1	2.1
LL [0-30]	775	5.4	26.2	14.4	3.7

Table 18 Subject scores descriptive statistics for the General Knowledge Test

	Subjects	N	Min	Max	Mean	SD
SCIENCE	ALG	775	2.0	20.0	13.3	3.4
	GEOM	775	2.6	20.0	9.2	3.6
	PHYS	771	1.0	18.2	11.1	3.1
	CHEM	775	3.2	18.5	10.1	2.8
	BIO	775	3.0	19.0	10.2	2.7
	INF	774	2.2	20.0	9.1	2.7
LANGUAGE	KAZ	775	2.3	19.5	11.8	3.4
	RUS	775	2.0	18.6	9.8	3.3
	TUR	770	3.0	37.7	17.6	6.2
	ENG	775	6.0	38.0	18.5	6.1
HUMANITIES	KAZ_LIT	775	2.0	20.0	10.4	3.1
	GEOG	775	3.3	17.0	10.3	2.2
	GH	775	1.5	17.0	8.3	2.3
	KZH	775	2.3	19.0	11.5	3.0

Using the independent samples t-test for the general school achievement of secondary school students, as seen, it is found that there is a significant difference between boys and girls $t(773)=-10.3$, $p<.001$. Looking at the t-test results, it can be concluded that girls are better than boys in general school achievement, and this population sets in given participants have intrinsic differences and these differences are not by chance. School achievement (GKT) statistics with gender are presented in Table 19.

Table 19 School achievement (GKT) overall statistics with gender

Gender	N	Mean	SD

Male	549	11.0	2.2
Female	226	12.8	2.2

School achievement (GKT) overall statistics with instruction language are presented in Table 20. Using the independent samples t-test for the general school achievement of secondary school students, it is found that there is a significant difference between the groups, of which the instruction language is Kazakh and Russian, where Kazakh group students are better than the Russian group students t (773)=4.77, p<.001.

Table 20 School achievement (GKT) overall statistics with instruction language

Instruction language	N	Mean	SD
Kazakh	713	11.6	2.3
Russian	62	10.2	2.1

The school achievement (GKT) for each school type (boys, girls and mixed) were compared by a separate one-way ANOVA test. The ANOVA result shows that there is a significant difference in the school achievement of secondary school students F(2,772)=62.03, p<.001. School achievement (GKT) by school types is shown in Table 21.

Table 21 GKT by school types

School type	Mean	N	SD
Boys	11.0	518	2.1
Girls	13.0	192	2.2
Mixed	10.6	65	2.2
Total	11.5	775	2.3

Table 22 SC for school types

School type	Mean	N	SD

Boys	10.2	518	2.2
Girls	11.6	192	2.2
Mixed	9.8	65	2.1
Total	10.5	775	2.3

Table 22 shows the subscale Science (SC) for school types, where girls schools did better than boys and mixed schools in science.

Table 23 LL for school types

School type	Mean	N	SD
Boys	13.5	518	3.2
Girls	17.1	192	3.4
Mixed	13.1	65	3.5
Total	14.4	775	3.7

Table 23 shows the subscale Language (LL) for school types, where girls schools' average score is higher than boys and mixed schools in language.

Table 24 HUM for school types

School type	Mean	N	SD
Boys	9.9	518	2.0
Girls	11.0	192	1.8
Mixed	9.4	65	2.0
Total	10.1	775	2.0

Table 24 shows the subscale Humanities (HUM) for school types, where girls schools' average score is higher than boys and mixed schools in language. The histograms in Appendix K show that the General Knowledge Test (GKT), its subscales Science (SC), Language (LL) and Humanities (HUM) are normally distributed.

4.2.3 Perception of the computational thinking skills

This section presents the statistical findings for the perception of computational thinking skills. The computational thinking scale (CTS) includes the following areas of computational thinking: creativity (CTS_CR), algorithmic thinking (CTS_AT), cooperation (CTS_CO), critical thinking (CTS_CT) and problem-solving (CTS_PS), as mentioned in Chapter 3. Table 25 presents descriptive statistics of perception of the computational thinking skills and its subscales.

Table 25 Perception of computational thinking skills and its subscales descriptive statistics

	N	Items	Mean	Score	SD
CTS_CR (Creativity)	775	4	3.8	15.3	3.1
CTS_AT (Algorithmic thinking)	775	4	3.6	14.1	3.9
CTS_CO (Cooperation)	775	4	4.0	15.8	4.0
CTS_CR (Critical thinking)	775	4	3.7	14.6	3.5
CTS_PS (Problem-solving)	775	6	2.5	14.5	5.1
CTS (Total)	775	22	3.7	74.3	12.3

Using the independent samples t-test for the perception of the computational thinking skills of secondary school students, it is found that there is no significant difference between boys and girls, $t(773)=0.85$, $p=0.4$. Table 26 presents the perception of the computational thinking skills (CTS) for each gender group, where boys' and girls' CTS scores were close to each other.

Table 26 Perception of the computational thinking skills (CTS) overall statistics with gender

Gender	N	Score	SD
Male	549	74.6	12.7
Female	226	73.7	11.5

Using the independent samples t-test for the perception of the computational thinking skills of secondary school students, it is found that there is no significant difference between the groups, of which the instruction language is Kazakh and Russian, $t(773)=1.21$, $p=0.23$. Table 27 presents statistics of perception of the computational thinking skills of the secondary school students for each of the instruction language groups (Kazakh/Russian).

Table 27 Perception of the computational thinking skills (CTS) overall statistics with instruction language of the secondary school students

Instruction language	N	Score	SD
Kazakh	713	74.5	12.4
Russian	62	72.5	11.9

The number of participants from boys' schools was much bigger than the girls' and mixed schools, as shown in Table 28.

Table 28 CTS for school types

School type	Score	N	SD
Boys	74.63	518	12.7
Girls	73.84	192	11.9
Mixed	73.23	65	10.1
Total	74.32	775	12.3

A one-way ANOVA test was conducted to find out if the perception of computational thinking skills of secondary school students varies by school type variables (boys, girls, mixed). There are 20 boys' schools, 5 girls' schools and three mixed schools that participated in this study. The one-way ANOVA found no significant difference in perception of computational thinking skills between school types $F(2,772)=0.561$, $p=0.571$.

4.3 Validating the instruments and IRT

How to measure the computational thinking performance of the BIL students by multiple-choice questions is addressed as the first research question. The test questions are constructed carefully and after data collection, the responses to the bespoke multiple-choice questions are analysed using Item Response Theory. Item response theory (IRT) is a modern test theory based on the relationship between individuals' performance on a test item and their level of performance on an overall measure of the ability the item intended to measure. This research uses both two-parameter (2PL) and three-parameter (3PL) models. The differences and similarities of the obtained results from both 2PL and 3PL models are discussed. The 2PL is a two-parameter model with "item difficulty" and "item discrimination" coefficients and the 3PL is a three-parameter model with an additional "guessing" parameter compared to the 2PL model, as mentioned in Chapter 3, section 3.10.1. The two-parameter model and three-parameter models are based on five quizzes (Logic narrative, Logic numbers, Abstraction, Decode, and Pattern) with 10 multiple-choice items each with a total of 50 items. All multiple-choice items are dichotomously scored, given binary values of 0 or 1, indicating the answer is correct (1) or not (0). As the two-parameter model (2PL) and three-parameter model (3PL) are used in this study, the following item characteristic parameters are used: student ability theta ($\boldsymbol{\Theta}$), the item discrimination parameter ***a***, the item difficulty parameter ***b*** and the guessing parameter ***c***. The item discrimination (a) measures the extent to which the item is measuring the construct (computational thinking). The item difficulty (b) measures the difficulty of answering the item correctly. The guessing parameter (c) is used to test the effects of guessing on the probability of answering the item correctly.

The multiple-choice test used in this study consists of 5 quizzes each of which has 10 questions. These multiple-choice questions focus on the logic, abstraction and generalisation concepts of computational thinking, which were chosen after exploring the national curriculum, school syllabi and informatics textbooks. Five quizzes are labelled as follows: Quiz1 (Logic narrative), Quiz2 (Logic numbers), Quiz3 (Abstraction), Quiz4 (Decode) and Quiz5 (Pattern).

4.3.1 Two-parameter model (2PL)

In the two-parameter model, the discrimination coefficients are above 1 in most cases, which means all item characteristic curves fit well except for three outliers- item 1 (X1 curve) and item 6 (X6 curve) in Quiz3 (Abstraction) and item 7 (X7 curve) in Quiz4 (Decode) with item difficulty coefficients of 6.2, 2.0 and 2.4 respectively. As mentioned in Chapter 3, section 3.10.1, the steeper the curve for the item, the less discriminable the item, correspondingly the flatter the curve, the better discriminable the item, which means that the item can suit examinees with different abilities. Item discrimination analysis shows how each question is related to overall test performance.

Quiz1 (Logic narrative): The item discrimination coefficients for items in Quiz1 are between the range of 0.85 and 2.31, the item difficulty coefficients are between the range of -0.73 and 1.12 as shown in Table 29.

Table 29 Item difficulty and discrimination coefficients for Quiz1 Logic narrative

	Item1	Item2	Item3	Item4	Item5	Item6	Item7	Item8	Item9	Item10
Difficulty	-0.7	-0.4	0.1	-0.5	0.1	-0.1	-0.3	0.4	1.12	1.1
Discrimination	0.9	1.2	1.3	1.6	1.2	1.9	2.3	1.4	0.9	1.2

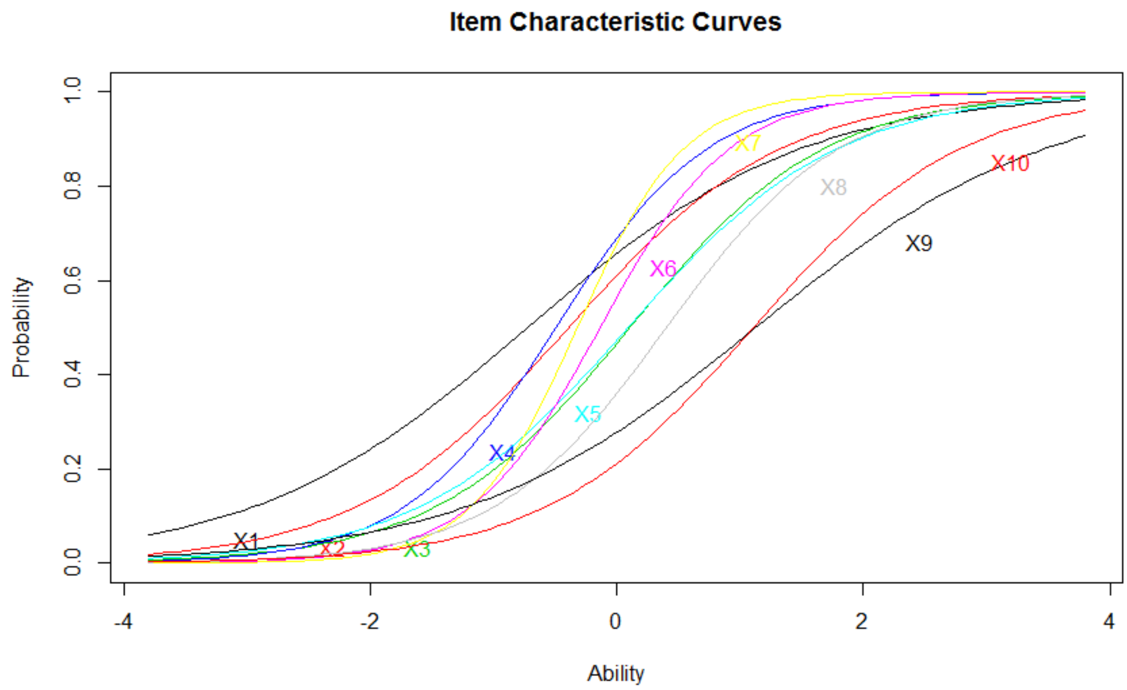


Figure 18 ICC for Quiz1 (2PL)

As can be seen in Figure 18 characteristic curves of items of Quiz1, item 1 (X1 curve) is the easiest item in Quiz1 with a positive discrimination value ***a***=0.9, showing a monotonically increasing S-curve in the range of ***Ө***ε[-4, +4]. The value along *Ө* scale for the curve middle point is ***b***=-0.7 which is the item difficulty coefficient, which means test takers whose ability value is -0.7 have a 50% probability of answering this item correctly. Item2 (X2 curve) displays a steeper curve than item1 (X1 curve) with a more discrimination value of ***a***=1.2 in the range of ***Ө***ε[-4, +4], with a more item difficulty coefficient of ***b***=-0.4. Curves X3, X4, X5, X6, X7 and X8 have similar positions in the plot with close values of item discrimination and item difficulty coefficients. The most difficult items are item9 and item10 (X9 and X10 curves) with discrimination values ***a***=0.9 and ***a***=1.2 in the range of ***Ө***ε[-4, +4] and positive item difficulty coefficients ***b***=1.1 and ***b***=1.2 respectively.

Quiz2 (Logic numbers): The item discrimination coefficients for items in Quiz2 are between the range of 1.0 and 2.3, the item difficulty coefficients are between the range of -1.1 and 0.8 as shown in Table 30.

Table 30 Item difficulty and discrimination coefficients for Quiz2 Logic numbers

	Item1	Item2	Item3	Item4	Item5	Item6	Item7	Item8	Item9	Item10
Difficulty	-1.1	-0.6	0.0	-0.2	-0.6	-0.1	0.8	-0.2	0.1	0.4
Discrimination	1	1	1.5	2.0	2.3	1.5	1.6	1.7	1.6	1.3

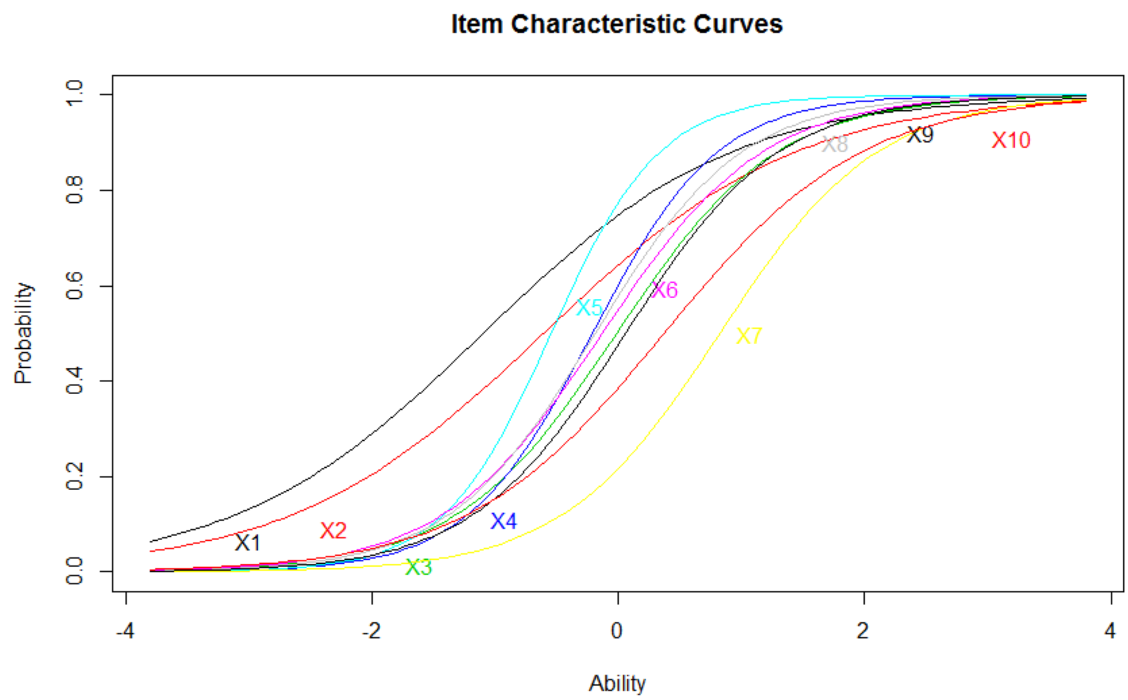


Figure 19 ICC for Quiz2 (2PL)

All items in Quiz2 fit quite well as can be seen in Figure 19. The easiest item here is item1 (X1 curve) with a positive discrimination value ***a***=1.0, displaying an increasing S-curve in the range of ***Ө***ε[-4, +4] with a negative item difficulty coefficient of ***b***=-1.1. The most difficult item is item7 (X7 curve) with an item discrimination coefficient of ***a***=1.6 in the range of ***Ө***ε[-4, +4], and item difficulty ***b***=0.8.

Quiz3 (Abstraction): The item discrimination coefficients for items in Quiz3 are between the range of 0.2 and 1.4, the item difficulty coefficients are between the range of -1.1 and 6.2 as can be seen in Table 31.

Table 31 Item difficulty and discrimination coefficients for Quiz3 Abstraction

	Item1	Item2	Item3	Item4	Item5	**Item6**	Item7	Item8	Item9	Item10
Difficulty	**6.2**	-1.1	-1.1	-0.3	0.2	**2.0**	-0.5	0.1	0.3	0.0
Discrimination	0.2	0.8	1.1	1.3	0.9	0.6	1.4	1.3	0.8	1.0

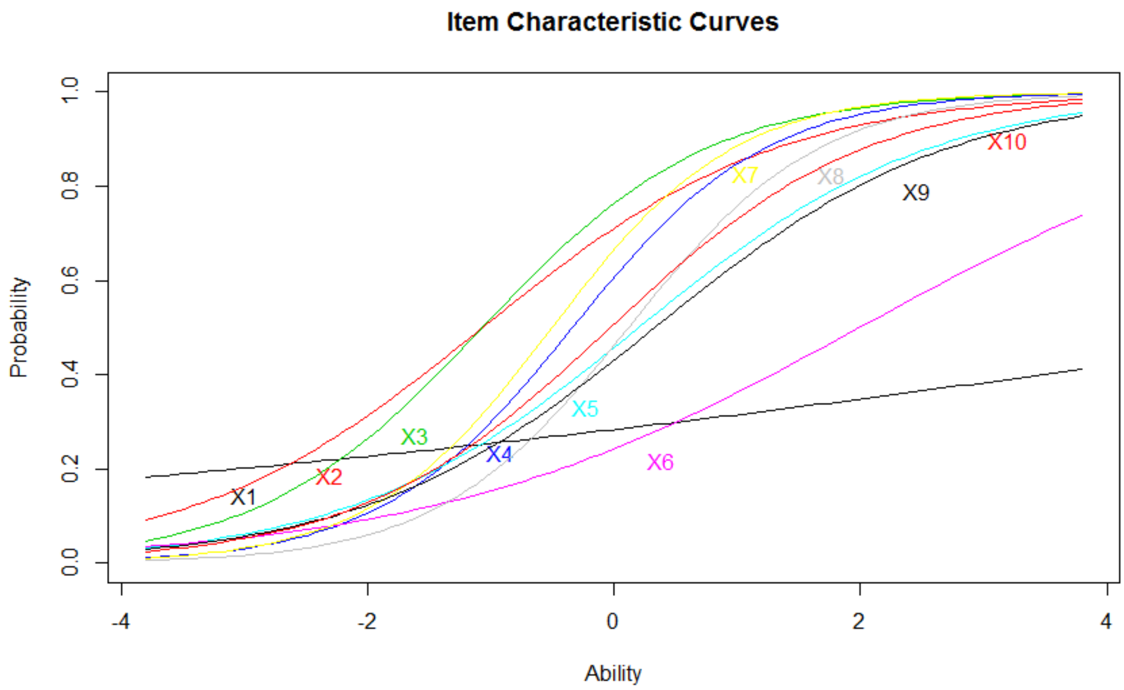


Figure 20 ICC for Quiz3 (2PL)

The most difficult item is item1 (X1 curve) with an item difficulty coefficient of ***b***=6.2 and an item discrimination coefficient of ***a***=0.2. Another item with a high difficulty coefficient is item6 (X6 curve) with an item difficulty coefficient of ***b***=2.0 and an item discrimination coefficient of ***a***=0.6 in the range of ***Ө***ε[-4, +4]. Other items in Quiz3 items have the item difficulty coefficients between -1.1 and 0.3; and their item discrimination coefficients are between 0.8 and 1.4 in the range of

Ɵε[-4, +4]. The most difficult two items X1 and X6 have the least item discrimination values 0.2 and 0.6 respectively in Quiz3 as shown in Figure 20.

Quiz4 (Decode): The item discrimination coefficients for items in Quiz4 are between the range of 0.5 and 2.5, the item difficulty coefficients are between the range of -0.5 and 2.4 as can be seen in Table 32.

Table 32 Item difficulty and discrimination coefficients for Quiz4 Decode

	Item1	Item2	Item3	Item4	Item5	Item6	**Item7**	Item8	Item9	Item10
Difficulty	0.8	0.0	0.1	-0.1	0.0	-0.5	**2.4**	-0.2	0.1	-0.1
Discrimination	1.3	1.7	2.2	2.1	2.5	1.6	**0.5**	1.4	2.2	1.6

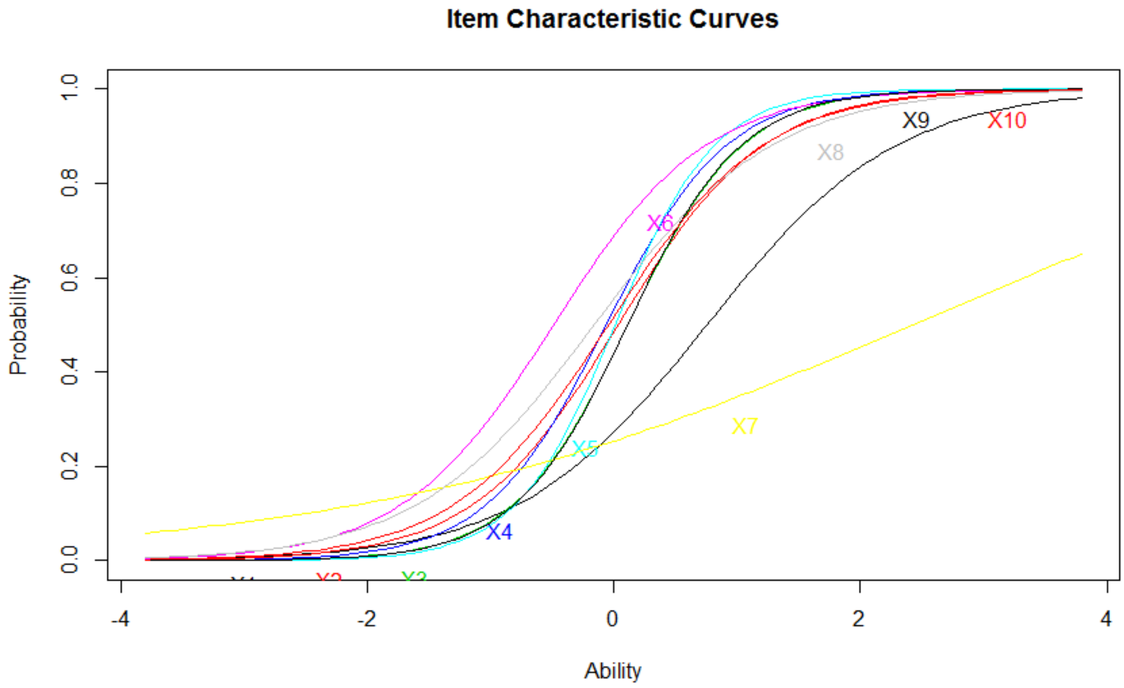


Figure 21 ICC for Quiz4 (2PL)

The most difficult item is item7 (X7 curve) with an item difficulty coefficient of ***b***=2.4 and an item discrimination coefficient of ***a***=0.5 in the range of ***Ɵ***ε[-4, +4]. Other items in Quiz4 items have the item difficulty coefficients between -0.5 and 0.8, item discrimination coefficients between 1.3 and 2.5 in the range of ***Ɵ***ε[-4, +4]. Item7 (X7 curve) has the least discrimination value of a=0.5 in Quiz4 as shown in Figure 21.

Quiz5 (Pattern): The item discrimination coefficients for items in Quiz5 are between the range of 0.8 and 1.5, the item difficulty coefficients are between the range of -0.4 and 1.1 as shown in Table 33.

Table 33 Item difficulty and discrimination coefficients for Quiz5 Pattern

	Item1	Item2	Item3	Item4	Item5	Item6	Item7	Item8	Item9	Item10
Difficulty	-0.5	0.2	0.5	0.3	-0.2	0.4	0.4	0.6	1.0	0.2
Discrimination	0.9	1.0	1.2	1.2	1.2	1.3	1.5	1.3	1.6	1.1

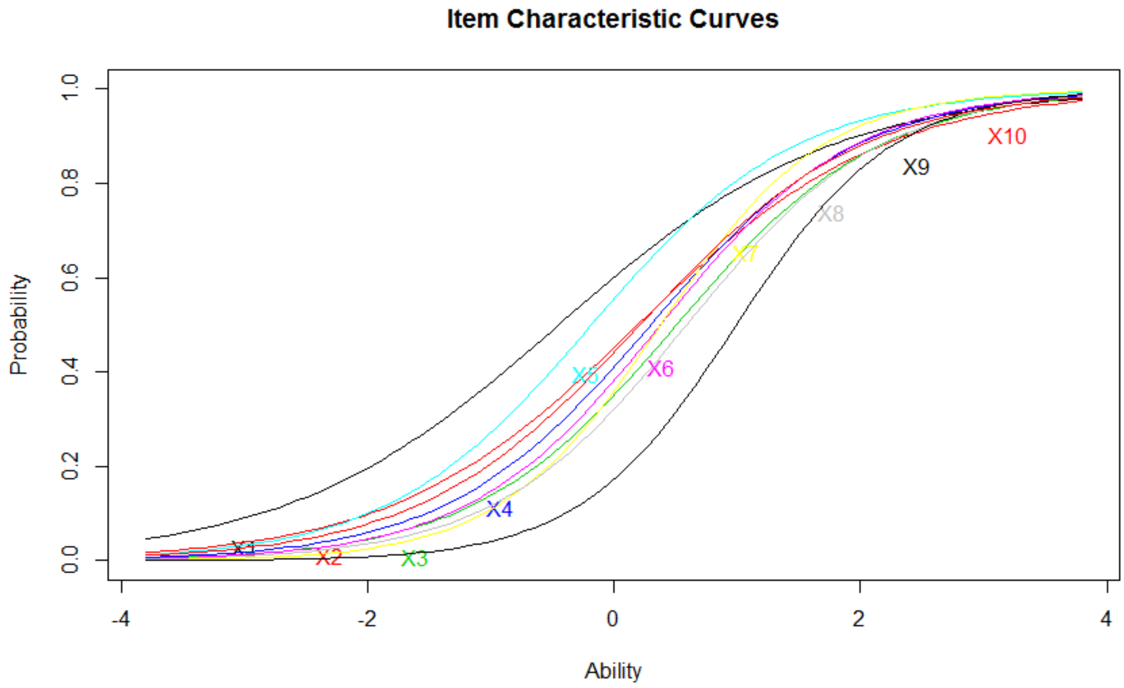


Figure 22 ICC for Quiz5 (2PL)

All items in Quiz5 fit quite well as can be seen in Figure 22. The easiest item here is item1 (X1 curve) with a positive discrimination value of ***a***=0.8, displaying an increasing S-curve in the range of ***ϴ***ε[-4, +4] with an item difficulty coefficient of ***b***=-0.4. The most difficult item is item9 (X9 curve) with an item discrimination coefficient of ***a***=0.4 in the range of ***ϴ***ε[-4, +4], and the item difficulty coefficient of ***b***=1.5.

4.3.2 Three-parameter model (3PL)

In the three-parameter model, the item guessing coefficients' range between 0 and 0.3 except for two items with the guessing parameter 0.5. The guessing parameter can also describe why low-level ability students respond correctly to an item. The item difficulty coefficients are found between -0.7 and 1.3 except three and all item characteristic curves fit well except for those three outliers, which are indicated.

Quiz1 (Logic narrative): The item discrimination coefficients for items (except the items 8, 9 and 10) in Quiz1 are between the range of 1.2 and 4.5. The item difficulty coefficients are between the range of -0.5 and 1.2 as shown in Table 34.

Table 34 Guessing, difficulty and discrimination coefficients for Quiz1 Logic narrative

	Item1	Item2	Item3	Item4	Item5	Item6	Item7	Item8	Item9	Item10
Guessing	0.3	0.0	0.2	0.0	0.3	0.1	0.2	0.2	0.2	0.1
Difficulty	-0.1	-0.4	0.4	-0.5	0.7	0.1	-0.1	0.7	1.2	1.0
Discrimination	1.3	1.2	2.0	1.7	4.5	2.5	3.4	19	24.2	6.9

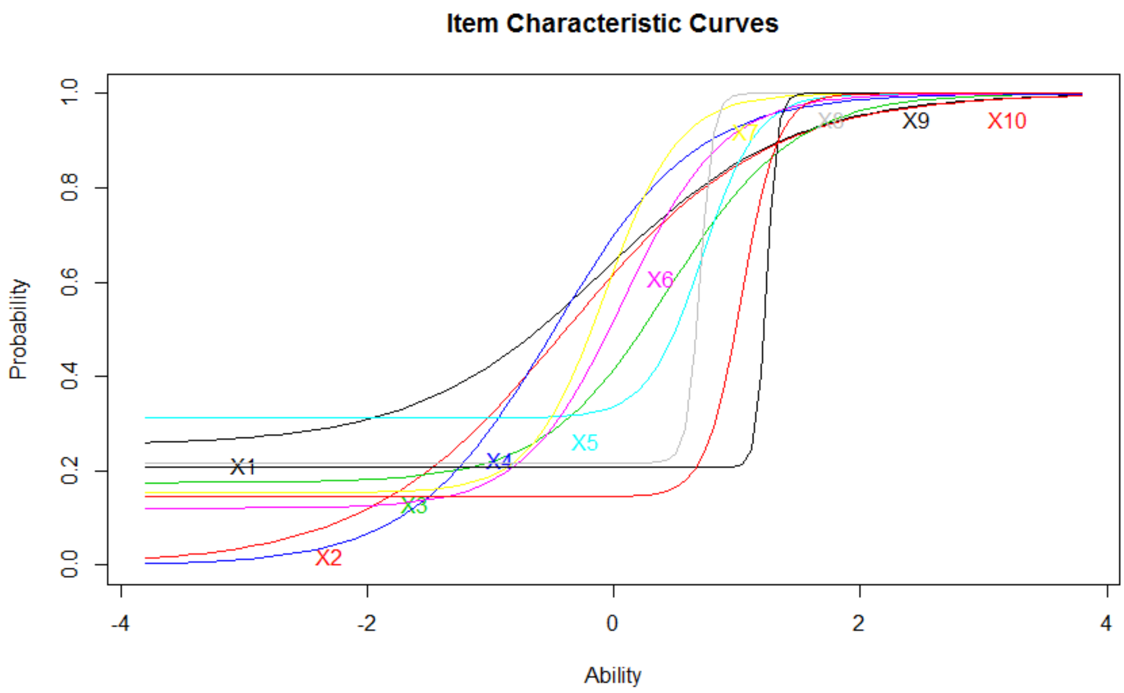


Figure 23 ICC for Quiz1 (3PL)

Although the item with the least difficulty coefficient is item4 (X4 curve) with a positive discrimination value of ***a***=1.4, its guessing coefficient is 0. Item8 (X8 curve) and item9 (X9 curve) display a very steep line as shown in Figure 23 with the discrimination values of ***a***=19 and ***a***=24.2 respectively, with equal item difficulty coefficients of ***b***=0.2. The most difficult items are item9 and item10 (X9 and X10 curves) with discrimination values of ***a***=24.2 and ***a***=6.9 and positive item difficulty coefficients ***b***=1.2 and ***b***=1, guessing coefficients of ***c***=0.2 and ***c***=0.1 respectively.

Quiz2 (Logic numbers): The item discrimination coefficients for items in Quiz2 are between the range of 1.7 and 6.4, the item difficulty coefficients are between the range of -0.3 and 0.8 as shown in Table 35.

Table 35 Item difficulty and discrimination coefficients for Quiz2 Logic numbers

	Item1	Item2	Item3	Item4	Item5	Item6	Item7	Item8	Item9	Item10
Guessing	0.5	0.3	0.2	0.0	0.1	0.3	0.1	0.0	0.1	0.1
Difficulty	0.2	0.2	0.4	-0.2	-0.3	0.4	0.8	-0.1	0.3	0.5
Discrimination	2.7	1.7	3.1	2.0	2.8	5.7	6.4	1.8	2.0	1.6

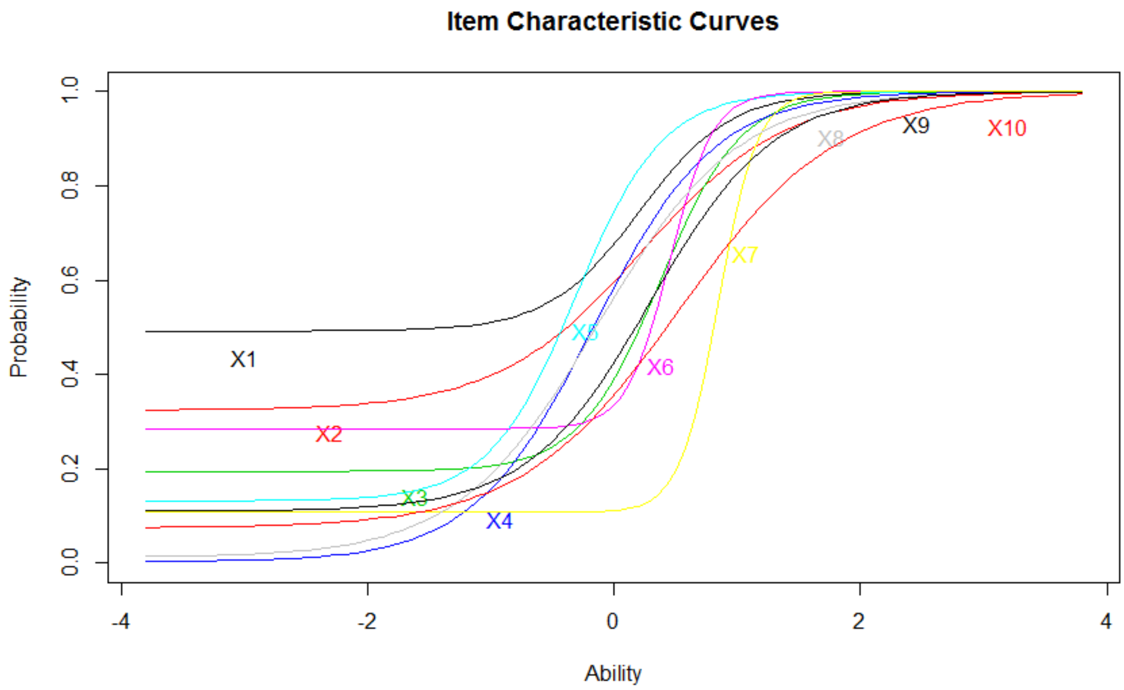


Figure 24 ICC for Quiz2 (3PL)

The easiest to guess item here is item1 (X1 curve) with guessing coefficient ***c***=0.5, discrimination value ***a***=2.7 with item difficulty coefficient of ***b***=0.2 as shown in Figure 24. The most difficult item is item7 (X7 curve) with guessing coefficient ***c***=0.1, item difficulty ***b***=0.8, item discrimination coefficient of ***a***=6.4. Item4 and item8 have the minimum guessing parameter of ***c***=0.

Quiz3 (Abstraction): The item discrimination coefficients for items in Quiz3 are between the range of 0.8 and 8.6, the item difficulty coefficients are between the range of -0. and 6.2 as can be seen in

	Item1	Item2	Item3	Item4	Item5	Item6	Item7	Item8	Item9	Item10
Guessing	**0.2**	0.5	0.2	0.1	0.0	**0.2**	0.0	0.0	0.0	0.1
Difficulty	**1.8**	0.4	-0.7	-0.1	0.2	**1.5**	-0.5	0.1	0.3	0.2
Discrimination	**4.1**	2.6	1.3	1.5	0.9	**8.6**	1.5	1.3	0.8	1.1

Table 36.

Table 36 Item difficulty and discrimination coefficients for Quiz3 Abstraction

	Item1	Item2	Item3	Item4	Item5	Item6	Item7	Item8	Item9	Item10
Guessing	**0.2**	0.5	0.2	0.1	0.0	**0.2**	0.0	0.0	0.0	0.1
Difficulty	**1.8**	0.4	-0.7	-0.1	0.2	**1.5**	-0.5	0.1	0.3	0.2
Discrimination	**4.1**	2.6	1.3	1.5	0.9	**8.6**	1.5	1.3	0.8	1.1

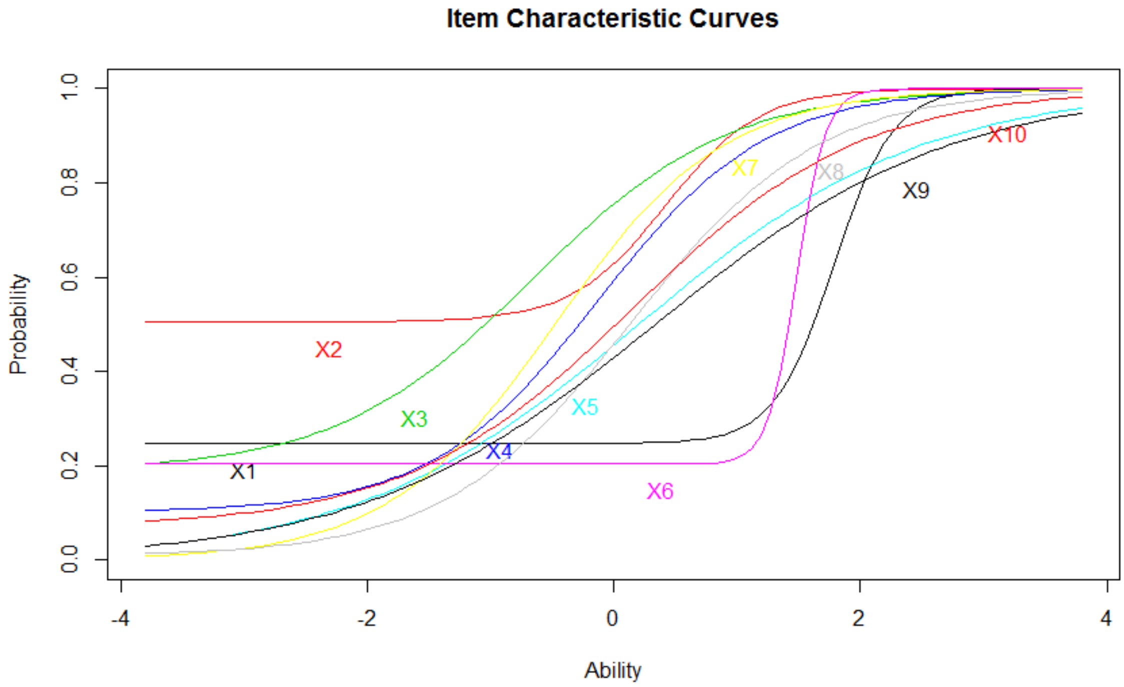


Figure 25 ICC for Quiz3 (3PL)

The most difficult item is item1 (X1 curve) with an item difficulty coefficient of ***b***=1.8 and item discrimination coefficient of ***a***=0.2, guessing coefficient of ***c***=0.2 as shown in Figure 25. The second difficult item is item6 (X6 curve) with an item difficulty coefficient of ***b***=1.5 and an item discrimination coefficient of ***a***=0.9. In Quiz3, items 5, 7, 8 and 9 have the minimum guessing coefficient of ***c***=0. The most difficult two items item1 and item6 have the most item discrimination values ***a***=4.1 and ***a***=8.6 respectively.

Quiz4 (Decode): The item discrimination coefficients for items in Quiz4 are between the range of 0.5 and 15.6, the item difficulty coefficients are between the range of -0.2 and 2.2 as can be seen in Table 37.

	Item1	Item2	Item3	Item4	Item5	Item6	Item7	Item8	Item9	Item10
Guessing	0.1	0.2	0.2	0.2	0.0	0.2	**0.0**	0.0	0.1	0.3
Difficulty	0.8	0.5	0.4	0.3	0.0	-0.1	**2.2**	-0.2	0.2	0.6
Discrimination	1.7	4.3	3.9	4.8	2.8	2.6	**0.5**	1.5	2.5	15.6

Table 37 Item difficulty and discrimination coefficients for Quiz4 Decode

	Item1	Item2	Item3	Item4	Item5	Item6	Item7	Item8	Item9	Item10
Guessing	0.1	0.2	0.2	0.2	0.0	0.2	**0.0**	0.0	0.1	0.3
Difficulty	0.8	0.5	0.4	0.3	0.0	-0.1	**2.2**	-0.2	0.2	0.6
Discrimination	1.7	4.3	3.9	4.8	2.8	2.6	**0.5**	1.5	2.5	15.6

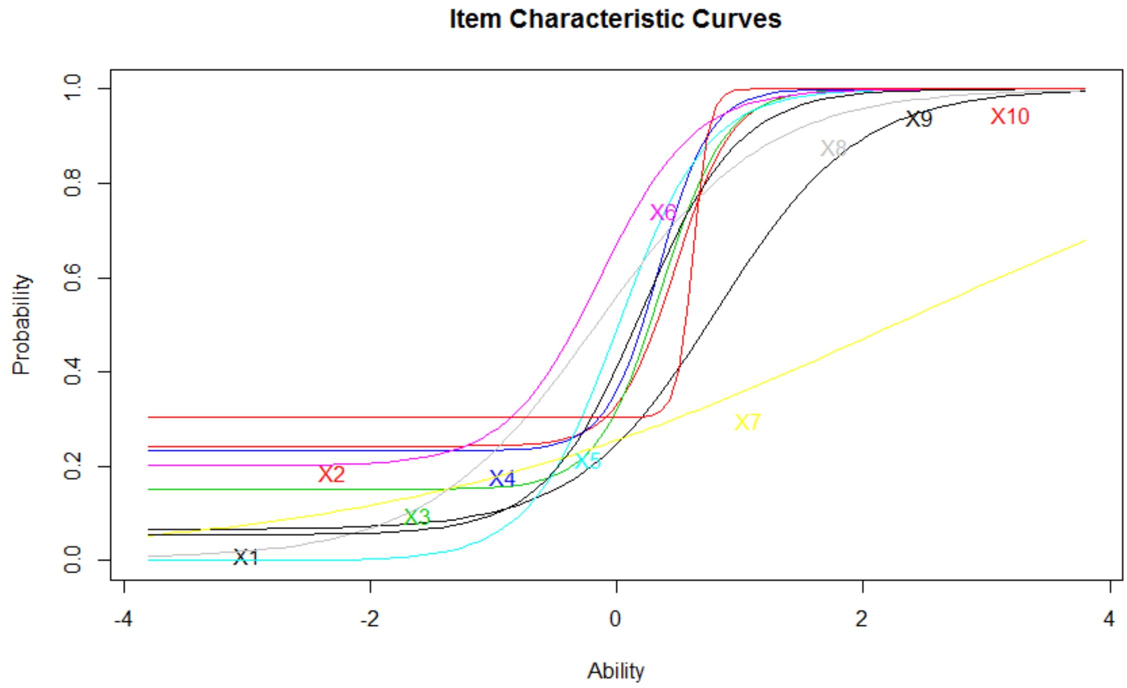


Figure 26 ICC for Quiz4 (3PL)

The most difficult item is item7 (X7 curve) with an item difficulty coefficient of ***b***=2.2, with a guessing coefficient of ***c***=0, item discrimination coefficient of ***a***=0.5 as shown in Figure 26. Other items' difficulty coefficients in Quiz4 differ between -0.2 and 0.8. Item7 (X7 curve) has the least discrimination value of a=0.5. Items 5, 7 and 8 have the minimum guessing coefficient of *c*=0 in Quiz4.

Quiz5 (Pattern): The item discrimination coefficients for items in Quiz5 are between the range of 1.6 and 9.8, the item difficulty coefficients are between the range of -0.1 and 1.3 as shown in Table 38.

Table 38 Item difficulty and discrimination coefficients for Quiz5 Pattern

	Item1	Item2	Item3	Item4	Item5	Item6	Item7	Item8	Item9	Item10
Guessing	0.3	0.1	0.2	0.2	0.0	0.3	0.2	0.2	0.1	0.3
Difficulty	0.3	0.5	1.0	0.8	-0.1	0.9	0.8	0.9	1.3	0.8
Discrimination	2.0	1.5	3.5	2.6	1.6	5.9	2.5	3.5	9.8	3.0

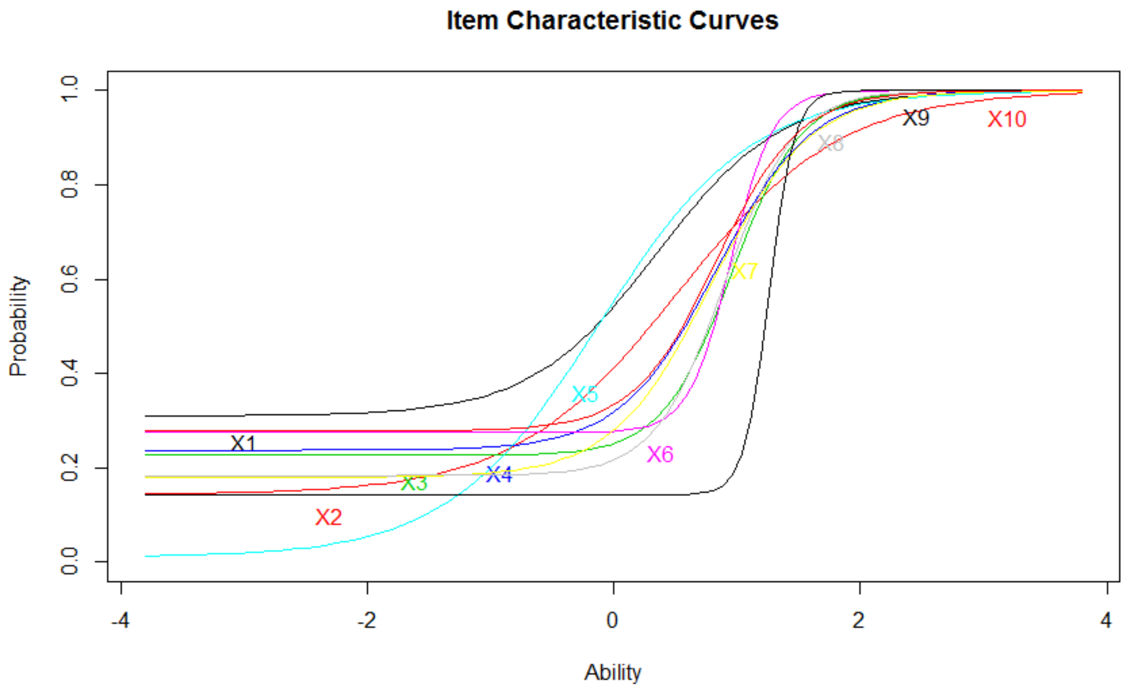


Figure 27 ICC for Quiz5 (3PL)

The perfect item in Quiz5 is item5 (X5 curve) with a positive discrimination value ***a***=1.6, displaying increasing S-curve with item difficulty coefficient of ***b***=-0.1, guessing coefficient of ***c***=0 as shown in Figure 27. The most difficult item is item9 (X9 curve) with the highest item discrimination coefficient of ***a***=9.8, item difficulty coefficient of ***b***=1.5 and guessing coefficient of ***c***=0.1.

4.3.3 IRT Conclusion

The ANOVA test results between the 2PL and 3PL models show that when the p-value is significant as shown in Appendix H, it means the second model (3PL) fit better than the first model (2PL). It is worth noting that 3PL models require a larger sample size (Lord, 1980), usually, the sample size of 1000-2000 participants are recommended (Gao & Chen, 2005). In this study, the number of participants is 775 and although the p values are significant for each quiz, which mean the 3PL models do significant improvement on the 2PL models, there is research indicating that the guessing parameter is less stable. Furthermore, the test information functions in Appendix I-J show that the 2PL model fits better than the 3PL model. Although the 3PL IRT model provides more parameters than other 1 or 2 parameter models, in the 3PL model lower-ability participants who answer correctly difficult items are more likely to have them guessed than the higher-ability participants (Chiu & Camilli, 2013). Therefore, the multiple-choice questions are examined by using both 2PL and 3PL models to test questions with both models.

Table 39 Difficulty and discrimination coefficients of outliers according to 2PL and 3PL IRT models

Questions		2PL		3PL		
Quiz	Item	Diff	Discr	Diff	Discr	Guess
Quiz3	Item1	6.2	0.2	1.8	4.1	0.2
Quiz3	Item6	2.0	0.6	1.5	8.6	0.2
Quiz4	Item7	2.4	0.5	2.2	0.5	0.0

Three outlier items are the most difficult questions, which are Item1 and Item6 in Quiz3 and Item 7 in Quiz4 as shown in Table 39. Item difficulty and item discrimination analysis help to establish the construct validity of these multiple-choice questions (Haladyna & Rodriguez, 2013; Violato, 1991). Besides, the item discrimination coefficients below 0.7 are considered not a good fit. All items' discrimination coefficients are found above 0.8, except for these three outliers. Three outlier items' discrimination coefficients are 0.2, 0.6 and 0.5 in the 2PL model. In summary, the majority of the items fit well but three items stood out as very difficult for secondary school students in both 2PL and 3PL IRT models. Carefully constructed multiple-choice questions that suit the BIL curriculum have been tested using two IRT models and as addressed in the initial research question these findings answer that research question, which will be discussed in the next chapter.

4.3.4 Reliability analysis of the questionnaire

For the measure of scale reliability of the computational thinking scale questionnaire, Cronbach's alpha is used. Cronbach's alpha is a measure of internal consistency, which tests how closely related a set of items are as a group. Cronbach's alpha of .70 and above is considered good.

Table 40 Computational Thinking Scale reliability analysis (Cronbach's alpha)

	N	Mean	Cronbach's alpha
CTS_CR (Creativity)	4	3.86	.599
CTS_AT (Algorithmic thinking)	4	3.55	.834
CTS_CO (Cooperation)	4	3.96	.841
CTS_CR (Critical thinking)	4	3.69	.738
CTS_PS (Problem-solving)	6	2.44	.749
CTS (Total)	22	3.40	.778

The Computational Thinking Scale is a five-point Likert type scale and consists of 22 items that could be collected under five factors. Each one of the items taking place in the factors has been scaled as never (1), rarely (2), sometimes (3), generally (4), always (5). Cronbach's alpha was calculated for the reliability test. The result of Cronbach's alpha for the perception of the computational thinking scale (CTS) for each factor are found as follows, as shown in Table 40: creativity=0.599, algorithmic thinking=0.834, cooperation=0.841, critical thinking=0.738, problem-solving=0.749 and total=0.778, which shows that the instrument is reliable. Participants' perception of computational thinking skills varies between 2.26 and 4.17, and the mean is 3.40. All factors' mean values are close to each other, except the problem-solving factor (CTS_PS) with the coefficient of 2.44, which might be the indication that students are less confident in their problem-solving skills. Both pilot and main phase results show that participants' problem-solving score was the minimum among all other subscales.

4.4 Relationship findings

This section discusses the relationship between computational thinking performance and other variables. First, correlational relationship findings are presented to concisely summarize the direction and strength of the relationships between the variables. Then, regression findings are discussed to see the more detailed analysis that can be used for prediction purposes. These correlations and regression analysis are related in the sense that both deal with relationships among given variables in this study.

4.4.1 Correlational relationship.

This section presents the Pearson correlation coefficients and scatterplots for the variables used in this study, such as CTP, GKT, SC, LL, HUM and CTS as shown in Figure 28.

Variables used in this study are listed as follows**: CTP** - Computational Thinking Performance, **CTS** - Computational Thinking Scale, **GKT** - General Knowledge Test that consists of the following subject variables: PHYS, CHEM, BIO, ENG, KAZ, KAZ_LIT, RUS, ALG, GEOM, CS, TUR, GH, KZH, GEOG, **SC** - Science (Physics, Chemistry, Algebra, Geometry and Computer Science), **LL** - Language Level (Kazakh, English, Russian and Turkish), **HUM** - Humanities (Kazakh Literature, general history, history of Kazakhstan and geography). The computational thinking performance (**CTP**) is the total of the following subscales: Logical reasoning (CTP_L), abstraction (CTP_A) and generalisation (CTP_G). The **GKT**-the score of the General Knowledge Test that consists of 14 subjects, which are grouped into the following three blocks: Science (**SC**), Language (**LL**) and Humanities (**HUM**) subjects achievement. Science (SC)- mean value of algebra (ALG), geometry (GEOM), physics (PHYS), chemistry (CHEM), biology (BIO) and informatics (INF) subjects; Language (LL) – mean value of Kazakh (KAZ), Russian (RUS), Turkish (TUR) and English (ENG) languages; Humanities (HUM) – mean value of Kazakh literature (KAZ_LIT), geography (GEOG), general history (GH) and Kazakh history (KZH) subjects. The perception of the computational thinking skills (**CTS**) is the total of the following: creativity (CTS_CR), algorithmic thinking (CTS_AT), cooperation (CTS_CO), critical thinking (CTS_CT) and problem-solving (CTS_PS).

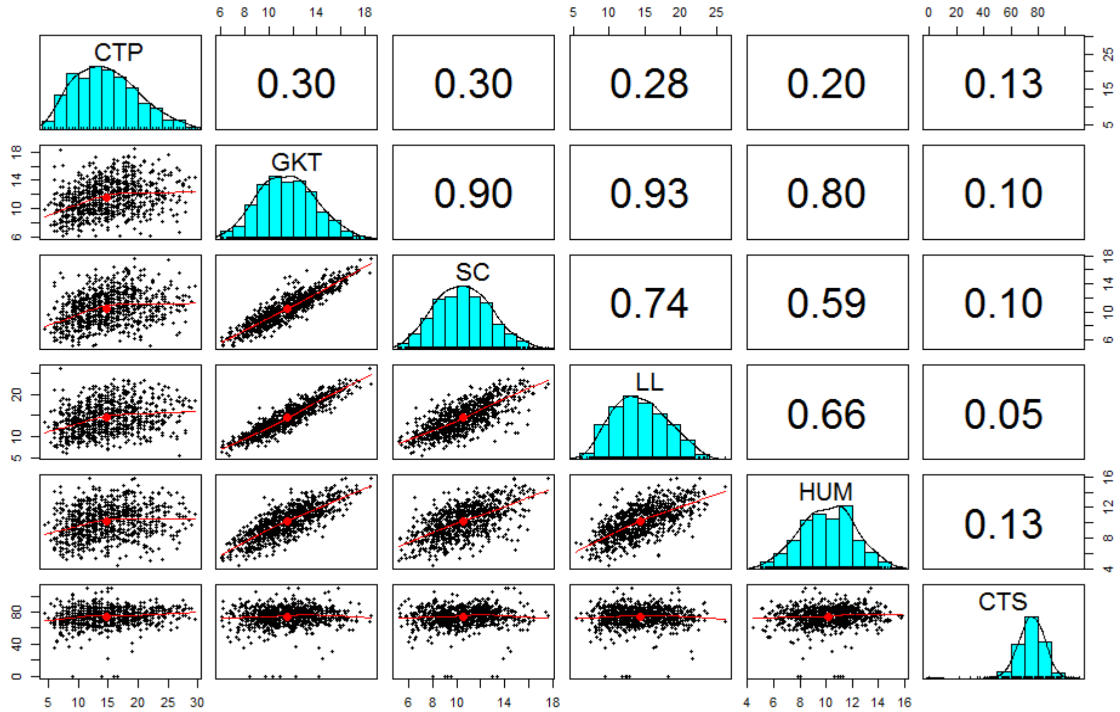


Figure 28 Correlations and plot for CTP

As can be seen from Figure 28, the Pearson correlation results show that the computational thinking performance (CTP) and the school achievement (GKT) are moderately positively correlated, r(773)=.30, p<.001. The computational thinking performance (CTP) and the science-subjects achievement (SC) are the most positively correlated direction among others, r(773)=.30, p<.001. The language-subjects achievement (LL) and the computational thinking performance (CTP) are found to be weakly positively correlated, r(773)=.28, p<.001. The humanities-subjects achievement (HUM) and the computational thinking performance (CTP) are also weakly correlated, r(773)=.20, p<.001. At last, the computational thinking performance (CTP) and the perception of the computational thinking skills (CTS) are found to be weakly positively correlated, r(773)=.13, p<.001. Since the science-subjects achievement (SC) is the most positively correlated among others, the correlations between computational thinking performance (CTP) and each science subject were calculated as shown in Figure 29**,** as usually STEM subjects are associated with computational thinking skills.

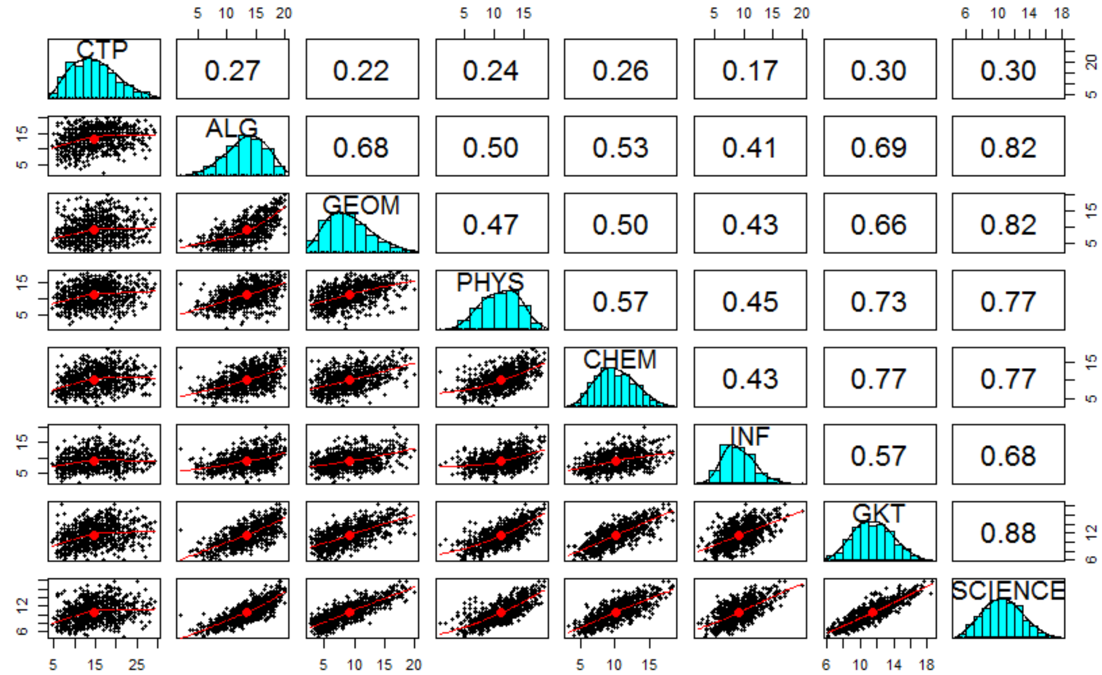


Figure 29 Correlations and plot for CTP with science subjects

Figure 29 shows the Pearson correlation results for each science subject, where the computational thinking performance (CTP) and the algebra (ALG) are the most positively correlated, $r(773)=.27$, $p<.001$. The computational thinking performance (CTP) and other science subjects are correlated as follows, chemistry (CHEM) $r(773)=.26$, $p<.001$, physics (PHYS) $r(773)=.24$, $p<.001$, geometry (GEOM) $r(773)=.22$, $p<.001$ and informatics (INF) $r(773)=.17$, $p<.001$ are positively correlated.

4.4.2 Multiple regression

Regression analysis identifies the relationship between a dependent variable and one or more independent variables. A simple or multiple regression model addresses the relationship among the dependent and independent variables. Dependent variables are usually the variables that are being measured in a study. Independent variables are the variable that is changed or controlled in a study to see the effects on the dependent variables. These relationships between one dependent variable and one or more independent variables are analysed by the multiple linear regression, as mentioned in Chapter 3. Predictors are the independent variables used to predict the dependent variable. In this study, the dependent variable is CTP and predictors are CTS, GKT, SC, LL, and HUM. Two stepwise regression equations were used to address the research questions 2a, 2b and 2c:

RQ 2a: To what extent is there a relationship between the computational thinking performance as measured by the multiple-choice questions and the science-subjects achievement of the secondary school students in Kazakhstan as measured by the General Knowledge Test?

RQ 2b: To what extent is there a relationship between the computational thinking performance as measured by the multiple-choice questions and the language-subjects achievement of the secondary school students in Kazakhstan as measured by the General Knowledge Test?

RQ 2c: To what extent is there a relationship between the computational thinking performance as measured by the multiple-choice questions and the humanities-subjects achievement of the secondary school students in Kazakhstan as measured by the General Knowledge Test?

The first stepwise regression analysis results are shown in Table 41, in which the computational thinking performance (CTP) is the dependent variable with two steps of independent variables. The first step of the regression analysis includes the perception of the computational thinking skills (CTS) and in the second step, the school achievement (GKT) is added. R is the measure of how well the predictors predict the dependent variable. R^2 (R square) is the square root of R that varies between 0 and 1 and gives the amount of variance in the dependent variable explained by the predictors. The R^2 change is tested by F-test which is referred to as the F-change. In the first step, the perception of computational thinking skills (CTS) has a positive and significant effect. CTS is included in the second step, where the GKT is added as the indicator of the school achievement of students. The school achievement can be a predictor of the computational thinking performance of the students, as by adding the GKT in the second step of the regression analysis, the R^2 is increased from 0.125 to 0.319. Both perception of computational thinking (CTS) and school achievement (GKT) are significant predictors of computational thinking performance (CTP).

The magnitude of the CTS coefficient initially was .053 but decreased to .041 after adding GKT as a predictor.

Table 41 Regression of computational thinking performance on the perception of computational thinking skills (CTS) and school achievement (GKT) of the students

	Predictors	*B*	*SE*	*S Beta*	*t*	*Sig*	*R2*
1	(Constant)	10.9	1.13		9.653	p<.001	
	CTS	0.053	0.015	0.125	3.517	p<.001	
							0.125
	F change = 12.368						
2	(Constant)	4.27	1.32		3.217	p<.001	
	CTS	0.041	0.014	0.097	2.827	0.005	
	GKT	0.653	0.076	0.295	8.595	p<.001	
							0.319
	F change = 73.882						

The second step of the hierarchical multiple regression results is shown in Table 42. Three separate variables SC, LL and HUM (Science, Language and Humanities achievements) that form the general school achievement are used to see their effect more precisely. As discussed in the previous regression analysis, the perception of computational thinking skills (CTS) is found to be significant and this is the same case as in the first step as shown in Table 41.

Table 42 Hierarchical multiple regression of computational thinking performance

	Predictors	B	SE	S Beta	t	Sig	R2
1	(Constant)	10.9	1.13		9.653	p<.001	
	CTS	0.053	0.015	0.125	3.517	p<.001	
							0.125
	F change = 12.368						
2	(Constant)	4.954	1.290		3.841	p<.001	
	CTS	0.040	0.014	0.096	2.806	0.005	
	SC	0.652	0.077	0.290	8.461	p<.001	
							0.315
	F change = 71.597						
3	(Constant)	4.436	1.297		3.419	0.001	
	CTS	0.042	0.014	0.100	2.915	0.004	
	SC	0.419	0.113	0.187	3.706	p<.001	
	LL	0.198	0.071	0.140	2.795	0.005	
							0.329
	F change = 7.809						
4	(Constant)	4.688	1.346		3.484	0.001	
	CTS	0.043	0.014	0.102	2.975	0.003	
	SC	0.436	0.113	0.194	3.772	p<.001	
	LL	0.221	0.071	0.157	2.836	0.005	
	HUM	-0.83	0.117	-0.33	-0.709	0.479	
							0.330
	F change = 0.502						

When science-subjects achievement is added in the second step the perception of computational thinking skills (CTS) is still significant and the R2 is increased from 0.125 to 0.315, which shows the better model as shown in

Table 42. The science-subjects achievement variable has a positive association with a significance of 0.29. When the language-subjects achievement variable is added in the third step, it does not change the relative importance of science-subject achievement, the significance of the perception of computational thinking skills (CTS) is changed from 0.005 to 0.004 and there is an insignificant growth in R2 from 0.315 to 0.329. The humanities-subjects achievement variable is excluded due to its non-significant coefficient of 0.479 and beta of -0.33. Therefore, based on the results of the study and multiple regression analysis, it can be concluded that the predictors of the computational thinking performance of secondary school students are the perception of computational thinking skills (CTS), science-subjects achievement (SC) and language-subjects achievement (LL).

In summary, the correlation and regression analysis findings have been presented and revealed several relationships. As can be seen, correlation and regression analysis have many similarities and some differences. Neither regression nor correlation analyses should be interpreted as establishing cause-and-effect relationships, as these relationships can only indicate to what extent variables are associated with each other (Field, 2009).

Chapter 5 Discussion

5.1 Outline

The discussion chapter consists of five sections. The first section covers the first research question "How to measure the computational thinking performance of secondary school students by multiple-choice questions?" and discusses the measurement of computational thinking performance of the Bilim Innovation Lyceum students. It discusses the construction of the multiple-choice questions and the descriptive statistics of computational thinking performance measured by the multiple-choice test. Later, the quality of the multiple-choice questions based on the Item Response Theory models are discussed, where both 2PL and 3PL IRT models show that the test questions are well designed and suit the purpose. The second section addresses the second research question, the relationship between computational thinking performance and the school achievement of secondary school students. Particularly, it discusses the relationship between computational thinking performance and the secondary school students' achievements in sciences, languages and humanities and shows how such a relationship varies across school types, instruction language groups and gender. Then, the third research question addressing the perception of computational thinking skills, whether perception can be a predictor of computational thinking performance is discussed. Finally, a concluding section the summarizes the findings discusses the research journey.

5.2 Multiple-choice test and its measurement

The first research question (RQ1) addresses the measurement of computational thinking performance of secondary school students by multiple-choice questions. A few studies have been developed to assess computational thinking skills by using multiple-choice questions. The assessment of computational thinking by using multiple-choice questions is somewhat hard as it requires high-order cognitive task measurement. There are several instruments used to measure computational thinking skills, which are either tool-specific or programming oriented (Fraillon et al., 2019). The participants' age group at this study was similar to Israel's nationwide examination, a large-scale assessment that aims to measure subjects knowledge and level of thinking using Bloom's taxonomy (Zur-Bargury, Pârv, & Lanzberg, 2013). Grover, Cooper and Pea (2014) used some questions from Israel's national examination that were based on Scratch and developed a formative assessment by multiple-choice questions, which covered decomposition, sequences, conditionals and loops. A thoroughly examined multiple-choice online test was developed by

Román-González, Moreno-León, et al. (2017), which consists of 28 items and covers the following concepts: Sequences, loops, if-else, complex conditionals and functions. The test is more likely programming oriented as it has some elements of loops, if-else and functions, where the assessment of thinking skills should not be limited to measuring programming language oriented (Naveh, 2006). A hybrid assessment by Wiebe et al. (2019) is a short version that contains 19 of the Román-González and Moreno-León, et al. (2017) test and 6 Bebras tasks, a more programming-free version that reduces the test duration and complexity.

The multiple-choice test used in this study was designed specifically not to use any tool-specific attributes such as block programming, pseudo-codes or scripts; and the focus of the study is narrowed down to three concepts of computational thinking as described in Chapter 1. The bespoke 50 multiple-choice questions in this study followed the framework by Diagnostic Questions, which is shown in Appendix F. The multiple-choice questions were developed to measure the computational thinking performance of 8th-grade students, taking into consideration the informatics national curriculum, annual informatics lesson plans with the syllabi and the Informatics textbooks (Shaniyev et al. 2017) at Bilim Innovation Lyceums. The instrument has been carefully constructed in line with the recommendations provided by the experts in constructing multiple-choice questions as presented in Chapter-2. The test questions were assessed by two reviewers with experience in assessing computational thinking. After the test questions were constructed, firstly they were reviewed by seven informatics teachers and piloted on more than a hundred 9th grade students. The multiple-choice test consists of 5 quizzes, where each quiz has 10 questions with a total of 50 questions. The computational thinking concepts included in these multiple-choice questions are abstraction, generalisation and logic with 10, 20 and 20 questions in each quiz respectively. Each question in this multiple-choice test has four response options, with one correct answer and three distractors. A sufficient amount of time and training is needed so that a satisfactory level of skills to be developed (Marina Umaschi Bers, Flannery, Kazakoff, & Sullivan, 2014; Clark, Tanner-Smith, & Killingsworth, 2016), and the data collection period in this study was during May, which is towards the end of the academic year of 8th grade. The results of the study indicate that the mean score of the computational thinking performance of 775 secondary school students is 14.8 out of 30, with a minimum score of 4.5 and a maximum of 29.5. The computational thinking performance results found no significant difference in gender, language instruction group (be it Russian or Kazakh) or school type (boys only, girls only or mixed).

When constructing the questions whether as a formative or summative assessment, the teachers usually may construct the questions as easy, medium or hard based on their personal views and experiences. However, how good or how hard the question is can only be answered by testing it. It is not easy to identify good quality questions without testing the items within a given sample of participants. That is where IRT models are used to see the item difficulty and discrimination parameters. The Item Characteristic Curves demonstrates how items are situated as S-shapes in a plot where the probability of answering correctly on the y-axis and the ability of students on the x-axis. The Test Information Function plots show which range of ability of the participants the overall test questions measures the best. The individual item information functions can be summarized across the quizzes to form test information functions. The test information functions for the quizzes can display quizzes that perform well or poorly. Low information for the quiz may show that the quiz items measure something different being perhaps out of context, or the items are not clearly written being, for instance, poorly worded or too complex for the secondary school students to comprehend.

Both 2PL and 3PL IRT models show that there are three items in the multiple-choice test that can be identified as the most difficult questions. The most difficult items are Item1 and Item6 in Quiz3 that focus on the Abstraction concept, and Item7 in Quiz4 that focuses on the Generalisation concept. The discrimination parameter of an item indicates how well an item discriminates between respondents below and above the item threshold parameter, as shown by the slope of the ICCs in section 4.3, Chapter 4; in other words, discrimination represents the strength of an item's ability to differentiate among participants at different levels along the trait continuum. The discrimination coefficients of multiple-choice items show that the items are quite well differentiated among secondary school students. The Test Information Functions also show that the quizzes overall test well the students with average ability. Item Characteristic Curves illustrate that all items are well fitted except for the most difficult three items. Three outliers are found to be the most difficult questions, which are Item1 and Item6 in Quiz3 and Item 7 in Quiz4 as discussed in section 4.3.3, Chapter 4. These outlier items out of fifty are found as a result of both 2PL and 3PL Item Response Theory models. In IRT the more the item difficulty coefficients, the more the item is difficult. Both 2PL and 3PL IRT models show that these outlier items are the most difficult ones as shown in Chapter 4, section 4.3.3. Although it was expected that the first item in Quiz3 would be one of the easiest questions, it turned out to be the most difficult one among all items. The issue with this item1 in Quiz3 can be the ambiguity and most of the students might have been misled by the shape of the figure at the stem of the question and could not pay attention to the direction of the arrow. There can be several reasons why these items were more difficult, not just ambiguity, but also wordiness and language barrier. Nevertheless, overall these three items out of fifty multiple-choice questions did not make a major effect on the test quality as can be seen from the test information functions in Appendix I and Appendix J. The item discrimination and difficulty coefficients, item characteristic curve plots, and test information plots for all items, except for three outliers in this test, show they all fit well. Thus, it can be concluded that the multiple-choice questions used in this test are suitable for measuring the computational thinking performance of 8th-grade students at Bilim Innovation Lyceums. The positive feedback from BIL informatics teachers who assisted during the data collection, for both pilot study and main data collection was confident that the questions are good enough to measure the deep understanding and thinking skills of students.

5.3 Relationship between computational thinking performance

One of the main goals of education is to support students with skills that can be applied outside the initial learning context. If the transfer is at the core of education, the subjects enhancing computational thinking should be well delivered. This section discusses the relationship between computational thinking performance and school achievement in three sub-scores: the science, language and humanities subjects. In this research study 20 boys' schools, 5 girls' schools and three mixed schools participated with a total of 518 students coming from the boys only schools, 192 from girls only and 65 from mixed schools. The instruction language groups are Kazakh and Russian. Although gender and language groups are not the focus of the study, uneven distribution of boys and girls, Kazakh and Russian groups in the sample should be noted.

RQ2a-The computational thinking performance and the science subjects: The first part of the second research question investigates the relationship between the computational thinking performance and the science-subjects achievement of the secondary school students, as measured by the multiple-choice questions and the General Knowledge Test.

No significant difference in computational thinking performance between school types, language instruction groups or gender was found. However, the independent samples t-tests for the general school achievement found a significant difference between boys and girls, where girls outperform boys, and Kazakh groups do better than Russian groups and also school types have an effect on the general school achievement of students. Although gender is not the main focus of this study, it is interesting to discover the difference between boys and girls. The results of this study support several studies in recent years which have shown that gender factor does not affect the computational thinking skills of students. (Fields, Kafai, & Giang, 2017; J. Lee, Jung, & Park, 2017; Werner et al., 2012). However, the results of this research are controversial to the findings by Prottsman (2011) who claims that women do better at acquiring computational thinking skills and that gender has an effect on computational thinking skills. The findings are also contrary to the study which state that the computational thinking performance of boys is higher than that of girls (Polat, Hopcan, Kucuk, & Sisman, 2021). The results of this study revealed a positive correlation between the computational thinking performance and general school achievement of the Bilim Innovation Lyceums students, which concurs with the results of the study by Gouws et al. (2013b), who studied the assessment of the computational skills of computer science students which suggest students with higher results in the computational thinking test did well in their class tests. The results were similar to the findings of the study by Polat et al. (2021), which found mathematics performance of students positively affected their computational thinking performance more than their IT (informatics) performance. Similarly, the research results of Ambrosio et al. (2014) show a high correlation between problem-solving and mathematics and students' academic success in a related field. Ham (2018) claims that computational thinking is natural and everyone is a computational thinker to some extent. Triggering, controlling and fostering the cognitive sides of computational thinking skills are what makes a person a better computational thinker. Ham (2018) also claims that computational thinking is integrally linked to spatial thinking and they both help us to become more creative. However, the findings of Román-González, Pérez-González and Jiménez-Fernández (2016) show that computational thinking skill is less linked with spatial ability than general mental ability. Lu and Fletcher (2009) claim that people become better at handling information when their computational thinking skills get better. Similarly, Hu (2011) states that developing computing has a positive effect on improvement in computational thinking. To improve classroom lessons and curriculum, there is a need for more research into the relationship between programming, problem-solving and computational thinking (Selby, 2012). The results of this study also support (Boom et al., 2018) results, where a significant and large positive linear relationship between computational thinking and intelligence was found. There can be several factors affecting the computational thinking performance of students such as computer-related educational activities, Project-Based Learning tasks, science experiments, as computational thinking skills are widely utilised in the problem formulation, data collection, data analysis, modelling, solution-based algorithmic steps and digital skills (Fraillon et al., 2019). The correlation results that algebra is the most correlated and informatics is less correlated are similar to the findings by Polat et al. (2021) who found that the IT course achievement affects students' computational thinking performance less than the achievement in mathematics.

Several studies were inconsistent with the results of this research study. Doleck et al. (2017) found no association between computational thinking skills and academic performance. The study results of (Hershkovitz et al., 2019) found no associations between computational thinking and creative thinking. Román-González et al. (2018) found a higher correlation between the achievement in IT courses and computational thinking skills, which also differs from the findings of this study. To understand the relationships between computational thinking and other areas more deeply, simply doing a correlation analysis does not reveal the picture precisely. This research does not touch on the causal effect but rather investigates how a change in one variable affects the other, which is regression analysis. Weintrop et al. (2016) point to the relationship between computational thinking and other subjects, such as mathematics, science, biology, chemistry and physics, especially mathematics, which cannot be separated from computational thinking. The general school achievement is divided into three variables: Science, Language and Humanities subjects achievements. The correlations in Chapter 4, Figure 28 show that science achievements are moderately positively correlated, the language and humanities subjects are weakly correlated. The results in Chapter 4 also show that computational thinking performance is most correlated with Algebra, Chemistry and Physics among science subjects. This correlation might be explained by the fact that computational thinking uses mathematical and engineering thinking skills to produce new products and artefacts (Ozcinar et al., 2017, p39) and as those science subjects heavily use operations with numbers, the concepts of numbers and operations are at the heart of mathematics (algebra) in middle schools (Kilpatrick, Swafford, & Findell, 2001). These results support Sussman's argument that computational thinking and mathematical thinking are not identical but closely related, as they both use abstraction and reasoning (Linn et al., 2010). The results of the multiple regression analysis revealed that the computational thinking performance of secondary school students can be predicted by achievements in science subjects and language subjects. The results of Román-González and Pérez-González et al. (2017) show a positive correlation between grade levels and CT skills by examining the relationship between grades and computational thinking skills. This finding also supports the assumption that CT skills are problem-solving skills. As mathematical thinking and engineering thinking are the foundations of computer science, computational thinking touches both mathematical and engineering thinking. In this study, correlation results presented in Chapter 4, section 4.4 show that the science subjects that correlate most closely with computational thinking performance are algebra, physics and chemistry. This might be the case when those subjects are closely related to problem-solving activities. It could be expected that informatics to be correlated among other science subjects but informatics was found to be the least closely correlated subject. This might be due to the fact that the topics or specific computer application skills and knowledge that were required by students in those periods during data collection did not require considerable thinking skills. This result also raises questions to address the informatics curriculum if it fits the aim and the objectives in the classroom environment. These findings are in line with the studies emphasizing that the development of computational thinking improves students' performance in problem-solving, mathematical modelling, reasoning abilities and academic performance (Calao, Correa, Moreno-Leon, & Robles, 2015). Likewise, the results of this research can support the idea that learning mathematics benefit from integrating computational thinking activities in school curricula (Barcelos, Munoz, Villarroel, Merino, & Silveira, 2018). These positive relationships between computational thinking and mathematics-related subjects can also be explained by commonalities between mathematics and problem-solving as stated by Polya (2004), where abstraction, generalization and problem decomposition skills are described as key skills in problem-solving tasks. These findings can be also be questioned as follows: Do these positive correlations in this study exist because science, algebra and mathematics provide a meaningful set of problems (Jona et al., 2014; Wilensky et al., 2014), in which computational thinking skills can be implemented easily? However, the transfer of computational thinking or applying computing ideas as a solution to daily life issues is a topic of debate. Programming skills has not been confirmed to result in generalised problem-solving skills (Mark Guzdial, 2015). As the answer to the research question 2a can support the statement that computational thinking augments and expands the realm of thinking, logic, mathematics and problem-solving (Sanford & Naidu, 2016), STEM subjects should be more emphasized at schools in order to develop and benefit from computational thinking. Especially BILs can make use of their technology-enriched environment to enhance the development of computational thinking skills of students.

RQ2b-The computational thinking performance and the language subjects: The next research question investigates the relationship between the computational thinking performance and the language-subjects achievement of the secondary school students, as measured by the multiple-choice questions and the General Knowledge Test. The study has taken the combination of all four Kazakh, English, Russian and Turkish languages as achievement in a language subject. The correlations in Chapter 4, Figure 28 show that the computational thinking performance and the language subjects achievement are weakly but positively correlated. The multiple regression analysis results in Chapter 4, Table 42 show that language achievement can be a predictor of computational thinking performance of secondary school students. Interestingly, the results show a certain relationship between computational thinking skills and language skills, therefore, it is worth investigating to what extent computational thinking is linked to linguistic structure (Linn et al., 2010). The results are consistent with the findings of the intervention study by Nesiba et al. (2015), where DISSECT classes demonstrated their abilities to use algorithms, algorithms through spatial reasoning and abstraction. Computational thinking skills enhances non-scientific courses such as English classes and when students can use computational thinking to improve their reading and writing skills, it is more likely they will do better in other disciplines (Nesiba et al., 2015). Computational thinking is a thought process that takes place in the human mind; language is a mode of expressing thoughts. Thoughts cannot be expressed without the help of language, and they cannot be separated. The fact that a language is not stable and monolithic, instead it has a dynamic nature should be taken seriously, especially when it comes to how languages are taught and learned (Counihan, 2008). Using a language that embraces sentences, words, grammar and thoughts, allows people to understand and code unfamiliar objects (Athreya & Mouza, 2017). To enrich the standard language arts curriculum, interactive journalism can be used as a vehicle to enhance the computational thinking skills of middle school students and teachers even for those who do not view themselves as ‘math type’ (Wolz, Stone, Pearson, Pulimood, & Switzer, 2011). In the language classes, computational thinking can be integrated in terms of modelling for primary and secondary schools (Sabitzer et al., 2018). It's important to study further the relation of the use of language for thinking and communication (Counihan, 2008). The sensitivity to both oral and written language skills, ability to learn and utilize these skills to reach certain goals is explained in linguistic intelligence, a part of Gardner's multiple intelligence theory (2011). Borensztajn (2011) underlines the main function of intelligence and memory as being to predict, as in Hawkins theory of memory prediction framework. Various research projects study how cognitive tasks are related to general language use, language used for thinking and language used for talking. Counihan (2008) remarks it cannot be simply implied that literacy in language usage causes an increase in logical reasoning. Some claim that people are born with language knowledge and built-in programs (Gardner, 2011). Whether the language skills are built-in or acquired and learned later, the fact that children learn language skills from very early ages while their problem-solving skills are relatively immature is hard to explain (Chomsky, 2002). The RQ2b reveals that language is an important factor associated with computational thinking, and the recent tendency in implementing three language education in BILs can support the development of computational thinking skills of students.

RQ2c-The computational thinking performance and the humanities subjects: The last part of the second research question investigates the relationship between the computational thinking performance and the humanities-subjects achievement of the secondary school students, as measured by the multiple-choice questions and the General Knowledge Test. The study has taken the combination of all the following subjects as a humanities subjects achievement: Kazakh literature, general history, history of Kazakhstan and geography. The correlations in Chapter 4, Figure 28 show that the computational thinking performance and the language achievements are weakly but positively correlated. The multiple regression analysis results in Chapter 4, Table 42 show that the humanities subjects achievement cannot be a predictor of computational thinking performance of secondary school students. The multiple-choice questions to measure computational thinking performance were designed to measure students' thinking levels. The humanities subjects contain many domain-related facts and concepts, isolated pieces of information, definitions, and specific details. It is might be the case that the humanities questions in the General Knowledge Test were designed to measure some factual and declarative knowledge as introduced by Anderson (1993) rather than students' thinking level, and no strong correlation could be found between students' thinking level and humanities score in this study. Learning a set of factual knowledge is not enough to simply improve comprehension (Glass & Sinha, 2013), and to reach the top of Bloom's Taxonomy, higher-order retrieval activities help best rather than simple factual knowledge retrievals (Agarwal, 2019). Are the humanities subject tests were constructed to measure higher-order thinking of BIL students? If yes, then we could expect a correlation between the humanities score and the bespoke multiple-choice test used in this research but the question is left unanswered. While relying on the General Knowledge Test results as an indicator of general school achievement of BIL students, the quality of the General Knowledge Test is left open for critique and questions. Testing the validity and reliability, and the quality of the General Knowledge Test questions is beyond the scope of this study.

5.4 Perception of computational thinking

The third research question (RQ3) addresses the perception of computational thinking skills of students, as measured by the computational thinking scale questionnaire. It investigates if the perception of computational thinking skills can be a predictor of the computational thinking performance of secondary school students. The results can be explained in light of the five factors included in the adapted version of the computational thinking scale questionnaire. These five factors in this questionnaire are ‘Creativity’ with 4 items, ‘Algorithmic thinking’ with 4 items, ‘Cooperation’ with 4 items, ‘Critical thinking’ with 4 items and ‘Problem-solving' with 6 items. For example, items in the creativity subscale refer to elements that are related to confidence and self-trust, such as believing in one’s ability to solve problems or having the confidence to carry out a plan. Along with the cooperation subscale with 4 items, some items in problem-solving subscales, such as developing one’s thinking and teaching others in a collaborative learning environment, are related to collaboration and group work. The items in algorithmic thinking mainly used mathematical expressions to express their thinking skills. As this questionnaire is self-reported, the results may not reflect the actual computational thinking level of the participants. However, perception is also an important factor to consider especially as an initial measurement in any intervention studies. The independent samples t-test for the perception of the computational thinking skills of students found no significant difference between boys and girls, Kazakh and Russian groups and also between school types. The computational thinking performance and the perception of the computational thinking skills are found to be positively but weakly correlated. The regression analysis results show that the perception of computational thinking skills can be a predictor of the computational thinking performance of secondary school students. The results show the score of the perception of computational thinking skills are 74.3 out of 110, with a minimum score of 22 and a maximum of 110. Despite the nature of the self-assessments instrument, where it is doubtful whether students can and will evaluate themselves honestly and accurately (Korucu, Gencturk, & Gundogdu, 2017), this study utilises the questionnaire and consider perception as an important factor in identifying computational thinking. This computational thinking scale questionnaire is based on the ISTE framework and covers computational thinking on a broader scale. Guggemos, Seufert and Román-González (2019) claim that subscales ‘Creativity’ and ‘Cooperation’ are not parts of actual computational thinking, and comment on the fact that the original version of the questionnaire was found to lack discriminant validity. Interestingly, the results of this study show that the highest scores among the five factors in this computational thinking scales questionnaire are cooperation and creativity with the value of 15.8 and 15.3, respectively. However, Cronbach's alpha for the creativity subscale (CTS_CR) was low 0.599, as shown in Table 40. Self-reported questionnaires are the easier and quicker way to measure creative thinking, but not the most robust one. Hershkovitz et al. (2019) showed no associations between creative thinking and computational thinking but found computational creativity is associated with computational thinking. The problem-solving subscale was the lowest score in the computational thinking scales (CTS). Perhaps students' self-confidence, self-efficacy in problem-solving or students' perception of ability to solve problems (Covington, 2009) were lower than their perception of other skills. It is worth noting that, Román-González et al. (2017) found significant correlations between computational thinking and self-efficacy perception relative to students' computational thinking performance. Cronbach's alpha for creativity was the lowest among others. Creativity is a construct that is hard to define and it keeps its unpredictable features to some extent, which makes it hard to measure (Kaufman & Stenberg, 2010). This study uses the shorter adapted version of the original questionnaire. The Cronbach's alpha for the overall computational thinking scales (CTS) was found as 0.778, which is close to the original author's adapted scale results (Korkmaz et al., 2015), which shows this instrument is reliable.

The results of the perception of the computational thinking skills found no significant difference between boys and girls, and they were consistent with the study findings by Korucu, Gencturk and Gundogdu (2017) and Durak and Saritepeci (2017) who found that the computational thinking levels of students do not differ in terms of their gender. Polat et al. (2021) found a significant but small correlation between computational thinking and perception of computational thinking, where computational thinking perception also did not differ in terms of gender. As shown in Chapter 4, Figure 28 the computational thinking performance is positively correlated with the perception of computational thinking scales. Although the instrument is a self-reported scale, the multiple regression analysis results shown in Chapter 4, Table 42 show that the perception of computational thinking skills can be a predictor of secondary school students' computational thinking performance.

Perception of computational thinking measured by the CTS questionnaire developed by Korkmaz et al. (2017) is found to be a predictor of computational thinking performance of Bilim Innovation Lyceum students. This research study used the adapted version of the original self-reported questionnaire to measure the perception of computational thinking skills of BIL students, which can be administered in a short time. This shorter version of the questionnaire is suitable as a quick application to get an initial indicator of students computational thinking skills.

5.5 Summary of the findings

An important aspect of the study is the measurement of the computational thinking performance of secondary school students. As shown in Figure 30 computational thinking performance of BIL students in Kazakhstan is measured by multiple-choice questions, which used IRT models for further item analysis.

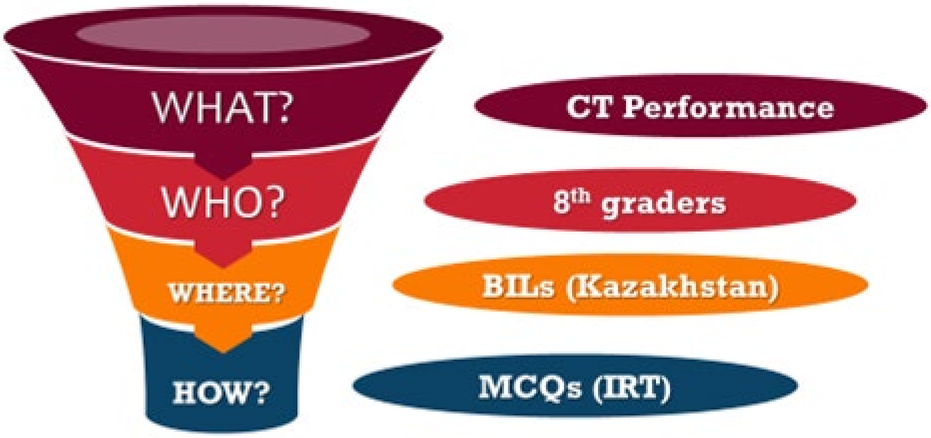


Figure 30 Measurement rationale

The results of this research study found that the 8th-grade BIL students' computational thinking performance score is 14.8 out of 30 as measured by the bespoke multiple-choice test. When multiple-choice questions are constructed correctly in line with the recommendations by the experts in assessment, it is possible to get a good instrument to measure computational thinking.

Computational thinking in the curriculum

The education world is constantly changing with emerging technologies. In recent years the integration of computational thinking can be seen in many countries. As introduced in Chapter-1 the Kazakhstani education system has gone through massive upgrades. By the major changes in the education system at secondary education, the national curriculum in Kazakhstan is updated, and computational thinking is integrated into the curriculum of informatics. In testing, multiple-choice questions are widely used in both school achievement tests and the Unified National Test (UNT), which is the entrance examination for HEIs. These changes in the education system include the assessment approaches where test questions have been redesigned so that they measure and focus more on higher-order thinking abilities (analysing, evaluating and creating) of students than lower-order thinking (remembering, understanding and applying). As discussed in Chapter 2, computational thinking in the curriculum contains the following subcategories: modelling, algorithms and programming. In addition, the main objectives of the curriculum are to teach students to tackle a variety of tasks through analysis, abstraction, modelling, and programming as well as enabling students to gain the abilities to think logically and algorithmically, find patterns, and think in terms of generalisation, decomposition and use evaluation (National Academy of Education, 2016). These objectives in the Kazakhstani curriculum have similarities with the UK programme (Csizmadia et al., 2015), thus the computational thinking concepts used in this research followed the UK's computational thinking model.

How do we see computational thinking?

Although the computational thinking elements had existed in the curriculum and were being practised in the classroom before being officially integrated into the updated curriculum, computational thinking has not been specifically stated as a separate section. Therefore, it can be said that computational thinking existed in the Kazakh curriculum but not formally. Since computational thinking is in the curriculum now, the question "How do we see computational thinking?" arises. As discussed earlier, the computational thinking concepts in the Kazakhstani curriculum are similar to the ones described by the UK system. To see the computational thinking in the curriculum and more importantly its existence in the classroom, the informatics new textbooks and informatics annual plans were studied. As discussed earlier in Chapter 2, 7th and 8th-grade BIL students are introduced to CISCO's "Get connected" course and they are also familiar with making flowcharts using the program FCPro. The fact that students are familiar with logic gates, algorithms and flowcharts are also in line with Chuang et al. (2015) and Piaget's Formal Operational Stage (Flavell, 1963; Piaget, 1958), where students are expected to show the ability to think logically, understand abstract concepts and develop their problem-solving skills. Based on these investigations, the focus of the study was on the following concepts of computational thinking: logic, abstraction and generalisation.

How do we measure computational thinking?

A high-quality assessment system is crucial in education as the data obtained from assessment impacts design, direction and decisions on future educational modifications. Some advantages of having a high-quality assessment of student learning are that it can provide live feedback on schools' academic level and can increase the students' achievements in any given school achievement and as well as giving guidance on curriculum development. However, some criticise that educational systems carry data-driven cultures that focus more on assessment rather than delivering the appropriate content; and over some time, the collected data start to rule your decisions on the amount of data and ways of data collections (Wiliam, 2014). Furthermore, these types of approaches have narrow instructions and low-level thinking strategies instead of higher-order thinking skills, which are essential in problem-solving, critical thinking, creativity (Conley, 2007) and computational thinking (Lai, 2019). The type of assessments and the extent to which these tests and instruments measure students' abilities, and monitor and predict the outcome are all topics of debate. The design of the assessment system and how students are assessed provide a roadmap for teaching and learning instructions. The fact that informatics is not a mandatory subject in the final year examinations at Kazakhstani secondary schools makes the delivery of computational thinking challenging, as informatics is the key subject in enhancing the computational abilities of students, and informatics remains the only subject that has specifically stated section of computational thinking in the national curriculum. If the government and educational bodies, as stated in the updated curriculum, aim to teach students to tackle a variety of tasks by using analysis, abstraction, modelling, and programming; and help students to gain the ability to think logically and algorithmically, see patterns and evaluate their solutions, then the assessment of computational thinking skills should become an educational priority. It is important to have standardised tests as an assessment of the computational thinking skills of students (Linn et al., 2010). The assessment is as important as the integration of computational thinking into curricula. Because without any valid assessment or evaluation, it is unlikely that computational thinking can successfully be integrated into any curricula. Therefore, having narrowed the focus of the research study to three concepts of computational thinking, the assessment and measurement of these concepts by specific instruments should be developed. As discussed in Chapter 2, many studies on computational thinking skills either have small sizes or use self-reported instruments that focus on perception and do not measure performance. If the studies measure performance and skills, they are tool-specific and are mainly based on particular knowledge of programming and coding skills. Computational thinking is not limited to programming or coding skills. When measuring thinking skills, the instrument should not be limited to measuring the programming language oriented thinking or programming oriented thinking but rather programming free (Naveh, 2006). Thus, it was important to develop a non-tool-specific instrument that could measure the computational thinking performance of secondary school students. Therefore, multiple-choice items are used to measure computational thinking performance as multiple-choice questions are considered suitable for the evaluation of higher-order cognitive skills and problem-solving skills (Downing & Haladyna, 2006). As listed in Chapter 2, multiple-choice questions have some advantages, such as applicability to a large audience, practicality in terms of time and money, and also objectivity and compatibility with the item response theory. In addition, multiple-choice questions are the most frequently used format of assessment in education (Downing & Haladyna, 2006; Reynolds et al., 2009).

The first research question addresses how to measure computational thinking performance by using multiple-choice questions. As discussed earlier, these multiple-choice questions are mainly focused on the following computational thinking concepts: logic, abstraction and generalisation. There are 5 quizzes with 10 questions each. Each quiz focuses on one concept of computational thinking. Also, quiz questions are ordered from simple to complex. Each question is constructed in a certain way that each distractor is plausible, which means in case a student answers wrong it still gives information about his/her understanding, not just right or wrong. Simple questions are expected to be easy and complex ones difficult. These difficulty and discrimination parameters of the questions, in short, the quality of the test, are measured using the Item Response Theory. The measurement method by using multiple-choice questions is more convenient for large samples.

As explored in the previous chapters, computational thinking is a cognitive process that can be considered higher-order thinking. Although this study could not cover all aspects of such a complex construct such as computational thinking, it narrowed its focus to concepts of computational thinking and developed bespoke multiple-choice questions to measure it.

Another aspect of this study is the relationship between computational thinking performance and school achievement. The findings of this research study show that there is a particular relationship between the computational thinking performance of Bilim Innovation Lyceums in Kazakhstan and their academic achievement. In the computational thinking performance test, boys outperform girls, but no significant difference was found between them. Intelligence or a factor ‘g’ as a general mental skill is the most effective known predictor of students’ school achievement as measured by IQ tests (Gottfredson, 1998), and the higher the IQ level the higher the computational thinking skills, as computational thinking and intelligence have common characteristics in the definitions related to problem-solving skills and abstract reasoning (Boom et al., 2018). According to McKeachie (1987), the 'g' factor is important in transfer, whether learned skills and knowledge in one area can be applied in another area or domain. The process of transfer depends on several variables (De Corte, 2003; Eraut, 2004). Eraut (2004) lists five stages of transfer and four key variables: the nature of what is being transferred is closely related to the knowledge domain and the functional context dimensions introduced by (Barnett & Ceci, 2002); the difference between the context is the transfer distance described by De Corte (2003); and the time devoted to the transfer process is close to the temporal context by Barnett and Ceci (2002), which refers to the time elapsed between training and applying. The science-subject scores, language-subject scores and the perception of computational thinking are found to be the predictors of the computational thinking performance of secondary school students. If the ‘g’ factor, higher-order thinking skills and academic achievement are all associated, then to what extent computational thinking skills affect the transfer is a question for further research. It would be interesting to carry out longitudinal research which allows discovering some factors in-depth, and it would be possible to see whether girls need more time in training than boys and eventually they reach the same level of computational thinking performance, as confirmed by Atmatzidou and Demetriadis (2015). These findings are important points to draw to the attention of policymakers and educators, as the results of this study shed light on the measurement of computational thinking and also the development of an enriched curriculum that prepares students for the rising demands of the 21st century. While the research findings of the work with the BILs may not be generalizable to other schools, the results can serve as a framework in both measurements of computational thinking and curriculum development with a high standard for thinking and learning. The implications and recommendations of the research are presented in the next chapter.

Chapter 6 Conclusion

6.1 Measurement and relationship

This study contributes to the body of knowledge both about the measurement of computational thinking skills and the relationship between computational thinking and other variables. Developing validated measurement tools on computational thinking and investigation of the relationship between computational thinking skills and various subjects help in designing a curriculum that supports students' thinking skills and contributes to their cognitive development. This study also contributes to the body of knowledge on the measurement of computational thinking performance by using a non-tool specific instrument, a multiple-choice test. Such measurement instruments and the results of the relationships can help educators, teachers, specifically Bilim Innovation Lyceum administration in designing better curricula to support students' cognitive development by adjusting subjects that enhance students' problem-solving skills. The findings of this research provide implications and recommendations that teachers, BIL administration, the education board can benefit in the measurement of thinking skills, assessment system and designing curriculum. The results of this research provide a better understanding of the relationship between computational thinking skills and school achievement of BIL students, highlighting the role of science and language subjects in schools, which are found to be predictors of computational thinking performance. The study also contributes to the question bank and the evaluation of the assessment processes of the Quantum project. The research also provides recommendations for constructing good quality multiple-choice questions in general and specifically to measure computational thinking performance.

Computational thinking is one of the important life-long skills for everyone in the 21st century. The idea of computational thinking in education was promoted by Papert in the 1980s but has been more central to the educational reform agenda over the last fifteen years. Although the transferability of computational thinking skills is a topic of debate, the power of computational thinking is undoubtedly essential educational outcomes. Because higher-order thinking and complex cognitive tasks, such as computational thinking, are hard to explicitly identify or assess, they might be deprioritized in the classroom environment or sometimes may not be provided with appropriate approaches and methods to deliver to students. As we can witness the integration of computational thinking massively by many countries and increasing movement towards computing in education sectors, many schools, institutes and education centres start to raise awareness of its importance. Figure 31 illustrates the research path beginning from the curriculum with integrated computational thinking, through computational thinking elements in a classroom environment, the relationship between computational thinking and school achievement, measuring computational thinking performance and examining the quality of the multiple-choice questions, and finally ending with the recommendations.

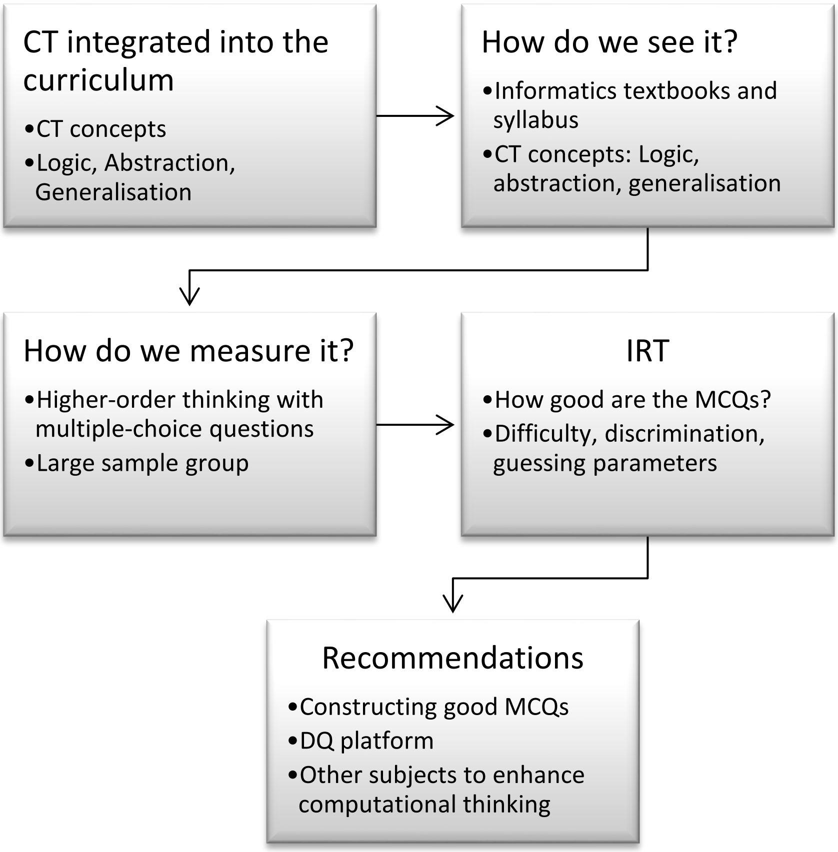


Figure 31 The measurement of computational thinking in this study and the implications

The study investigated the measurement of thinking skills of students at Bilim Innovation Lyceums in Kazakhstan focusing on the computational thinking concepts that students are already familiar with. As a measurement instrument, bespoke multiple-choice questions were constructed. The Item Response Theory models (2PL and 3PL) show that the test items are good enough to measure computational thinking. The results of this research study revealed that the computational thinking performance of 8th grade Bilim Innovation Lyceum students in Kazakhstan is correlated with school achievement. The general school achievement is categorised into three directions: science subjects, language subjects and humanities subjects. Achievement in science and language subjects are found to be the predictors of secondary school students' computational thinking performance, whilst that in the humanities is not. The perception of computational thinking skills can also be a predictor of computational thinking performance. Most of the earlier studies were tool-specific instruments based on specific coding or programming skills, thus they failed to eliminate pre-acquired skills, but, as Naveh (2006) suggests, it was avoided measuring programming language oriented thinking, when measuring thinking skills.

6.2 Limitations of the study

First of all, this study uses single-point primary data, a multiple-choice test, that was taken only once. However, computational thinking is a multi-dimensional construct and it is not easy to measure it even with multiple-point data collection. In general, multiple-choice tests have certain limitations such as lack of assessment and reflection of complex performance and lack of diversity (Gayef et al., 2014; Hancock, 1994; Martinez, 1999; Paxton, 2000; Simkin & Kuechler, 2005). Although multiple-choice tests are criticized for these limitations, in the academic community they are widely accepted as a testing instrument due to their validity and reliability. The quality of the multiple-choice questions used in this study is found as good to measure computational thinking performance. The next limitation is the fact of being context-based, which might give different results, as data collection was only in Kazakhstan; and only included the participants from selective schools, the Bilim Innovation Lyceums (BIL), which means the participants' abilities and performance on average are higher than the general school population in Kazakhstan. The study narrowed its focus to logic, abstraction and generalisation concepts of computational thinking for the 8th-grade level. Therefore, the bespoke instrument used to measure computational thinking performance in this research might not fit other sample groups. Another possible limitation is the language barrier. The participants' first language is not English, but all the participants speak it as their second or third language. The participants' science lessons (algebra, geometry, physics, chemistry, biology and informatics) are conducted in English at BILs. It is also important to note that every single question of the computational thinking performance test was carefully written using simple English, as the test overall was designed both for both native and non-native speakers. One more limitation of this study can be the fact that it cannot demonstrate causality and is limited with the correlational and regression analysis. These listed limitations show areas in which the study can be further developed or modified for future research. Considering all the mentioned limitations of the study, the implications of this research might also vary.

6.3 Implications

It should be noted that this study investigates the relationship between computational thinking and general school achievement and also the perception of computational thinking skills, but not the pedagogical aspects of delivering computational thinking or the cognitive structures developing in the students' mind during the process of computational thinking.

1. **Measurement implications.**

The measurement of thinking skills is an important part of the education system, classroom environment and designing curriculum, especially when a construct such as computational thinking has been integrated. This research firstly narrowed down the focus to three concepts of computational thinking, then constructed bespoke multiple-choice questions for measuring computational thinking performance. The findings of this research study generate discussions about measurement, relationships with subjects and perception of computational thinking. The primary data analysis can also be useful to identify BIL students' learning objectives and support their progress. When measuring the computational thinking skills of students, educators and teachers should keep in mind that enough amount of training time is required to develop a satisfactory level of computational thinking skills (Marina Umaschi Bers et al., 2014; Clark et al., 2016). The results of the study open up discussions with researchers about the measurement of computational thinking performance by using multiple-choice questions. The Item Response Theory models, both 2PL and 3PL show the constructed multiple-choice questions are good enough to measure students' thinking skills. In many research studies, computational thinking skills are measured by instruments that require knowledge of specific tools or programming languages, which raises questions about whether they are measuring computational thinking or knowledge and skills in a specific programming language. A person with some experience in a particular tool or programming language might perform better than the one who is less experienced in that tool, while the computational thinking skills might be better. The multiple-choice questions serve as a non-tool-specific instrument, and at the same time, they can be applied to large sample groups.

2. **Relationship implications.**

Little to no data is available on the relationship between computational thinking and general school achievement of secondary school students, particularly students at Bilim Innovation Lyceums. However, numerous studies with different methods have highlighted a positive correlation between computational thinking skills and intelligence and reasoning abilities of students (Boom et al., 2018; Boucinha et al., 2019; Gouws et al., 2013b). A few findings are available on the relationship between computational thinking skills and the academic achievement of students. As mentioned earlier, in most cases the computational thinking skills are measured by tool-specific instruments, and computational thinking skills are associated with mathematical knowledge and achievement. Chapter 2 discussed the commonality of computational thinking and other thinking types, such as mathematical thinking, logical thinking, engineering thinking, design thinking and creative thinking. The correlational test results of this study show that computational thinking performance is positively correlated with science-subjects achievement and language-subjects achievement. The most correlated science subjects with computational thinking performance are mathematics, physics and chemistry, whereas informatics subject was the least correlated among the science-subjects direction. The lower correlation of informatics and computational thinking performance might be because informatics topics stated in the syllabus covered some specific areas or some knowledge and skills in certain computer applications, where the multiple-choice test was designed to measure thinking skills. Since correlation does not imply causation, no causal inferences can be drawn from this correlation test results of this study. However, the science subjects and language subjects are found to be predictors of the computational thinking performance of the participants. These findings should draw the attention of the stakeholders, educators, school administration and students if such higher-order thinking skills as computational thinking are being integrated and brought into a classroom environment. Computational thinking is not just a matter of informatics courses or computer science classes. Instead, it is a required skill of the 21st century, which needs both a technology-rich environment and a well-supported curriculum. A few studies investigate the relationship between computational thinking and other domains (Durak & Saritepeci, 2017; Román-González, Pérez-González, & Jiménez-Fernández, 2016; Román-González et al., 2018; Yildiz, Yilmaz, & Yilmaz, 2017), which show the importance of a comprehensive approach in fostering computational thinking. This study supports the link between computational thinking and general school achievement or academic progress in certain courses. Computational thinking performance and school achievement are significantly dependent and independent variables, meaning that secondary school students with high school achievement generally have higher computational thinking performance or/and vice versa. Since the science and language subjects scores and the perception of computational thinking skills are the predictors of computational thinking, students with lower performance in computational thinking can be supported by improving those missing areas to help students to become better at tackling different tasks and gain abilities to think computationally (National Academy of Education, 2016). Each Bilim Innovation Lyceums can benefit from the obtained data set as it can help with student-level analysis from a different measurement perspective such as thinking skills, which can support further development of students' computational thinking. Each participant school can benefit from administering the assessment of students' thinking skills and using the research findings to guide individual educational instructions. Computational thinking should not be considered as a stand-alone skill for students to be developed and trained, instead, it should be perceived as a cognitive approach that can utilise digital age capabilities in a proper context to benefit from it.

6.4 Recommendations

Several recommendations for further research can be listed based on the findings of the current study. The results of this research shed light on the measurement of computational thinking performance of secondary school students by using multiple-choice questions and relationships between computational thinking and other variables. This study contributes to the body of knowledge on measurement methods of computational thinking by multiple-choice questions, as there is a lack of studies about measuring computational thinking by using non-tool-specific instruments. The guidelines on constructing good multiple-choice questions and the outcome of the Item Response Theory models give valuable information for teachers as a practical application of assessment. When teachers follow the recommendations on how to construct good quality multiple-choice questions they may get a good assessment tool. However, having that assessment tool, even if a teacher followed all the recommendations on how to construct good quality multiple-choice questions does not guarantee the quality of the questions can meet the expectations. This is why the outcome of the Item Response Theory is crucial for teachers as it is not always easily available to get item analysis individually. There should be a quality-oriented system for each question behind each assessment. The Diagnostic Questions used in this research study (www.diagnosticquestions.com) is a rich platform with lots of questions both for teachers and students that provides better quality questions in many subjects.

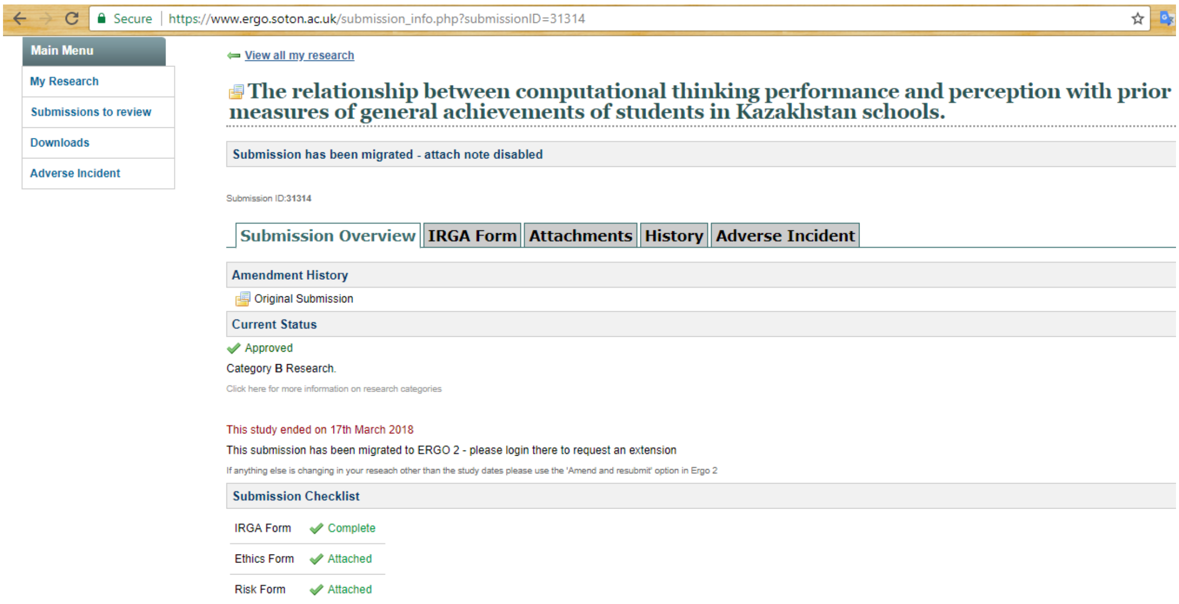
Since this research is quantitative with a large number of participants, subsequent research can follow a longitudinal approach to investigate students' computational thinking skills as related to the variables in this study to explore in greater depth the changes over periods. It is also recommended to replicate this research using the same instruments on a sample of public schools' students in Kazakhstan which could help in gaining a richer picture as those students represent a major percentage of the Kazakhstani student population than Bilim Innovation Lyceums do; and as those public schools have different academic achievement and different level of intelligence. The Bilim Innovation Lyceums in the sample represent the BIL students across Kazakhstan, where the majority of the participants are boys. Although no significant differences were found between boys and girls, language groups or school types in regard to the level of computational thinking performance, girls performed better than boys, and the Kazakh language groups were better than the Russian language groups at school achievement. The ICILS 2018 results provide a larger picture of the computer literacy of 8^{th} graders at the country level, where

Kazakhstan participated for the first time and scored the lowest among 14 participant countries (ICILS, 2020). There is room to grow for Kazakhstan and these findings provide a valuable guide to follow. The results of this study suggest carrying out research further on science and language subjects that could enhance the computational thinking skills of students. BIL administration can make use of these findings with regards to the general school achievement, as the school atmosphere at BILs is the key factor that plays a great role in students' academic achievement (Zhussipbek, 2019). Not only BILs but any school administration should work to align their teaching strategies with the learning outcomes and assessments. Coherent curriculum alignment helps students understand how different subjects fit together to enhance their thinking skills, which in turn helps them learn. These recommendations are for the educators, teachers, school administrators, and stakeholders, the education board and the Ministry of Education and Science of the Republic of Kazakhstan.

If we see computational thinking as a stand-alone special branch of science (Athreya & Mouza, 2017) and accept it as an umbrella term relevant to a subset of related cognitive skills used in computational tasks and activities (Saxena & Basnet, 2017), then where their versatility is shown in practice, a clear roadmap is needed to develop and transfer computational thinking skills, and generate appropriate studies and research strategies to make the best use of these skills. As Wolz et al. (2011) proposed, by respecting the culture of each school, and carefully engaging the school teachers - by infusing (not injecting) curriculum - it is possible to spread confidence in and enthusiasm for computational thinking even among those who do not consider themselves to be good computational thinkers. Schools are building the foundation of future professionals with good computational thinking skills that can be transferred to any professional area and benefit in various settings. To reach this goal, it is important to make thinking visible, embrace both student thinking and expert thinking, where each step of a student's thinking is visible and techniques and strategies are carefully chosen (Bransford et al., 2000). The computer world is diverse, so it is truly hard to fit computational thinking with all the possible human thinking modes, ways and styles, their combinations and products into a restricted definition box. It is more important to comprehend the fact that computational thinking can involve more areas than we might think of, than generating and using certain definitions. The idea behind computational thinking is open to be explored and studied further by neuroscience, cognitive psychology and education, as the process of human thinking is yet to be investigated, whereas the possibilities of computers are growing.

This quantitative study examines the computational thinking performance of 8th-grade students at Bilim Innovation Lyceums (BIL) in Kazakhstan, and the relationship between computational thinking performance and school achievement, as measured by multiple-choice questions and the General Knowledge Tests. The research results provide a general picture of students' computational thinking level, the methods to measure the computational thinking performance, students' perception of computational thinking and the relationship between the academic achievement, which can help school teachers, school administrators, stakeholders and education boards in shaping the lesson guidelines, and also to reflect on student assessment practices, and develop curricula. Briefly, the quality of multiple-choice questions to measure the computational thinking performance, the relationship between computational thinking performance and general school achievement, and perception of computational thinking skills of secondary school students were studied and discussed. The key findings here are the predictors of computational thinking can be the science subjects, language subjects and the perception of computational thinking skills; and when special recommendations are followed and tests using the IRT models are implemented, valid and reliable multiple-choice questions can be constructed to measure higher-order thinking skills.

Appendix A ERGO approval by the University of Southampton


Secure | https://www.ergo.soton.ac.uk/submission_info.php?submissionID=31314

Main Menu
- My Research
- Submissions to review
- Downloads
- Adverse Incident

View all my research

The relationship between computational thinking performance and perception with prior measures of general achievements of students in Kazakhstan schools.

Submission has been migrated - attach note disabled

Submission ID:31314

Submission Overview | IRGA Form | Attachments | History | Adverse Incident

Amendment History

Original Submission

Current Status

Approved

Category B Research.

Click here for more information on research categories

This study ended on 17th March 2018

This submission has been migrated to ERGO 2 - please login there to request an extension

If anything else is changing in your reseach other than the study dates please use the 'Amend and resubmit' option in Ergo 2

Submission Checklist

IRGA Form	Complete
Ethics Form	Attached
Risk Form	Attached

Appendix B Participant Information Sheet

(Online multiple-choice questions and online questionnaire)

Study Title: The relationship between computational thinking performance and perception with prior measures of general achievements of students in Kazakhstan schools.

Researcher: Yerkhan Mindetbay

ERGO number: 31314

Please read this information carefully before deciding to take part in this research. It is up to you to decide whether or not to take part. If you are happy to participate you will need to tick the box below and click start.

What is the research about?

I am a PhD researcher at the University of Southampton. The research I am doing is part of my PhD degree thesis that is about computational thinking skills. By online multiple-choice questions on www.diagnosticquestions.com, computational thinking skills will be measured. This study will look for the relationship between computational thinking level and general performance of 9th-grade secondary school students in Kazakhstan.

Why have I been asked to participate?

The participants of this study are 9th grade Bilim Innovation School (BIL) students in Kazakhstan. As the educational reform is ongoing now in Kazakhstan and BIL experiences are widely shared among all other schools, it is very important to research BIL students. With all your school innovations and achievements, your voluntary participation is highly valuable in this data collection for this research.

What will happen to me if I take part?

If you agree to participate in this research project, you will have a questionnaire with 22 questions. Then your teachers will arrange 100 minutes long online multiple-choice quiz with 50 questions. Both the multiple-choice quiz and the questionnaire will be taken once only. Your Performance Assessment Test (PAT) result and Admission Test results will be used as secondary data in this study.

Are there any benefits in my taking part?

These online multiple-choice questions might be interesting and challenging for you. All the questions are in-line with your current curriculum and an annual

program. However, you will not get any marks/points for this test. Regardless of what result you will get you will not be punished or marked for that. This test might give you positive insights into your logical, algorithmic and finally computational thinking. The result of this large research will benefit teachers, the Kazakhstan government and researchers with its findings.

Are there any risks involved?

There are no real risks to being involved in this research.

Will my participation be confidential?

All the information obtained in this research will be kept strictly confidential, on a password-protected computer and will not be available publicly. Every participant has a unique ID number, so your names will not be published anywhere. The only access to the collected data will have the researcher and his supervisors. Data will be kept safe in line with the UK Data Protection Act and University of Southampton policy.

What should I do if I want to take part?

If you want to take part, let your class teacher or CS teacher know about it. If you are ready to participate, tick the box below and click start to begin the questionnaire.

What happens if I change my mind?

Participation in this research project is voluntary. You can withdraw before beginning, during or after completing the online questionnaire and multiple-choice questions. If you change your mind and want to withdraw your data after the questionnaire and test, you can let me or your class teachers know about it within 24 hours after completing the test. My contact details are provided below. I will be in contact with your class teachers during the data collection period.

What will happen to the results of the research?

The result of this study will be published in my PhD thesis. If you are interested in the results of this research, you can let me or class teachers know and I can share the findings later on.

Where can I get more information?

If you have any questions about this research, you can contact me:

Yerkhan Mindetbay

Southampton Education School

University of Southampton

Email: Y.A.Mindetbay@soton.ac.uk or yerkhan@gmail.com

Phone: Kazakhstan: +7 707 7183332 UK: +44 7514 158561

Supervisor: Dr Christian Bokhove C.Bokhove@soton.ac.uk

What happens if something goes wrong?

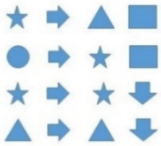
If you have any concerns or complaints about this study, you may contact:

Research Integrity and Governance Manager, rginfo@soton.ac.uk, phone +44 (0) 238 595058

Thank you for your time and cooperation.

Appendix C Multiple-choice questions sample (Pattern)

VIVA
FIFA
LIFE
FIVE



Each letter is represented by a certain figure.
Identify the letters that are represented by the combination

A

VLIAF

B

FLIAL

C

VLIFA

D

VLIAV

₸
$
£

a
b
c

Each symbol is represented by a certain number. Identify the correct combination of numbers to represent

£ ₸ ₸
₸ $ ₸
$ ₸ £

A

a a c
b c b
c a c

B

c a a
a c a
b a c

C

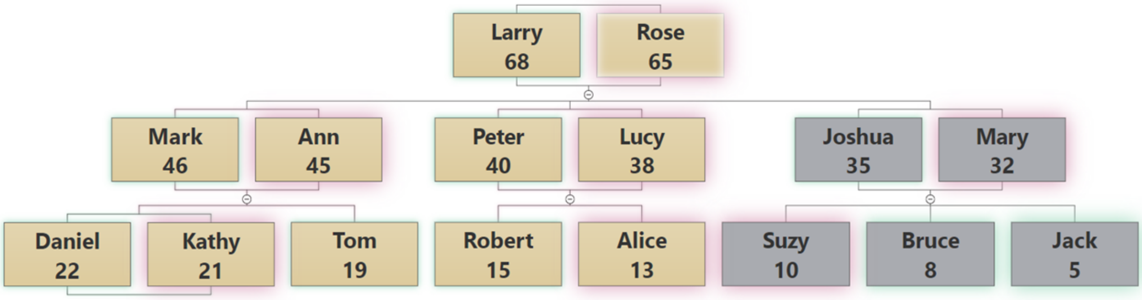
c a a
a b a
b a c

D

c a a
a b a
c a b

Appendix D Multiple-choice questions sample (Logic)

Family Tree



Identify all persons whose names contain both “e” and “i” or who are between 23 and 34 years old.

A Daniel, Alice, Mary

B Mary, Claire, Daniel, Alice

C Daniel, Kathy, Alice, Mary

D Alice, Mary, Peter, Daniel

Appendix E Evaluation studies on computational thinking

	Title	Participants	Environment/Tools	Authors/Year	Strengths	Limitations
1	The Fairy Performance Assessment by Carnegie Mellon Centre for Computational Thinking	10-14-year-old pupils	Programming environments Alice2, Storytelling Alice; and Webb, AgentCubes.	Werner, Denner, Campe, 2012	Engaging projects for children Data collected through 2 years	Not a validated assessment
2	Introducing Computational Thinking in Education Courses	The elementary and secondary education major students	Multiple-choice questionnaire and open-ended questions before and after the special CT course.	Yadav, Zhou, Mayfield, Hambrusch, Korb, 2011	Narrowed target topics	Tool-specific
3	Robotics and Engineering for Middle and High School Students to Develop Computational Thinking	Ten 13-year-old students	Exploratory and descriptive research was carried out during a week-long Robotics and Engineering workshop. Pre-post interviews were taken to measure the elements and dimensions of computational thinking verbally expressed by participants.	Grover 2011	A detailed explanation of the work done by children.	Limited sample size
4	New frameworks for studying and assessing the development of computational thinking	Children from Scratch community.	Scratch – a programming environment where children can create their own interactive stories, games, and simulations.	Brennan, Resnick 2012	Engaging, interactive	Tool-specific
5	Computer games created by middle school girls	59 middle school girls	Stagecast Creator	Denner, Werner, Ortiz 2012	Detailed instruction to students on how to program	Small sample size, only one programming environment
6	REACT (Real Time Evaluation and Assessment of Computational Thinking)	108 middle school students.	Programming tools	Koh, Basawapatna, Nickerson, Repenning,	Real-time assessment system with predictive	Programming focused rather than general concepts.

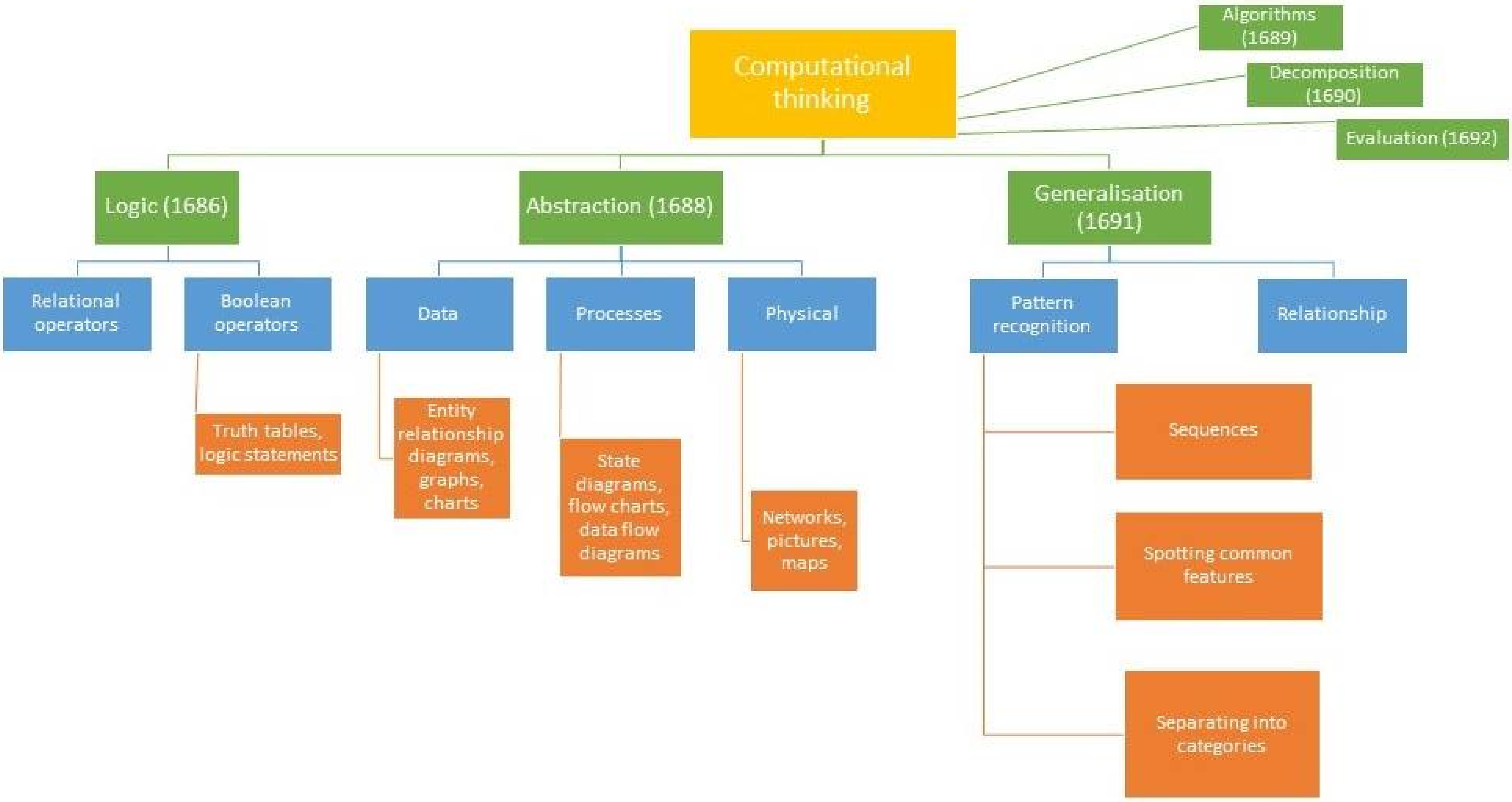
				2012	formative assessment.	
7	App Inventor for Android	7 students aged 13.	App Inventor, programming environment	Grover, Pea 2013	Engaging programming environment with more complex computational thinking concepts	Small sample size.
8	First-Year Student Performance in a Test for Computational Thinking	First-year students at the Rhodes University in South Africa	Multiple Choice Questions	Gouws, Bradshaw, Wentworth 2013	The six distinct classifications of CT skills and practices are introduced.	Applied in one institution only. The same test is planned for the second time.
9	Computational thinking in educational activities		Light-Bot, an educational Flash game developed by Armor Games	Gouws, Bradshaw, Wentworth 2013	Computational thinking framework	
10	Ethnocomputing: a mental model approach to the design and assessment of computational thinking in a culture-based learning environment	Middle School pupils	Kente Cloth, a programmable Culturally Situated Design Tool	Babbitt 2013	Cultural content approach	Region-specific assessment of computational thinking
11	Exploring Core Cognitive Skills of Computational Thinking	12 freshmen students at the University of Minho, Portugal	• general intelligence • spatial reasoning • mathematical reasoning • attention to details	Ambrosio, Almeida, Franco, Macedo, 2014	Validated cognitive tests. Correlation with academic success.	Small sample size.
12	“Systems of Assessments” for Deeper Learning of Computational Thinking in K-12	28 7th and 8th-grade students, in North Carolina, USA	Foundations for Advancing Computational Thinking (FACT) CS course	Grover 2015	A multidimensional approach to the solution	Small sample size

			Pre post-tests Test for transfer of skills on algorithmic thinking on text-based programming language			
13	Automated Personalised Assessment of Computational Thinking MOOC Assignments		Online smart assessment system MindReader to automatically assess programme codes	Jamil 2017	Multi-language supported fully automated assessment system.	Tool-specific
14	Mixed Methods for the Assessment and Incorporation of Computational Thinking in K-12 and Higher Education	4-6 grade pupils	Blockly data and students programming behaviours are qualitatively analysed	2016	Mixed method	Tool-specific
15	Comparing Computational Thinking Development Assessment Scores with Software Complexity Metrics	Different projects from the Scratch repository	Halstead's and McCabe's Cyclomatic Complexity	Moreno-Leon, Robles, Roman-Gonzalez. 2016	Narrowed focus	Self-reported, lack of validation
16	Comparing Students' Scratch Skills with Their Computational Thinking Skills in Terms of Different Variables	31 5th grade students in Kastamonu, Turkey	Dr. Scratch and computational thinking level scale survey	Oluk, Korkmaz 2016	Various concepts covered	Self-reported, tool specific
17	Analysis of the relation between computational thinking skills and various variables with the structural equation model	156 students from 5-12 grades in Ankara, Turkey	Personal Information Form. Computational Thinking Skills Scale questionnaire. A relational screening model was used.	Durak, Saritepeci. 2017	Various concepts covered	Self–reported, perception.

18	Does computational thinking correlate with personality?	99 5-10 grade school students in Spain	Computational thinking multiple-choice test. Big Five Questionnaire for Children	Roman-Gonzalez, Perez-Gonzalez, Moreno-Leon, Robles. 2016	Non-cognitive sides of computational thinking	Self-reported
19	Which cognitive abilities underlie computational thinking? Criterion validity of the Computational Thinking Test	1,251 Spanish students from 5th to 10th grade	Computational Thinking Test Primary Mental Abilities test(PMA) RP30 problem-solving test	Roman-Gonzalez, Perez-Gonzalez, Jimenez-Fernandez. 2016	Large-scale multi-dimensional	
20	A validity and reliability study of the computational thinking scales	1306 university students in Turkey	Computational thinking scales survey	Korkmaz, Çakir, Özden. 2017	Large-scale, different concepts of computational thinking	Self-reported
21	Mobile Computational Thinking Test	All major university students	MIT App Inventor	2015	Deeper and in detail testing	Restricted with mobile apps design
22	North Kentucky University CT test on Informatics perspectives		California Critical Thinking Skills Test by Insight Assessment		California Critical Thinking Skills Test	The same questions for pre-test and post-test.
23	UNED (Spain) CT 28 item MCQ test, based on	8th-grade pupils	Multiple Choice Questions	Marcos Román González, 2017	Good quality multiple-choice test	Follows a subject-specific approach, such as Scratch.
24	Diagnostic Questions (project Quantum)	High school pupils	Computer-based MCQ	2016	Automated assessment. Quality diagnostic questions.	MCQ may not reflect as multiple level assessment
25	ICILS 2018 (with computational thinking assessment)	8th-grade pupils	Computer-based multiple-choice questions with drag-and-drop	Expected 2018	Authentic, large-scale and international test.	
26	Bebras Challenge	All level pupils	Computer-based MCQ with drag-and-drop	Since 2005	Authentic and automated assessment of	

					computational thinking	
27	Principled Assessment of Computational Thinking (PACT) by SRI International	High school students	Authentic questions (free-response and multiple-choice questions)	2015	Authentic tests	
28	A nationwide exam	High school students	Multiple-choice questions	Zur-Bargury 2013	Large scale study. Measure learning and reflect on the curriculum.	
29	A comprehensive assessment of secondary school students' computational thinking skills	Secondary school students	Computational Thinking Test (CTt) by González, 2015 and Computational thinking levels scale (CTLS) by Korkmaz et al., (2015)	Elif Polat, Sinan Hopcan, Burak Sisman 2021	Combination of MCQ test and questionnaire with detailed analysis for large sample size	
30	BCTt: Beginners Computational Thinking Test	Secondary school students	MCQs with sequences, loops and conditionals	María Zapata Cáceres, Estefanía Martín-Barroso, Juan Carlos and Marcos Román-González, 2020	Non-programming specific, validated and reliable instrument.	

Appendix F Computational thinking in computing taxonomy by Diagnostic Questions



Appendix G CTP, CTS and GKT scores for each school

School		CTP	CTS	GKT
1	Mean	16.2	74.9	11.6
	N	33.0	33.0	33.0
	SD	5.2	7.2	2.6
2	Mean	15.9	77.2	12.7
	N	24.0	24.0	24.0
	SD	5.5	8.1	1.9
3	Mean	12.4	72.2	10.3
	N	42.0	42.0	42.0
	SD	5.3	8.3	1.8
4	Mean	14.8	75.2	11.7
	N	16.0	16.0	16.0
	SD	3.9	6.8	2.4
5	Mean	15.1	71.8	12.7
	N	31.0	31.0	31.0
	SD	3.4	25.4	2.0
6	Mean	**20.9**	**78.1**	**12.3**
	N	48.0	48.0	48.0
	SD	6.2	9.0	1.8
7	Mean	14.7	71.2	12.6
	N	42.0	42.0	42.0
	SD	4.2	11.5	2.3
8	Mean	17.4	74.5	10.4
	N	19.0	19.0	19.0
	SD	5.4	10.8	1.8
9	Mean	16.2	74.0	13.3
	N	29.0	29.0	29.0
	SD	3.9	8.4	2.1
10	Mean	12.9	71.6	11.4
	N	23.0	23.0	23.0
	SD	3.3	8.3	1.8
11	Mean	**11.8**	**83.8**	**10.7**

	N	20.0	20.0	20.0
	SD	2.9	11.0	2.3
12	Mean	15.1	80.8	12.7
	N	43.0	43.0	43.0
	SD	4.6	13.0	2.2
13	Mean	14.3	74.3	9.7
	N	39.0	39.0	39.0
	SD	5.6	13.2	1.9
14	Mean	**9.4**	**73.0**	**10.3**
	N	10.0	10.0	10.0
	SD	1.6	10.2	1.6
15	Mean	14.7	77.4	10.2
	N	49.0	49.0	49.0
	SD	5.3	7.4	1.8
16	Mean	15.8	76.9	12.8
	N	13.0	13.0	13.0
	SD	3.9	8.5	1.1
17	Mean	16.6	71.3	10.9
	N	45.0	45.0	45.0
	SD	5.2	10.3	2.1
18	Mean	14.2	71.4	13.2
	N	26.0	26.0	26.0
	SD	4.6	12.2	2.3
19	Mean	17.2	74.3	11.0
	N	38.0	38.0	38.0
	SD	5.1	11.0	1.8
20	Mean	10.3	74.0	11.3
	N	29.0	29.0	29.0
	SD	3.3	12.5	1.6
21	Mean	12.8	76.0	11.6
	N	24.0	24.0	24.0
	SD	3.4	18.8	2.3
22	Mean	13.4	72.0	11.7
	N	16.0	16.0	16.0
	SD	4.0	8.5	2.0

23	Mean	9.6	81.3	9.0
	N	6.0	6.0	6.0
	SD	2.9	8.7	1.0
24	Mean	15.0	73.0	8.3
	N	6.0	6.0	6.0
	SD	5.7	7.7	1.4
25	Mean	**13.5**	**71.4**	**13.4**
	N	**52.0**	**52.0**	**52.0**
	SD	4.0	11.0	2.3
26	Mean	14.5	69.5	10.1
	N	43.0	43.0	43.0
	SD	5.3	18.5	1.8
27	Mean	13.4	79.3	11.5
	N	**4.0**	**4.0**	**4.0**
	SD	3.5	2.2	1.5
28	Mean	**10.5**	**69.2**	**7.1**
	N	5.0	5.0	5.0
	SD	3.1	10.8	0.3
TOTAL	Mean	14.8	74.3	11.5
	N	775.0	775.0	775.0
	SD	5.2	12.4	2.3

Appendix H ANOVA test results for two IRT models

```
anova(2PL,3PL)
```

```
                    AIC     BIC  log.Lik   LRT df p.value
Abstraction_2PL   9583.76 9676.74 -4771.88
Abstraction_3PL 9566.15 9705.62 -4753.08 37.61 10  <0.001
```

```
            AIC     BIC  log.Lik   LRT df p.value
Decode_2PL 8816.56 8909.52 -4388.28
Decode_3PL 8768.29 8907.72 -4354.14 68.28 10  <0.001
```

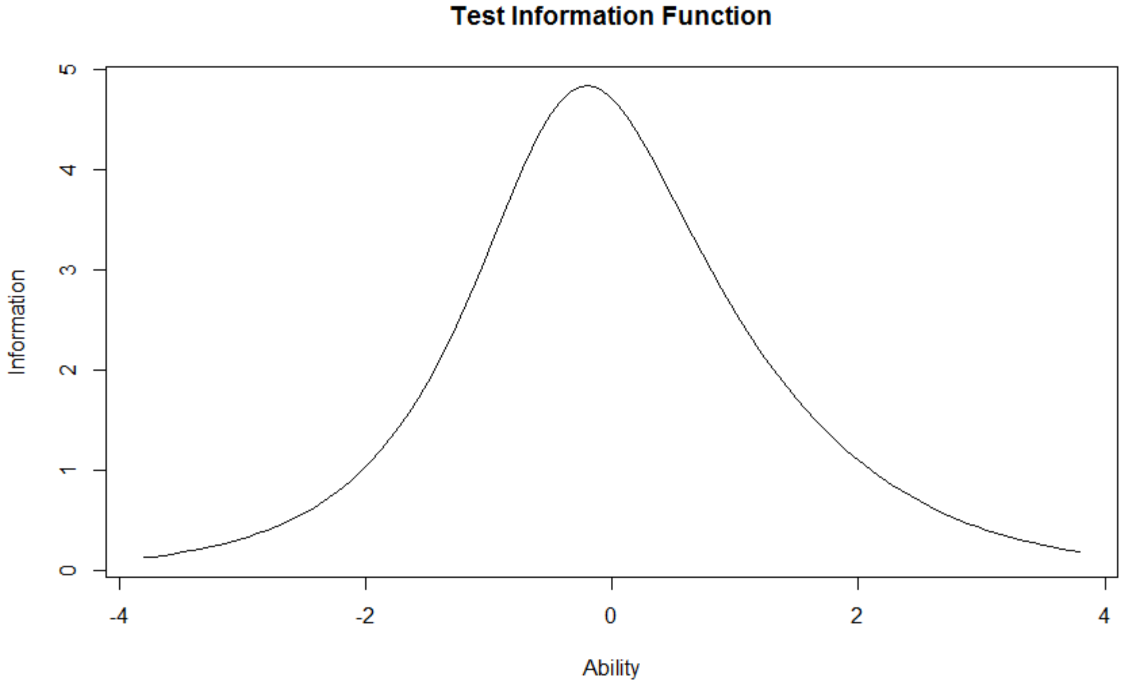
```
            AIC     BIC  log.Lik    LRT df p.value
LogicNarrative_2PL 9235.00 9327.98 -4597.50
LogicNarrative_3PL 9087.95 9227.42 -4513.98 167.04 10  <0.001
```

```
            AIC     BIC  log.Lik    LRT df p.value
LogicNumbers_2PL 8562.44 8654.44 -4261.22
LogicNumbers3PL  8471.58 8609.57 -4205.79 110.86 10  <0.001
```

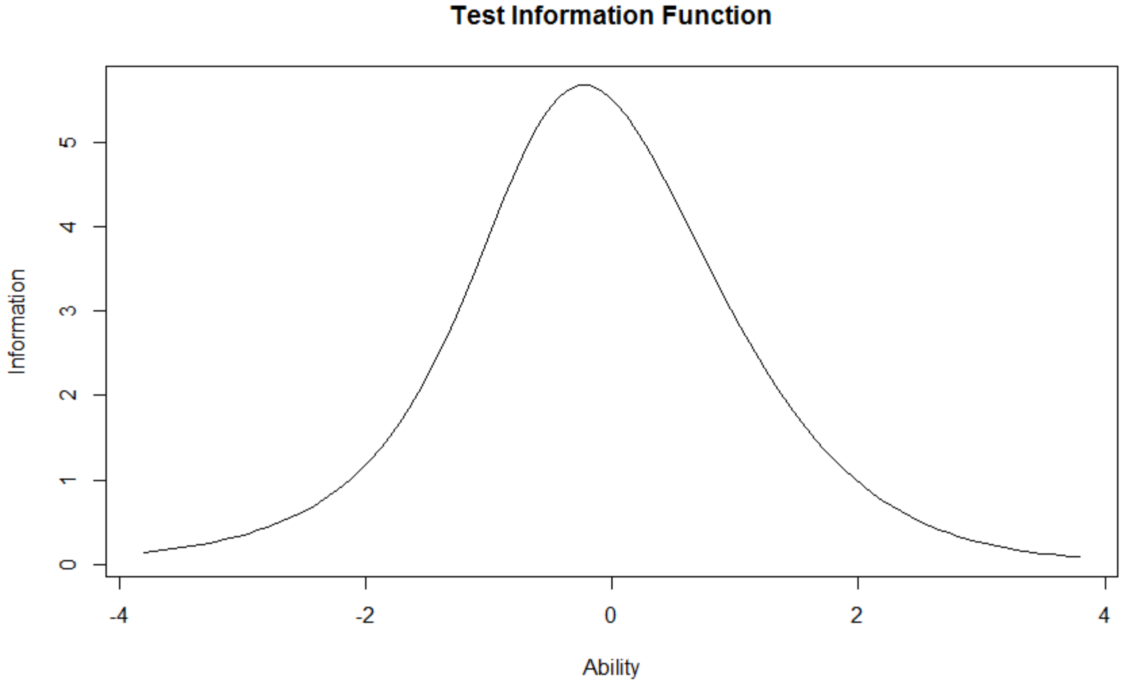
```
            AIC     BIC  log.Lik    LRT df p.value
Pattern_2PL 9443.35 9536.33 -4701.67
Pattern_3PL 9341.36 9480.83 -4640.68 121.99 10  <0.001
```

Appendix I Test Information Functions for 2PL model

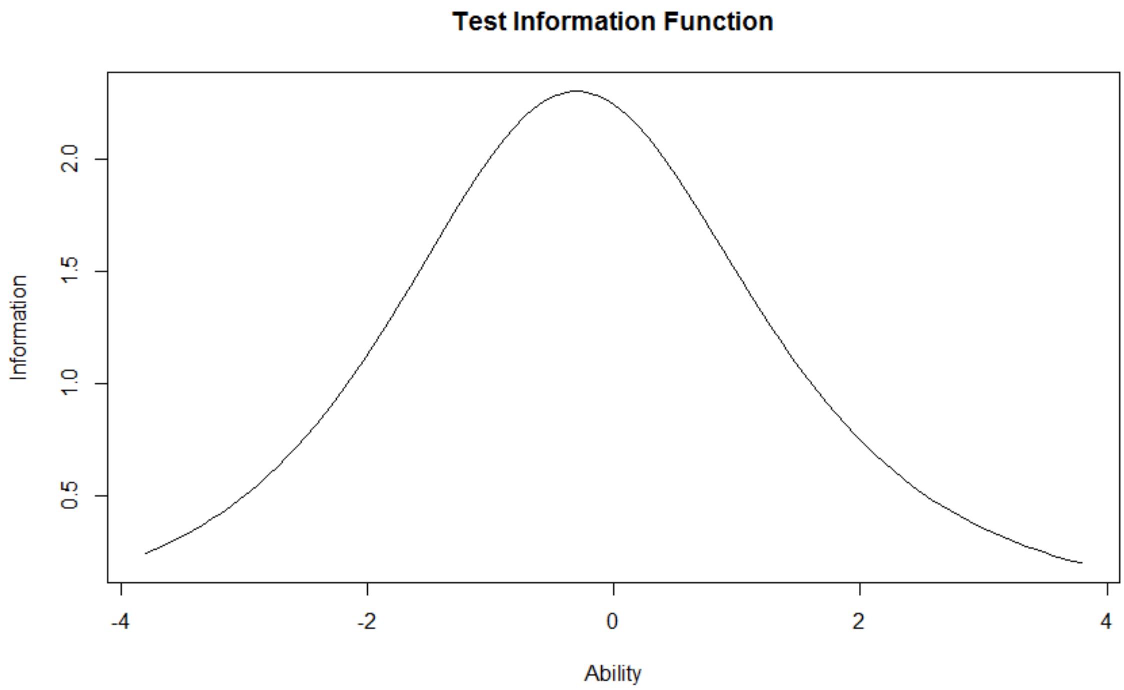
Test Information Function for Quiz1- Logic narrative (2PL).



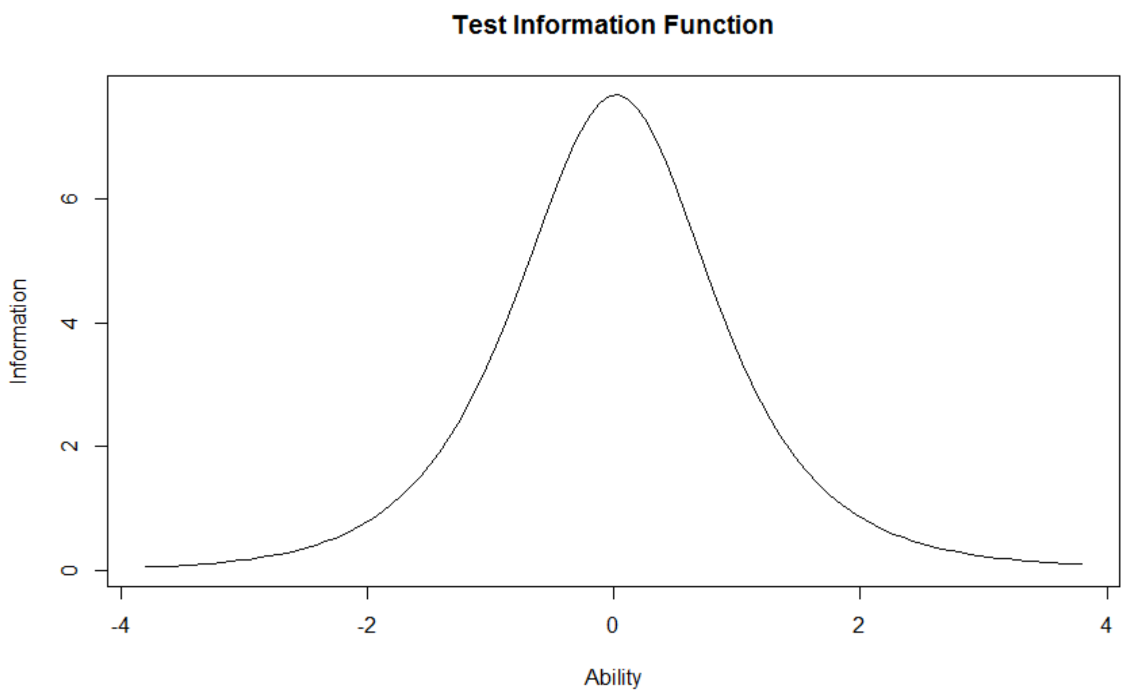
Test Information Function for Quiz2- Logic numbers (2PL).



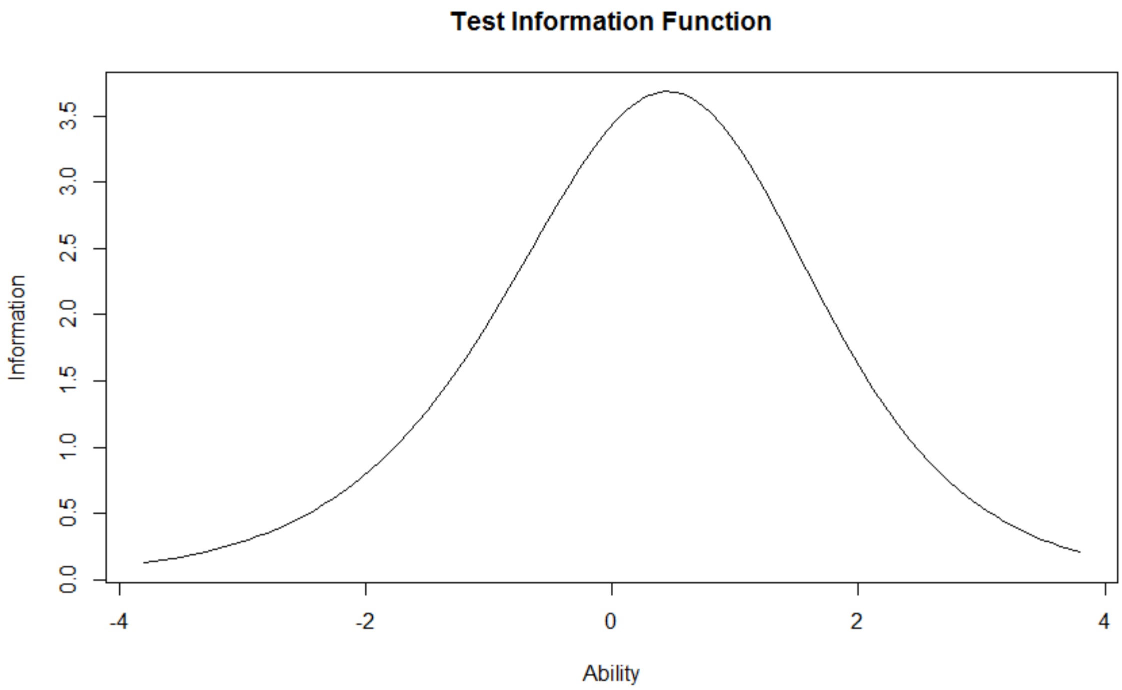
Test Information Function for Quiz3- Abstraction (2PL).



Test Information Function for Quiz4- Decode (2PL).

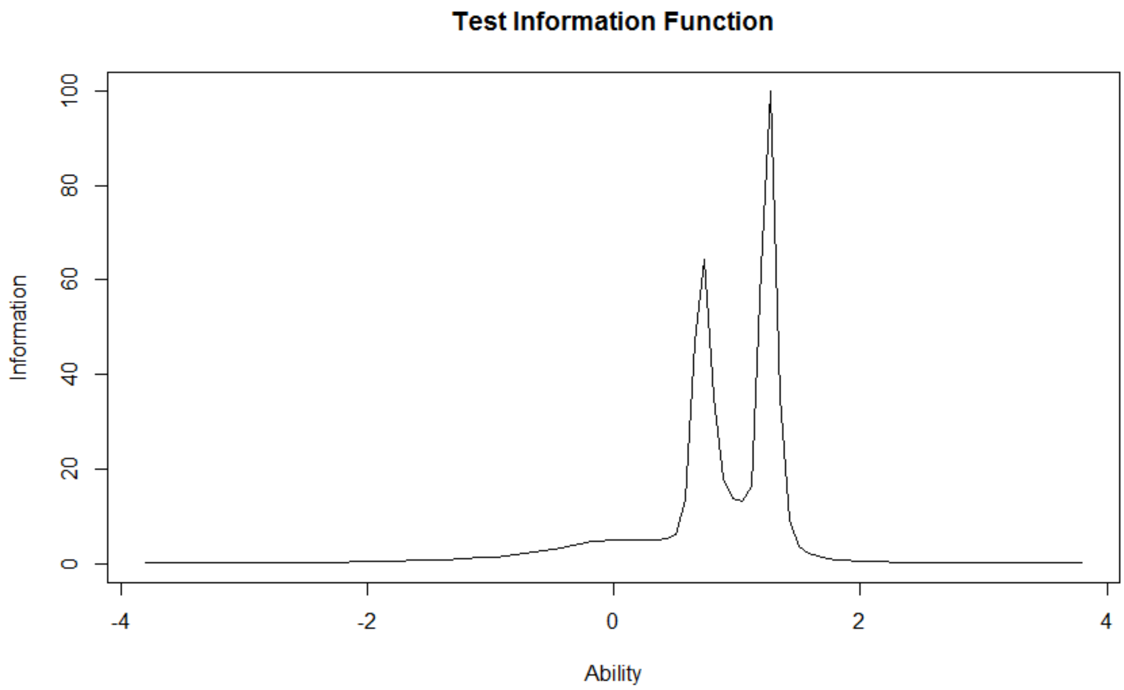


Test Information Function for Quiz5- Pattern (2PL).

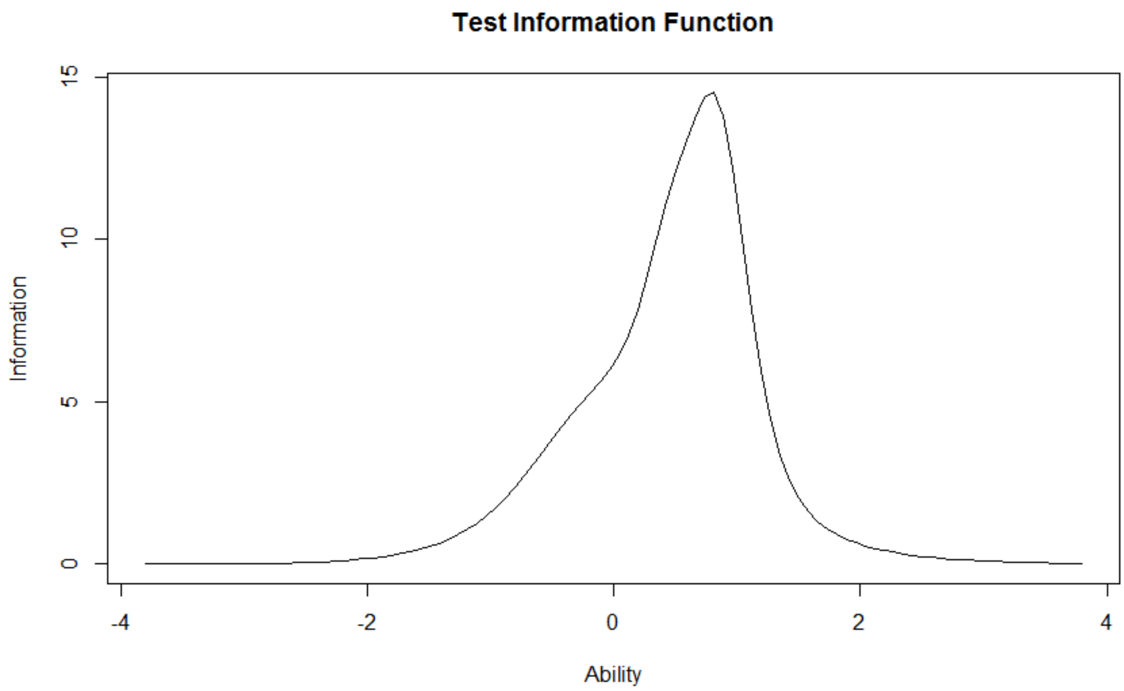


Appendix J Test Information Functions for 3PL model

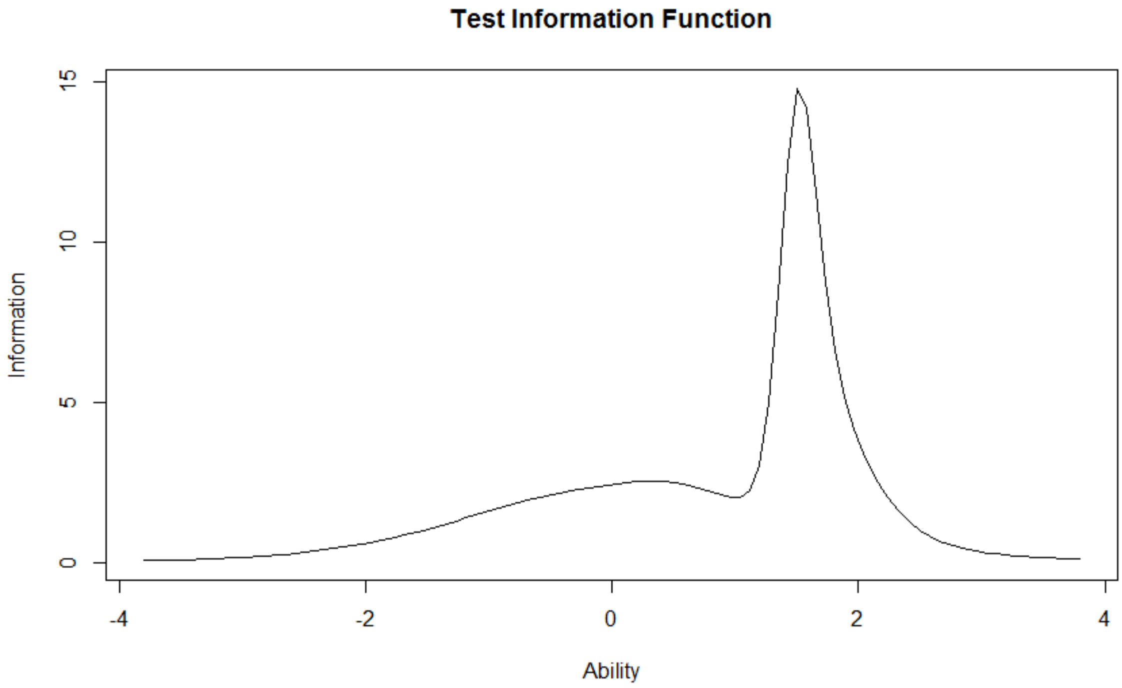
Test Information Function for Quiz1- Logic narrative (3PL).



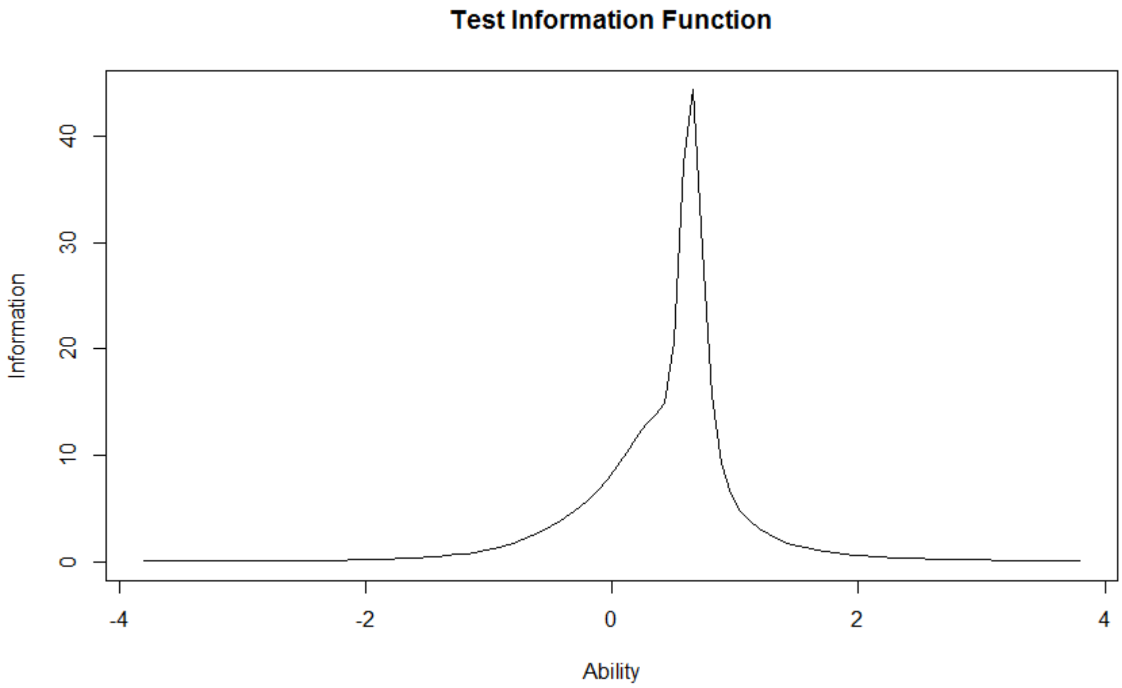
Test Information Function for Quiz2- Logic numbers (3PL).



Test Information Function for Quiz3- Abstraction (3PL).

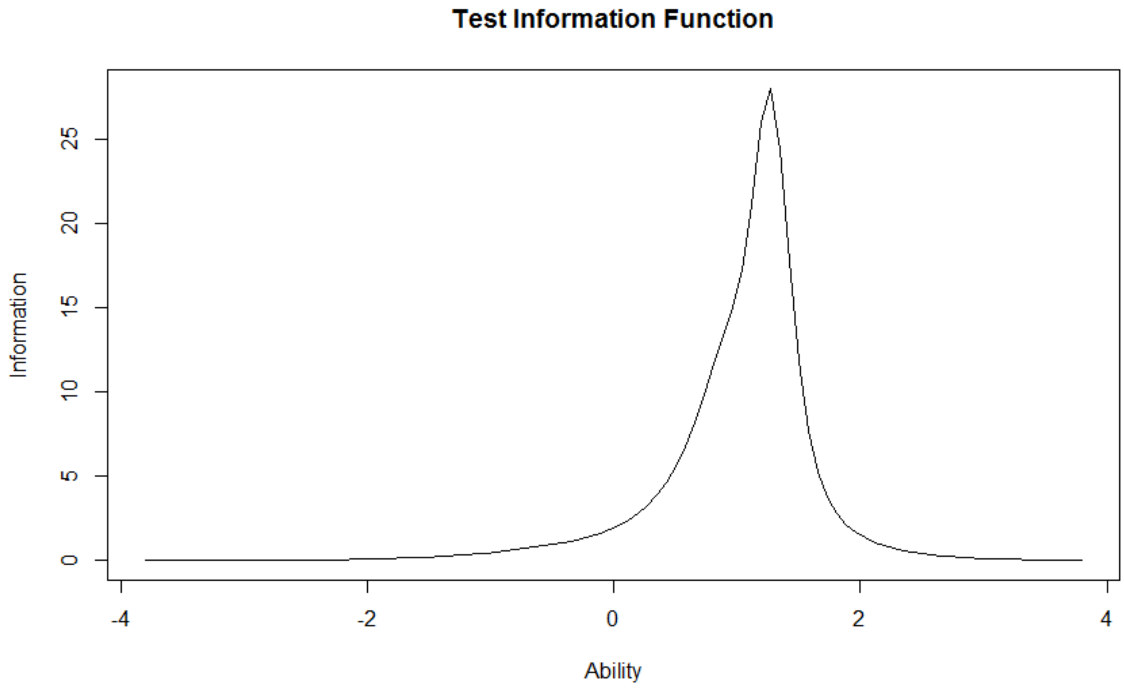


Test Information Function for Quiz4- Decode (3PL).

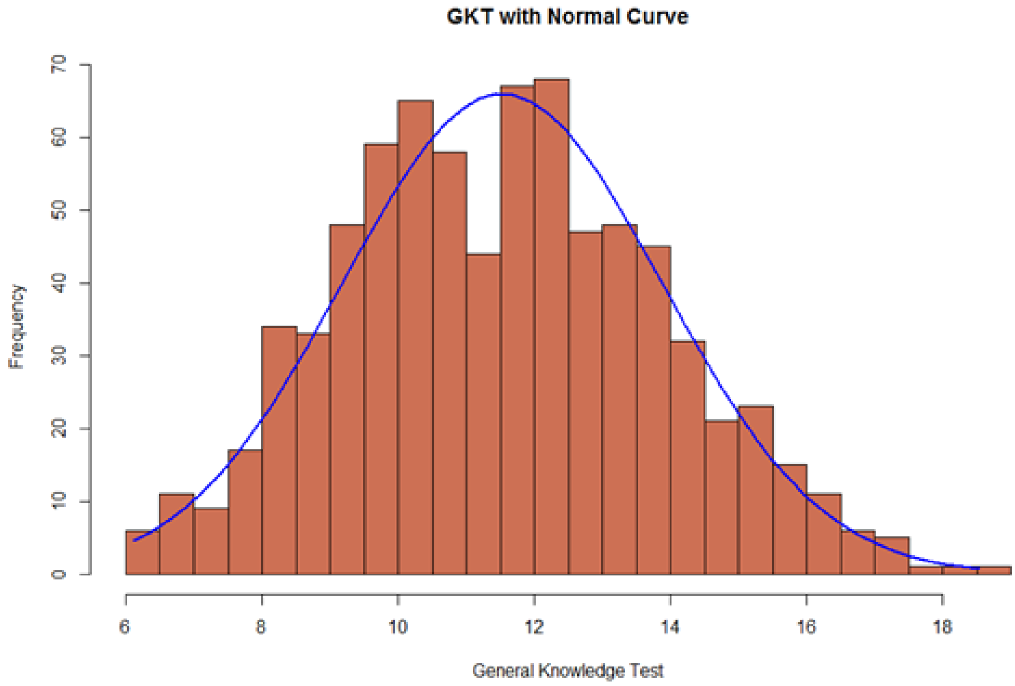


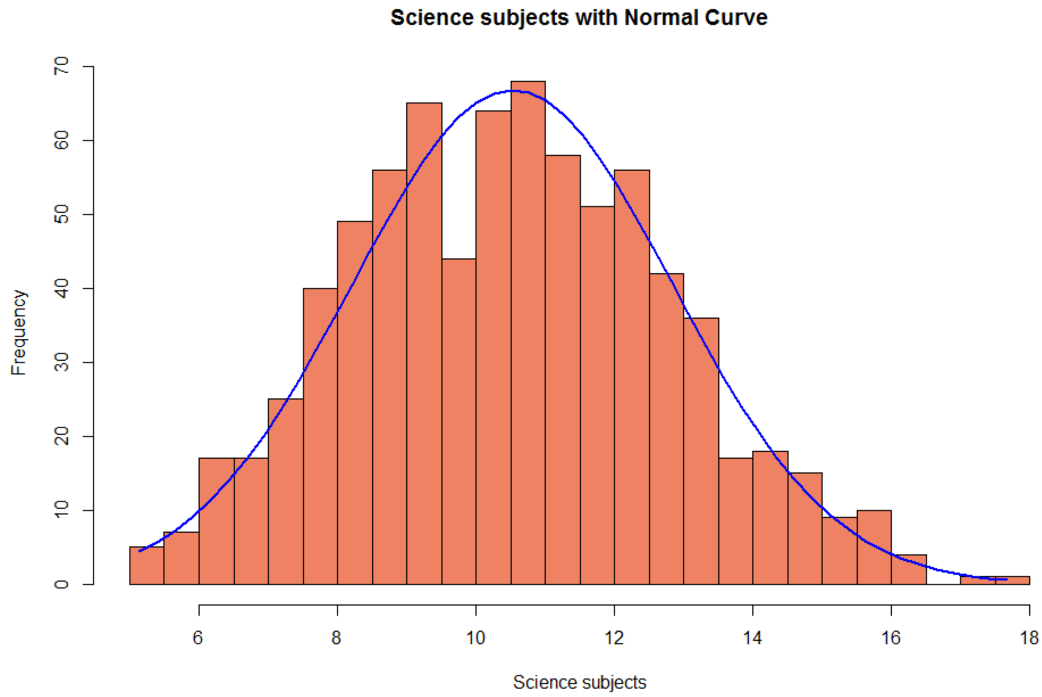
Appendix J

Test Information Function for Quiz5- Pattern (3PL).

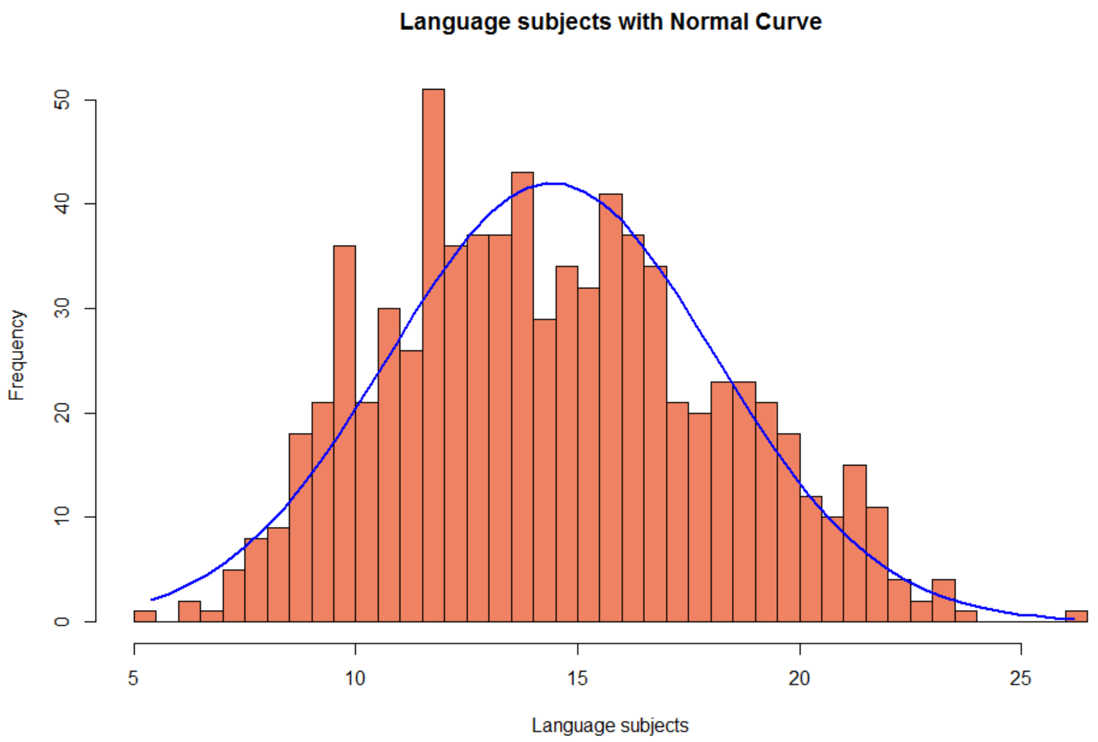


Appendix K Normal curves

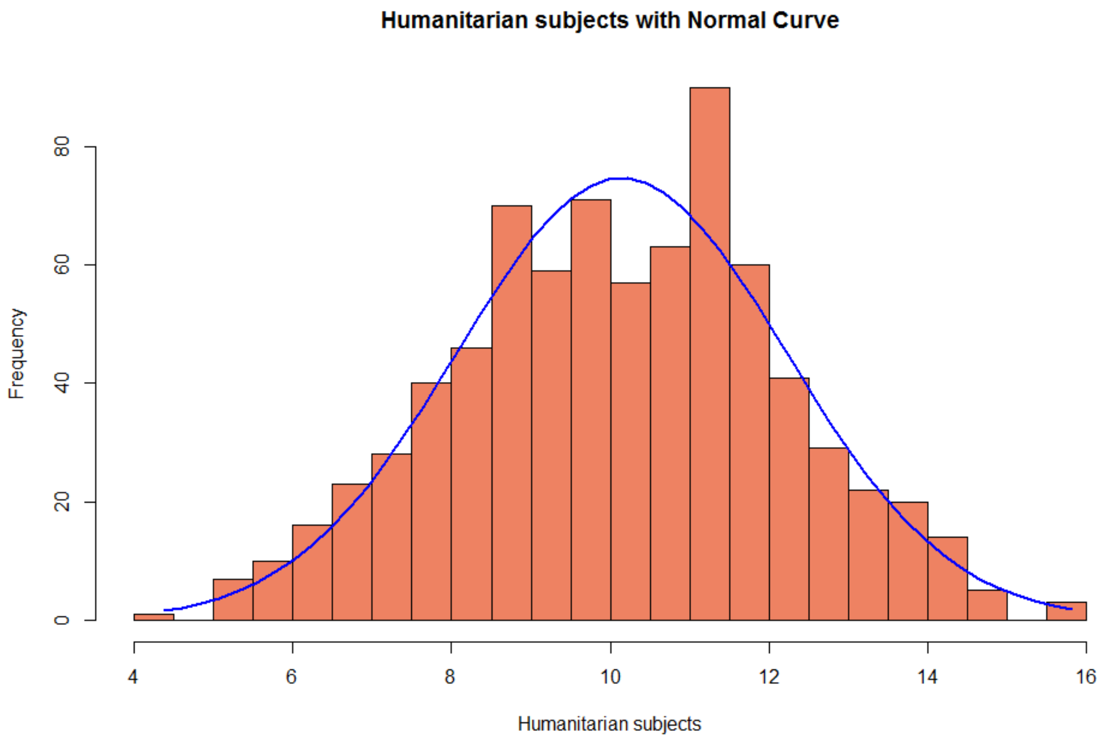




Appendix K



Language subjects with Normal Curve
Frequency
0
10
20
30
40
50
5
10
15
20
25
Language subjects



Humanitarian subjects with Normal Curve
Frequency
0
20
40
60
80
4
6
8
10
12
14
16
Humanitarian subjects

List of References

Abylkasymova, A. E., Nurmukhamedova, Z. M., Nurbaeva, D. M., & Zhumalieva, L. D. (2016). "The Turkish Vector" Influence on Teaching the Exact Disciplines in Modern Educational System of Kazakhstan: on the Example of Teaching Algebra and Mathematics. *Global Journal of Pure and Applied Mathematics*, *12*(4), 973–1768.

ACM, Code.org, CSTA, Cyber Innovation Center, & National Math and Science Initiative. (2016). *K-12 Computer Science Framework*. 297. Retrieved from https://dl.acm.org/citation.cfm?id=3079760

Agarwal, P. K. (2019). Retrieval practice & bloom's taxonomy: Do students need fact knowledge before higher order learning? *Journal of Educational Psychology*, *111*(2), 189–209. https://doi.org/10.1037/edu0000282

Aho, A. V. (2012). Computation and computational thinking. *Computer Journal*, *55*(7), 833–835. https://doi.org/10.1093/comjnl/bxs074

Allsop, Y. (2019). Assessing computational thinking process using a multiple evaluation approach. *International Journal of Child-Computer Interaction*, *19*, 30–55. https://doi.org/10.1016/j.ijcci.2018.10.004

Ambrosio, A. P., Almeida, L. da S., Franco, A., & Macedo, J. (2014). Exploring Core Cognitive Skills of Computational Thinking. *PPIG 2014 - 25th Annual Workshop*. Sussex: University of Sussex.

Ambrosio, A. P., Xavier, C., & Georges, F. (2015). Digital ink for cognitive assessment of computational thinking. *Proceedings - Frontiers in Education Conference, FIE*, (February).

American Educational Research Association., American Psychological Association., National Council on Measurement in Education., & Joint Committee on Standards for Educational and Psychological Testing (U.S.). (1999). *Standards for educational and psychological testing*. American Educational Research Association.

Anderson, J. (1993). *Rules of the mind*. New Jersey: Lawrence Erlbaum Associates.

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Pintrich, P. R., Raths, J., & Wittrock, M. C. (2001). *A Taxonomy for Learning, Teaching, and Assessing. A Revision of Bloom's Taxonomy of Educational Objectives* (Abridged E). London: Addison Wesley Longman, Inc.

Anderson, R., Thier, M., & Pitts, C. (2017). Interpersonal and intrapersonal skill assessment alternatives: Self-reports, situational-judgment tests, and discrete-choice experiments. *Learning and Individual Differences*, *53*, 47–60. https://doi.org/10.1016/j.lindif.2016.10.017

Andrews-Hanna, J. R. (2012). The Brain's Default Network and Its Adaptive Role in Internal Mentation. *The Neuroscientist*, *18*(3), 251–270.

Ater-Kranov, A., Bryant, R., Orr, G., Wallace, S., & Zhang, M. (2010). Developing a Community Definition and Teaching Modules for Computational Thinking: Accomplishments and Challenges. *SIGITE '10: Proceedings of the 2010 ACM Conference on Information Technology Education*, 143–148. https://doi.org/10.1145/1867651.1867689

Athreya, B. H., & Mouza, C. (2017). *Thinking Skills for the Digital Generation. The Development of Thinking and Learning in the Age of Information.* Philadelphia: Springer.

Atmatzidou, S., & Demetriadis, S. (2015). Advancing students' computational thinking skills through educational robotics: A study on age and gender relevant differences. *Robotics and Autonomous Systems*, *75*, 661–670. https://doi.org/10.1016/j.robot.2015.10.008

Babbie, E. R. (1998). *The practice of social research*. Belmont CA: Wadsworth Pub. Co.

Babbitt, W. (2013). *Ethnocomputing: a mental model approach to the design and assessment of computational thinking in a culture-based learning environment*. Rensselaer Polytechnic Institute.

Barcelos, T. S., Munoz, R., Villarroel, R., Merino, E., & Silveira, I. F. (2018). Mathematics learning through computational thinking activities: A systematic literature review. *Journal of Universal Computer Science*, *24*(7), 815–845.

Barefoot. (2014). *Computational thinking*. Retrieved from http://barefootcas.org.uk/barefoot-primary-computing-resources/concepts/computational-thinking/

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, *128*(4), 612–637. https://doi.org/10.1037/0033-2909.128.4.612

Baron, J. (1993). Why Teach Thinking? An Essay. *Applied Psychology: An International Review*, *42*(3), 191–237.

Basawapatna, A., Han Koh, K., Repenning, A., Webb, D. C., & Sekeres Marshall, K. (2011). Recognizing Computational Thinking Patterns. *SIGCSE'11*.

BBC. (2021). Introduction to computational thinking. Retrieved April 4, 2021, from https://www.bbc.com/education/guides/zp92mp3/revision

BCS - The Chartered Institute for IT. (2017). *Network of Teaching Excellence in Computer Science DfE Project Report for year 2 Q2*.

Bebras Community. *Statues*. , Pub. L. No. RC 3, 1 (2015).

Becker, W. E., & Johnston, C. G. (1999). The Relationship between Multiple Choice and Essay Response Questions in Assessing Economics Understanding. *Economic Record*, *75*(231), 348–357. https://doi.org/10.1111/j.1475-4932.1999.tb02571.x

Beheshti, E., Weintrop, D., Orton, K., Horn, M., Jona, K., Trouille, L., & Wilensky, U. (2013). *Bringing Expert Computational Practices into High School Science Classrooms*.

Beisenbayeva, L., & Gelişli, Y. (2016). Comparison of social studies, Turkish Republic Revolution History and Kemalism, History of Kazakhstan and World History curricula in the secondary education in Turkey and the Republic of Kazakhstan. *International Journal of Human Sciences*, *13*(1), 532. https://doi.org/10.14687/ijhs.v13i1.3571

Bell, T., Witten, I. H., & Fellows, M. (2015). CS Unplugged. In *Computer Science Unplugged*. https://www.csunplugged.org.

Berland, M., & Duncan, S. (2016). Computational Thinking in the Wild: Uncovering Complex Collaborative Thinking Through Gameplay. *Educational Technology*, *56*, 29–35. https://doi.org/10.2307/44430490

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. https://doi.org/10.1007/s10956-015-9552-x

Bers, Marina U. (2010). The TangibleK robotics program: Applied computational thinking for young children. *Early Childhood Research and Practice*, *12*(2), 1–20.

Bers, Marina Umaschi, Flannery, L., Kazakoff, E. R., & Sullivan, A. (2014). Computational thinking and tinkering: Exploration of an early childhood robotics curriculum. *Computers and Education*, *72*, 145–157. https://doi.org/10.1016/j.compedu.2013.10.020

Bienkowski, M., Snow, E., Rutstein, D., & Grover, S. (2015). *Assessment Design Patterns for Computational Thinking Practices in Secondary Computer Science: A first look (SRI technical report)*. Retrieved from http://pact.sri.com/resources.html

Bilim Innovation Lyceum. (2021). Retrieved May 16, 2021, from www.bil.edu.kz

Blackwell, A. F., Church, L., & Green, T. R. G. (2008). The abstract is' an enemy': Alternative perspectives to computational thinking. *Proceedings PPIG*, *8*, 34–43.

Bloom, B. S. (1956). The Classification of Educational Goals. In *Taxonomy of educational objectives, Handbook 1 Cognitive Domain*. New York: McKay.

Boom, K.-D., Bower, M., Arguel, A., Siemon, J., & Scholkmann, A. (2018). Relationship between computational thinking and a measure of intelligence as a general problem-solving ability. *Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education - ITiCSE 2018*, 206–211. https://doi.org/10.1145/3197091.3197104

Borensztajn, G. (2011). *The neural basis of structure in language*. University of Amsterdam.

Boucinha, R. M., Augusto, D., Barone, C., Teresinha, J., Brackmann, C. P., & Schneider, A. M. (2019). Relationship between the Learning of Computational thinking and the Development of Reasoning. *International Journal of Advanced Engineering Research and Science (IJAERS)*, *6*(6), 623–631. Retrieved from https://dx.doi.org/10.22161/ijaers.6.6.71

Bransford, J. D., Brown, A., & Cocking, R. (2000). How people learn: Mind, Brain, Experience, and School. In *National Research Council* (Expanded E). Washington DC: National Academy Press.

Brennan, K., & Resnick, M. (2012). New frameworks for studying and assessing the development of computational thinking. *American Educational Research Association*, pp. 1–25. https://doi.org/10.1.1.296.6602

Brichacek, A. (2014). Thinking, Computational thinking boosts students' higher-order skills. Retrieved April 13, 2017, from https://www.iste.org/explore/articleDetail?articleid=232&category=Featured-videos&article=

Bridgeman, B. (1992). A Comparison of Quantitative Questions in Open-Ended and Multiple-Choice Formats. *Journal of Educational Measurement*, *29*(3), 253–271. https://doi.org/10.1111/j.1745-3984.1992.tb00377.x

British Educational Research Association [BERA]. (2018). *Ethical Guidelines for Educational Research*. *forth edit*.

Brundrett, C., & Rhoades, M. (2013). *Theories of Educational Research The purposes of research*. 12–23. Retrieved from https://www.sagepub.com/sites/default/files/upm-binaries/58936_Brundett_&_Rhodes.pdf

Buechley, L., Eisenberg, M., Catchen, J., & Crockett, A. (2008). The LilyPad Arduino. *Proceeding of* *the Twenty-Sixth Annual CHI Conference on Human Factors in Computing Systems - CHI '08*, 423. https://doi.org/10.1145/1357054.1357123

Bundy, A. (2007). Computational Thinking is Pervasive. *Journal of Scientific and Practical Computing*, *1*(2), 67–69.

Burkhalter, N., & Shegebayev, M. R. (2012). Critical thinking as culture: Teaching post-Soviet teachers in Kazakhstan. *International Review of Education*, (58), 55–72. https://doi.org/10.1007/s11159-012-9285-5

Calao, L. A., Correa, H. E., Moreno-Leon, J., & Robles, G. (2015). Developing Mathematical Thinking with Scratch An Experiment with 6th Grade Students. In *Design for teaching and learning in a networked world* (pp. 17–27). https://doi.org/10.1007/978-3-319-24258-3

Carey, J. (1997). Everyone knows that E=MC^2 now, who can explain it? *Business Week*, *3547*, 66–68.

Caspersen, M. E., Gal-Ezer, J., McGettrick, A., & Nardelli, E. (2019). Informatics as a fundamental discipline for the 21 st century. *Communications of the ACM*, *62*(4), 58–58. https://doi.org/10.1145/3310330

Catlin, D., & Woollard, J. (2014). Educational Robots and Computational Thinking. *Proceedings of 4th International Workshop Teaching Robotics, Teaching with Robotics & 5th International Conference Robotics in Education*, 144–151.

Chan, N., & Kennedy, P. E. (2002). Are Multiple-Choice Exams Easier for Economics Students? A Comparison of Multiple-Choice and "Equivalent" Constructed-Response Exam Questions. *Southern Economic Journal*, *68*(4), 957. https://doi.org/10.2307/1061503

Chaoui, N. (2011). *Finding Relationships Between Multiple-Choice Math Tests and Their Stem-Equivalent Constructed Responses* (Claremont Graduate University). https://doi.org/10.5642/cguetd/21

Chiu, T. W., & Camilli, G. (2013). Comment on 3PL IRT Adjustment for Guessing. *Applied Psychological Measurement*, *37*(1), 76–86. https://doi.org/10.1177/0146621612459369

Chomsky, N. (2002). *On Nature and Language* (A. Belletti & L. Rizzi, Eds.). https://doi.org/10.1017/CBO9780511613876

Chuang, H. C., Hu, C. F., Wu, C. C., & Lin, Y. T. (2015). Computational thinking curriculum for K-12 education - A Delphi survey. *Proceedings - 2015 International Conference on Learning and Teaching in Computing and Engineering, LaTiCE 2015*, 213–214.

https://doi.org/10.1109/LaTiCE.2015.44

Città, G., Gentile, M., Allegra, M., Arrigo, M., Conti, D., Ottaviano, S., … Sciortino, M. (2019). The effects of mental rotation on computational thinking. *Computers and Education*, *141*(January), 0–10. https://doi.org/10.1016/j.compedu.2019.103613

Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital Games, Design, and Learning: A Systematic Review and Meta-Analysis. *Review of Educational Research*, *86*(1), 79–122. https://doi.org/10.3102/0034654315582065

Cohen, L., Manion, L., & Morrison, K. (2011). Research Methods in Education. In *Research methods in education* (6th ed.). Abingdon: Routledge.

Computer Science Teachers Association (CSTA), & International Society for Technology in Education (ISTE). (2011). *Computational Thinking in K–12 Education leadership toolkit*. 43. Retrieved from http://www.iste.org/learn/computational-thinking

Computing At School Scotland. (2018). About PLAN C. Retrieved February 22, 2018, from http://www.cas.scot/plan-c/

Conley, D. (2007). Redefining College Readiness. *Educational Policy Improvement Center*, *3*. https://doi.org/10.4103/0972-2327.29204

Cooper, S., & Dann, W. (2015). Programming: A Key Component of Computational Thinking in CS Courses for Non-Majors. *ACM Inroads*, *6*(1), 50–54. https://doi.org/10.1145/2723169

Counihan, M. (2008). *Looking for logic in all the wrong places: an investigation of language, literacy and logic in reasoning* (University of Amsterdam). Retrieved from http://dare.uva.nl/record/286594%5Cnhttp://dare.uva.nl/document/118220

Covington, M. (2009). Self-worth theory: Retrospection and prospects. In *K. R. Wenzel & A. Wigfield (Eds.), Educational psychology handbook series. Handbook of motivation at school* (pp. 141–169). Retrieved from https://psycnet.apa.org/record/2009-24219-008

Creswell, J. (2012). Educational research: Planning, conducting, and evaluating quantitative and qualitative research. In *Educational Research* (Vol. 4). https://doi.org/10.1017/CBO9781107415324.004

Creswell, J. (2014). *Research Design- Qualitative, Quantitative and Mixed Methods Approaches* (4th ed.). SAGE Publications.

CS Unplugged. (2016). What is Computational Thinking. Retrieved January 29, 2018, from

Computational Thinking and CS Unplugged website: https://cs-unplugged.appspot.com/en/computational-thinking/

Csizmadia, A., Curzon, P., Dorling, M., Humphreys, S., Ng, T., Selby, C., & Woollard, J. (2015). *Computational thinking A guide for teachers* (p. 18). p. 18. Computing At School.

Curzon, P., Dorling, M., Selby, C., & Woollard, J. (2014). *Developing computational thinking in the classroom: a framework*. Retrieved from http://eprints.soton.ac.uk/369594/10/DevelopingComputationalThinkingInTheClassroomaFramework.pdf

Curzon, P., & McOwan, P. W. (2017). *The power of computational thinking*. London: World Scientific Publishing Europe Ltd.

Czerkawski, B. (2018). A Research Agenda for Computational Thinking. In *Computational Thinking in Primary Education* (pp. 65–77). https://doi.org/10.4018/978-1-5225-3200-2.ch004

Dagiene, V., & Stupuriene, G. (2016). Bebras - a Sustainable Community Building Model for the Concept Based Learning of Informatics and Computational Thinking. *Informatics in Education*, *15*(1), 25–44. https://doi.org/10.15388/infedu.2016.02

Dawson, C. (2009). *Introduction to research methods* (4th ed.). Oxford: How To Content.

Dawson, C. (2013). *Introduction to Research Methods. A practical guide for anyone undertaking a research project.* (4th ed.). https://doi.org/10.1017/CBO9781107415324.004

De Ayala, R. (2009). The theory and practice of Item Response Theory. In *Change for the Better: Personal development through practical psychotherapy*. https://doi.org/10.4135/9781526438447.n17

de Bono, E. (1970). Lateral Thinking: A Textbook of Creativity. In *Penguin Books* (Vol. 3). https://doi.org/11.0002/bies.10094

De Corte, E. (2003). Transfer as the Productive Use of Acquired Knowledge, Skills, and Motivations. *Current Directions in Psychological Science*, *12*(4), 142–146. https://doi.org/10.1111/1467-8721.01250

Denmars, C. (2010). *Item Response Theory, Understanding Statistics, Measurement*. Oxford: Oxford University Press.

Denner, J., Werner, L., & Ortiz, E. (2012). Computer games created by middle school girls: Can they be used to measure understanding of computer science concepts? *Computers and*

Education, *58*(1), 240–249. https://doi.org/10.1016/j.compedu.2011.08.006

Denning, P. J. (2009). The profession of IT Beyond computational thinking. *Communications of the ACM*, *52*(6), 28. https://doi.org/10.1145/1516046.1516054

Denning, P. J. (2017). Remaining trouble spots with computational thinking. *Communications of the ACM*, *60*(6), 33–39. https://doi.org/10.1145/2998438

Department for Education. (2013a). *Computing programmes of study: Key Stages 3 and 4* (p. 3). p. 3. https://doi.org/FE-00190-2013

Department for Education. (2013b). The National Curriculum in England, Framework Document. Retrieved from https://www.gov.uk/topic/schools-colleges-childrens-services/curriculum-qualifications

DevTech Research Group, & Playful Invention Company. (2018). ScratchJr - Home. Retrieved April 6, 2018, from https://www.scratchjr.org/

Dewey, J. (1910). How We Think: A Restatement of the Relation of Reflective Thinking to the Educative Process. In *D.C. Heath & Co.* Boston.

Diagnostic Questions. (2016). *Advice for Writing Multiple-Choice Questions*.

Diagnostic Questions. (2020). Diagnostic Questions. Retrieved from diagnosticquestions.com

DiSessa, A. (2000). *Changing Minds: Computers, Learning, and Literacy.* Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED441394&site=eds-live

Doleck, T., Bazelais, P., Lemay, D. J., Saxena, A., & Basnet, R. B. (2017). Algorithmic thinking, cooperativity, creativity, critical thinking, and problem solving: exploring the relationship between computational thinking skills and academic performance. *Journal of Computers in Education*, *4*(4), 355–369. https://doi.org/10.1007/s40692-017-0090-9

Dolowits, D., Buckler, S., & Sweeney, F. (2008). *Researching Online* (1st ed.). Basingstoke: Palgrave Macmillan.

Dorling, M., & Walker, M. (2014). *Computing Progression Pathways*. Computing At School.

Downing, S. M., & Haladyna, T. M. (2006). *Handbook of test development* (1st ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Dufresne, R. J., Leonard, W. J., & Gerace, W. J. (2002). Marking sense of students' answers to multiple-choice questions. *The Physics Teacher*, *40*(3), 174–180.

https://doi.org/10.1119/1.1466554

Durak, H. Y., & Saritepeci, M. (2017). Analysis of the relation between computational thinking skills and various variables with the structural equation model. *Computers & Education*, *116*, 191–202. https://doi.org/10.1016/j.compedu.2017.09.004

Ebel, R., & Frisbie, D. (1991). Essentials of Educational Measurement. In *Educational Researcher*. https://doi.org/10.2307/1175572

EGOV.KZ. (2021). SNT in Kazakhstan: preparation and order of conduction. Retrieved March 30, 2021, from https://egov.kz/cms/en/articles/about_ent

Embretson, S., & Reise, S. (2000). *Item Response Theory for Psychologists*. New Jersey: Lawrence Erlbaum Associates.

EPDC. (2014). *National Education Profile*. Retrieved from http://www.epdc.org/sites/default/files/documents/EPDC NEP_Kazakhstan.pdf

Epstein, M. L., & Brosvic, G. M. (2002). Immediate Feedback Assessment Technique: multiple-choice test that "behaves" like an essay examination. *Psychol Rep*, *90*(1), 226. https://doi.org/10.2466/pr0.90.1.226-226

Eraut, M. (2004). Transfer of knowledge between education and workplace settings. *Workplace Learning in Context*, (January 2009), 201–221. https://doi.org/10.4324/9780203571644

Ershov, A. P. (1981). Programming, the second literacy. *Microprocessing and Microprogramming*, *8*(1), 1–9. https://doi.org/10.1016/0165-6074(81)90002-8

Fenna, D. S. (2004). Assessment of Foundation Knowledge: Are Students Confident in Their Ability? *European Journal of Engineering Education*, *29*(2), 307–312. https://doi.org/10.1080/0304379032000157277

Field, A. (2009). *Discovering Statistics Using IBM SPSS Statistics* (3rd ed.). London: SAGE Publications Ltd.

Fields, D. A., Kafai, Y. B., & Giang, M. T. (2017). Youth Computational Participation in the Wild. *ACM Transactions on Computing Education*, *17*(3), 1–22. https://doi.org/10.1145/3123815

Flannery, L. P., Kazakoff, E. R., Bontá, P., Silverman, B., Bers, M. U., Resnick, M., … Resnick, M. (2013). Designing ScratchJr: Support for Early Childhood Learning Through Computer Programming. *Proceedings of the 12th International Conference on Interaction Design and Children (IDC '13)*, 1–10. https://doi.org/10.1145/2485760.2485785

Flavell, J. H. (1963). The Developmental Psychology of Jean Piaget. In *D. Van Nostrand Company*. https://doi.org/10.2307/3118952

Fraillon, J., Ainley, J., Schulz, W., Friedman, T., & Duckworth, D. (2019). *Preparing for life in a digital world: IEA International Computer and Information Literacy Study 2018 International Report*. Retrieved from https://www.iea.nl/sites/default/files/data-repository/ICILS/ICILS2018/ICILS2018_IDB_SPSS.zip

Frank B. Baker. (2001). The Basics of Item Response Theory. In *ERIC Clearinghouse on Assessment and Evaluation* (2nd ed.). United States of America: ERIC Clearinghouse on Assessment and Evaluation.

Frauenberger, C., & Purgathofer, P. (2019). Ways of thinking in informatics. *Communications of the ACM*, *62*(7), 58–64. https://doi.org/10.1145/3329674

Frey, B. B., Petersen, S., Edwards, L. M., Pedrotti, J. T., & Peyton, V. (2005). Item-writing rules: Collective wisdom. *Teaching and Teacher Education*, *21*(4), 357–364. https://doi.org/10.1016/j.tate.2005.01.008

Funk, S. C., & Dickson, K. L. (2011). Multiple-Choice and Short-Answer Exam Performance in a College Classroom. *Teaching of Psychology*, *38*(4), 273–277. https://doi.org/10.1177/0098628311421329

Furber, S. (2012). Shut down or restart? The way forward for computing in UK schools. In *Technology*.

Gaipov, D., Yaylaci, Y., Çiğ, K., & Guvercin, S. (2013). Formation of Multilingual Educational System in Kazakhstan: Kazakh-Turkish High Schools. *Procedia - Social and Behavioral Sciences*, *103*, 416–424. https://doi.org/10.1016/j.sbspro.2013.10.355

Gander, W., Petit, A., Berry, G., Demo, B., Vahrenhold, J., McGettrick, A., … Meyer, B. (2013). Informatics Education: Europe Cannot Afford to Miss the Boat. In *ACM Europe: Informatics Education Report*. Retrieved from http://europe.acm.org/iereport/ACMandIEreport.pdf

Gao, F., & Chen, L. (2005). Bayesian or non-Bayesian: A comparison study of item parameter estimation in the three-parameter logistic model. *Applied Measurement in Education*, *18*(4), 351–380. https://doi.org/10.1207/s15324818ame1804_2

Gardner, H. (2011). *Frames of Mind* (2nd ed.). New York: Basic Books.

Gayef, A., Oner, C., & Telatar, B. (2014). Is asking same question in different ways has any impact on student achievement? *Procedia - Social and Behavioral Sciences*, *152*(212), 339–342.

https://doi.org/10.1016/j.sbspro.2014.09.206

Gierl, M. J., Bulut, O., Guo, Q., & Zhang, X. (2017). Developing, Analyzing, and Using Distractors for Multiple-Choice Tests in Education: A Comprehensive Review. *Review of Educational Research*, *86*(6), 1082–1116. https://doi.org/10.3102/0034654317726529

Glass, A. L., & Sinha, N. (2013). Providing the answers does not improve performance on a college final exam. *Educational Psychology*, *33*(1), 87–118. https://doi.org/10.1080/01443410.2012.723864

Google for Education. (2015). Exploring Computational Thinking. Retrieved April 3, 2017, from https://edu.google.com/resources/programs/exploring-computational-thinking/#!home

Gottfredson, L. S. (1998). The General Intelligence Factor. *Scientific American Presents*, 24–30.

Gouws, L., Bradshaw, K., & Wentworth, P. (2013a). Computational thinking in educational activities. *Proceedings of the 18th ACM Conference on Innovation and Technology in Computer Science Education - ITiCSE '13*, 10. https://doi.org/10.1145/2462476.2466518

Gouws, L., Bradshaw, K., & Wentworth, P. (2013b). First year student performance in a test for computational thinking. *ACM International Conference Proceeding Series*, 271–277. https://doi.org/10.1145/2513456.2513484

Greenberger, M. (1962). *The Computer in the University. Computers and the World of the Future.* Retrieved from https://archive.org/details/managementcomput00gree

Gretter, S., & Yadav, A. (2016). Computational Thinking and Media & Information Literacy: An Integrated Approach to Teaching Twenty-First Century Skills. *TechTrends*, *60*(5), 510–516. https://doi.org/10.1007/s11528-016-0098-4

Grover, S. (2011). *Robotics and Engineering for Middle and High School Students to Develop Computational Thinking Robotics and Engineering for Middle and High School Students to Develop Computational Thinking*. (650), 1–15.

Grover, S. (2015). "Systems of Assessments" for Deeper Learning of Computational Thinking in K-12. *Annual Meeting of the American Educational Research Association*, (650).

Grover, S., Cooper, S., & Pea, R. (2014). Assessing computational learning in K-12. *Proceedings of the 2014 Conference on Innovation & Technology in Computer Science Education - ITiCSE '14*, 57–62.

Grover, S., & Pea, R. (2013a). Computational Thinking in K–12: A Review of the State of the Field.

Educational Researcher, *42*(1), 38–43. https://doi.org/10.3102/0013189X12463051

Grover, S., & Pea, R. (2013b). Using a Discourse-Intensive Pedagogy and Android's App Middle School Students. *SIGSCE'13*, 723–728. https://doi.org/10.1145/2445196.2445404

Guggemos, J., Seufert, S., & Román-González, M. (2019). Measuring computational thinking - Adapting a performance test and a self-assessment instrument for German-speaking countries. *16th International Conference on Cognition and Exploratory Learning in Digital Age, CELDA 2019*, (Celda), 183–191.

Guzdial, M. (2008). Education Paving the way for computational thinking. *Communications of the ACM*, *51*, 25–27. https://doi.org/10.1145/1378704.1378713

Guzdial, Mark. (2015). Learner-Centered Design of Computing Education: Research on Computing for Everyone. *Synthesis Lectures on Human-Centered Informatics*, *8*(6), 1–165. https://doi.org/10.2200/S00684ED1V01Y201511HCI033

Guzdial, Mark. (2019). A new definition of Computational Thinking: It's the Friction that we want to Minimize unless it's Generative. Retrieved May 22, 2019, from https://computinged.wordpress.com/2019/04/29/what-is-computational-thinking-its-the-friction-that-we-want-to-minimize/amp/

Haladyna, T. M., & Downing, S. M. (1993). How Many Options is Enough for a Multiple-Choice Test Item? *Educational and Psychological Measurement*, *53*(4), 999–1010. https://doi.org/10.1177/0013164493053004013

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, *15*(3), 309–333. https://doi.org/10.1207/S15324818AME1503

Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. New York: Routledge.

Haladyna, T. M., & Steven, M. (1989). Validity of a Taxonomy of Multiple-Choice Item- Writing Rules. *Applied Measurement in Education*, *2*(1), 37–50. https://doi.org/10.1207/s15324818ame0201

Ham, D. A. (2018). Spatial Thinking as a Path Towards Computational Thinking. In *Computational Thinking in Primary Education* (pp. 103–122). https://doi.org/10.4018/978-1-5225-3200-2.ch006

Hamilton, L. S. (1994). *An Investigation of Students' Affective Responses to Alternative Assessment*

Formats. Washington DC.

Hancock, G. R. (1994). Cognitive complexity and the comparability of multiple-choice and constructed-response test formats. *Journal of Experimental Education*, *62*(2), 143–157. https://doi.org/10.1080/00220973.1994.9943836

Hattie, J. A. C., & Donoghue, G. M. (2016). Learning strategies: a synthesis and conceptual model. *Science of Learning*, *1*(1). https://doi.org/10.1038/npjscilearn.2016.13

Hattie, J., & Gregory, D. (2018). A model of learning: optimizing the effectiveness of learning strategies. In *Contemporary theories of learning: Learning theorists ... in their own words* (2nd ed., pp. 97–113). Oxon: Routledge.

Haynie, W. (1994). Effects of multiple-choice and short-answer tests on delayed retention learning. *Journal of Technology Education*, *6*(1), 1–7. Retrieved from http://scholar.lib.vt.edu/ejournals/JTE/v6n1/haynie.jte-v6n1

Heim, A. Q., & Watts, K. P. (1967). Open-Ended Answering in a Vocabulary Test. *British Journal of Educational Psychology*, *37*(3), 339–346.

Heintz, F., Mannila, L., & Farnqvist, T. (2016). A review of models for introducing computational thinking, computer science and computing in K-12 education. *Proceedings - Frontiers in Education Conference, FIE*, *2016-Novem*. https://doi.org/10.1109/FIE.2016.7757410

Hershkovitz, A., Sitman, R., Israel-Fishelson, R., Eguíluz, A., Garaizar, P., & Guenaga, M. (2019). Creativity in the acquisition of computational thinking. *Interactive Learning Environments*, *27*(5–6), 628–644. https://doi.org/10.1080/10494820.2019.1610451

Holder, W. W., & Mills, C. N. (2001). Pencils down, computer up: The new CPA exam. *Journal of Accountancy*, *191*(3), 57–60.

Hoskey, A., & Zhang, S. (2017). Computational thinking: what does it really mean for the K-16 computer science education community. *Journal of Computing Sciences in Colleges*, 129–135.

Hu, C. (2011). Computational Thinking – What It Might Mean and What We Might Do about It. *ITiCSE '11: Proceedings of the 16th Annual Joint Conference on Innovation and Technology in Computer Science Education*, 223–227. https://doi.org/10.1145/1999747.1999811

ICILS. (2020). IEA Releases International Results of the International Computer and Information Literacy Study 2018. Retrieved March 21, 2021, from https://www.iea.nl/sites/default/files/2019-11/ICILS Results Press Release.pdf

Immordino-Yang, M. H., Christodoulou, J., & Singh, V. (2012). Rest Is Not Idleness: Implications of the Brain's Default Mode for Human Development and Education. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *7*(4), 352–364.

International Association for the Evaluation of Educational Achievement. (2015). The IEA's International Computer and Information Literacy Study (ICILS) 2018. Retrieved August 15, 2018, from IEA website: https://www.iea.nl/sites/default/files/studies/Computational thinking leaflet.pdf

International Association for the Evaluation of Educational Achievement. (2018). ICILS. International Computer and Information Literacy Study | IEA. Retrieved February 27, 2018, from http://www.iea.nl/icils

International Society for Technology in Education (ISTE), & Computer Science Teachers Association (CSTA). (2011). Operational Definition of Computational Thinking.

Jamil, H. M. (2017a). Automated Personalised Assessment of Computational Thinking MOOC Assignments. *Proceedings - IEEE 17th International Conference on Advanced Learning Technologies, ICALT 2017*, 261–263. https://doi.org/10.1109/ICALT.2017.147

Jamil, H. M. (2017b). Smart Assessment of and Tutoring for Computational Thinking MOOC Assignments using MindReader. *ArXiv Preprint*. Retrieved from http://arxiv.org/abs/1705.00959

Jona, K., Wilensky, U., Trouille, L., Horn, M., Orton, K., Weintrop, D., & Beheshti, E. (2014). Embedding Computational Thinking in Science , Technology , Engineering , and Math (CT-STEM). *Future Directions in Computer Science Education Summit Meeting*, (2002), 1–5.

Juškevičiene, A., & Dagiene, V. (2018). Computational thinking relationship with digital competence. *Informatics in Education*, *17*(2), 265–284. https://doi.org/10.15388/infedu.2018.14

Kalelioglu, F., Gulbahar, Y., & Kukul, V. (2016). A Framework for Computational Thinking Based on a Systematic Research Review. *Baltic Journal of Modern Computing*, *4*(3), 583–596.

Kallia, M. (2017). *Assessment in Computer Science courses : A Literature Review*. London.

Karpicke, J. D., & Roediger, H. L. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, *57*(2), 151–162. https://doi.org/10.1016/j.jml.2006.09.004

Kaufman, J., & Stenberg, R. (2010). *The Cambridge Handbook of Creativity*. Retrieved from http://library1.nida.ac.th/termpaper6/sd/2554/19755.pdf

Kazakhstani results in the ICILS-2018 International Computer and Information Literacy Study. (2020). Retrieved March 21, 2021, from http://iac.kz/en/events/kazakhstani-results-icils-2018-international-computer-and-information-literacy-study

Khenner, E. K. (2016). Computational Thinking. *The Education and Science Journal*, *2*(2), 18–33. https://doi.org/10.17853/1994-5639-2016-2-18-33

Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up*. Washington DC: National Academy of Sciences.

Koh, K. H., Basawapatna, A., Nickerson, H., & Repenning, A. (2014). Real time assessment of computational thinking. *Proceedings of IEEE Symposium on Visual Languages and Human-Centric Computing, VL/HCC*, 49–52. https://doi.org/10.1109/VLHCC.2014.6883021

Kong, S., & Li, P. (2016). A Case Study Illustrating Coding for Computational Thinking Development. *24thInternational Conference on Computers in Education. India: Asia-Pacific Society for Computers in Education*, (October).

Korkmaz, Ö., Çakir, R., & Özden, M. Y. (2017). A validity and reliability study of the computational thinking scales (CTS). *Computers in Human Behavior*, *72*, 558–569. https://doi.org/10.1016/j.chb.2017.01.005

Korkmaz, Ö., Çakır, R., & Özden, Y. (2015). Computational Thinking Levels Scale (CTLS) Adaptation for Secondary School Level. *Gazi Journal of Education Sciences*, 143–162.

Korucu, A. T., Gencturk, A. T., & Gundogdu, M. M. (2017). Examination of the Computational Thinking Skills of Students. *Journal of Learning and Teaching in Digital Age*, *2*(1), 11–19.

Koshegulova, M., & Mindetbay, Y. (2020). The Impact of the Flipped Classroom on Students' Academic Achievements in Secondary Schools. *IGI Global*, 175–192. https://doi.org/10.4018/978-1-7998-2551-7.ch010

Lacourly, N., Martin, J. S., Silva, M., & Uribe, P. (2018). *IRT scoring and the principle of consistent order*. 1–15. Retrieved from http://arxiv.org/abs/1805.00874

Lai, R. (2019). *What Underlies Computational Thinking: Exploring its Cognitive Mechanism and Educational Implications*. Hong Kong: The Education University of Hong Kong.

Lau, W. W. F., & Yuen, A. H. K. (2009). Exploring the effects of gender and learning styles on computer programming performance: Implications for programming pedagogy. *British*

Journal of Educational Technology, *40*(4), 696–712. https://doi.org/10.1111/j.1467-8535.2008.00847.x

Lee, I., Martin, F., Denner, J., Coulter, B., Allan, W., Erickson, J., … Werner, L. (2011). Computational thinking for youth in practice. *ACM Inroads*, *2*(1), 32–37. https://doi.org/10.1145/1929887.1929902

Lee, J., Jung, Y., & Park, H. (2017). Gender Differences in Computational Thinking, Creativity, and Academic Interest on Elementary SW Education. *Journal of The Korean Association of Information Education*, *21*(4), 381–380.

Lenssen, P. (2005). Why Good Programmers Are Lazy and Dumb. Retrieved January 31, 2020, from http://blogoscoped.com/archive/2005-08-24-n14.html

Leonard, J., Buss, A., Gamboa, R., Mitchell, M., Fashola, O. S., Hubert, T., & Almughyirah, S. (2016). Using Robotics and Game Design to Enhance Children's Self-Efficacy, STEM Attitudes, and Computational Thinking Skills. *Journal of Science Education and Technology*. https://doi.org/10.1007/s10956-016-9628-2

Lewis, A., & Smith, D. (1993). Defining Higher Order Thinking. *Theory into Practice*, *32*(3), 131–137. https://doi.org/10.1080/00405849309543588

Lin, V. J. (2015). *Computational Thinking and Technology Toys*. 2–101. Retrieved from http://repository.wellesley.edu/thesiscollection/307/

Linden, W. J. van der. (2018). Handbook of Item Response Theory. In *CRC Press* (Vol. 3). California.

Linn, M. C., Aho, A. V., Blake, M. B., Constable, R., Kafai, Y. B., Kolodner, J. L., … Bradley, S. (2010). Report of a Workshop on The Scope and Nature of Computational Thinking. In *Thinking* (Vol. 1). https://doi.org/10.17226/12840

Livingstone, I., & Saeed, S. (2017). *Hacking the Curriculum: Creative Computing and the Power of Play*. Woodbridge: John Catt Educational Ltd.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lu, J. J., & Fletcher, G. H. L. (2009). Thinking About Computational Thinking Categories and Subject Descriptors. *Sigcse*, 260–264. https://doi.org/10.1145/1539024.1508959

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*, Vol. 41, pp.

51–61. https://doi.org/10.1016/j.chb.2014.09.012

Magno, C. (2009). Demonstrating the difference between Classical Test Theory and Item Response Theory using derived test data. *The International Journal of Educational and …*, *1*(1), 1–11. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1426043

Malyn-Smith, J., & Lee, I. (2012). Application of the Occupational Analysis of Computational Thinking-Enabled STEM Professionals as a Program Assessment Tool. *Editor: Steven Gordon Associate Editors: Thomas*, *3*(1), 2–10.

Mannila, L., Dagiene, V., Demo, B., Grgurina, N., Mirolo, C., Rolandsson, L., & Settle, A. (2014). Computational Thinking in K-9 Education. *Proceedings of the Working Group Reports of the 2014 on Innovation 38; Technology in Computer Science Education Conference*, 1–29. https://doi.org/10.1145/2713609.2713610

Martinez, M. E. (1999). Cognition and the question of test item format. *Educational Psychologist*, *34*(4), 207–218. https://doi.org/10.1207/s15326985ep3404_2

Marton, F. (2006). Sameness and Difference in Transfer. *The Journal of the Learning Sciences*, *15*(4), 499–535. https://doi.org/10.1207/s15327809jls1504

McCarthy, J. (2016). PBL, Kazakhstan, and Global Success Skills. *Edutopia*, 1–8.

McKeachie, W. J. (1987). Cognitive skills and their transfer: Discussion. *International Journal of Educational Research*, *11*(6), 707–712. https://doi.org/10.1016/0883-0355(87)90010-3

Merrel, J. D., Cirillo, P. F., Schwartz, P. M., & Webb, J. A. (2015). Multiple-Choice Testing Using Immediate Feedback—Assessment Technique (IF AT®) Forms: Second-Chance Guessing vs. Second-Chance Learning? *Higher Education Studies*, *5*(5), 50–55. https://doi.org/10.5539/hes.v5n5p50

Miller, G. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine*, *65*(9), S63–S67.

Mindetbay, Y., Bokhove, C., & Woollard, J. (2019a). The Measurement of Computational Thinking Performance Using Multiple-choice Questions. *Proceedings of International Conference on Computational Thinking Education 2019*, 176–179. Hong Kong: The Education University of Hong Kong.

Mindetbay, Y., Bokhove, C., & Woollard, J. (2019b). What is the Relationship between Students' Computational Thinking Performance and School Achievement? *International Journal of Computer Science Education in Schools*, *March*(Working Group), 3–19.

https://doi.org/https://doi.org/10.21585/ijcses.v0i0

Ministry of Education and Science of the Republic of Kazakhstan. (2010). SPED. Retrieved December 13, 2016, from http://www.akorda.kz/upload/SPED.doc

Ministry of Education and Science of the Republic of Kazakhstan. (2014). *Country Background Report for Kazakhstan*. Retrieved from http://iac.kz/ru/

Ministry of National Economy of the Republic of Kazakhstan Committee on Statistics. (2018). Number of the pupils in general education schools by kind of studying language. Retrieved April 2, 2018, from http://stat.gov.kz

Ministry of Science and Education. (2017). Unified National Test. Retrieved May 15, 2017, from http://edu.gov.kz/kz/page/ulttik_biringai_testileu

Minsky, M. (1988). *The society of mind*. New York: Simon & Schuster.

Moreno-Leon, J., Robles, G., & Román-González, M. (2016). Comparing Computational Thinking Development Assessment Scores with Software Complexity Metrics. *2016 IEEE Global Engineering Education Conference (EDUCON)*, 1040–1045. https://doi.org/10.1109/EDUCON.2016.7474681

Moreno-leon, J., Román-González, M., & Robles, G. (2018). On Computational Thinking as a Universal Skill. *2018 IEEE Global Engineering Education Conference (EDUCON)*, 1690–1695. Retrieved from 10.1109/EDUCON.2018.8363437

Moreno, R., Martínez, R. J., & Muñiz, J. (2015). Guidelines based on validity criteria for the development of multiple choice items. *Psicothema*, *27*(4), 388–394. https://doi.org/10.7334/psicothema2015.110

National Academy of Education. (2016). Updated curriculum by the National Academy of Education.

National Testing Centre. (2017a). Online Trial Testing. Retrieved December 12, 2017, from http://www.testcenter.kz/schools/voudso-online/

National Testing Centre. (2017b). Unified National Test. Retrieved May 15, 2017, from http://testcenter.kz/entrants/ent/

Naveh, N. (2006). *Developing Algorithmic Thinking in Middle School Pupils in Israel*. Anglia Ruskin University.

Nazarbayev, N. *The Strategy for development of the Republic of Kazakhstan until the year 2030*. , (1997).

Nesiba, N., Pontelli, E., & Staley, T. (2015). DISSECT: Exploring the relationship between computational thinking and English literature in K-12 curricula. *Proceedings - Frontiers in Education Conference, FIE*. https://doi.org/10.1109/FIE.2015.7344063

Oakley, B. (2014). *A mind for numbers* (Version 1). New York, NY, USA: Penguin Group (USA) LLC.

Oates, T., Coe, R., Jones, S. P., Scratcherd, T., & Woodhead, S. (2016). *Quantum : Tests Worth Teaching To* (pp. 1–7). pp. 1–7. Retrieved from http://community.computingatschool.org.uk/files/7256/original.pdf

OECD. (2015). Reviews of School Resources. In *World Bank Publications*. https://doi.org/http://dx.doi.org/10.1787/9789264245891-en

OECD. (2018). *PISA 2021 Mathematics Framework (Second Draft)* (Vol. 19). Vol. 19. Retrieved from https://www.oecd.org/pisa/sitedocument/PISA-2021-mathematics-framework.pdf

OECD. (2019). *PISA 2018 results*. Retrieved from http://www.oecd.org/pisa/PISA-results_ENGLISH.png

OECD. (2020a). Developing a national assessment that supports Kazakhstan's education goals. *OECD Education Policy Perspectives*, (27).

OECD. (2020b). Strengthening national examinations in Kazakhstan to achieve national goals. *OECD Education Policy Perspectives*, (24). https://doi.org/https://doi.org/10.1787/0bf8662b-en.

OECD, & World Bank. (2014). *Reviews of National Policies for Education: Secondary Education in Kazakhstan*. https://doi.org/10.1787/9789264030206-en

Ohlsson, S. (2012). Stellan Ohlsson: Deep Learning: How the Mind Overrides Experience. In *Science & Education* (Vol. 21). https://doi.org/10.1007/s11191-012-9449-5

Oluk, A., & Korkmaz, Ö. (2016). Comparing Students' Scratch Skills with Their Computational Thinking Skills in Terms of Different Variables. *International Journal of Modern Education and Computer Science*, *8*(11), 1–7. https://doi.org/10.5815/ijmecs.2016.11.01

Ozcinar, H., Wong, G., & Ozturk, H. T. (2017). *Teaching Computational Thinking in Primary Education* (H. Ozcinar, G. Wong, & H. T. Ozturk, Eds.). Hershey: IGI Global.

Palmer, E. J., & Devitt, P. G. (2007). Assessment of higher order cognitive skills in undergraduate education: Modified essay or multiple choice questions? Research paper. *BMC Medical*

Education, *7*, 1–7. https://doi.org/10.1186/1472-6920-7-49

Papert, S. (1980). Mindstorm. In *Journal of Chemical Information and Modeling*. https://doi.org/10.1017/CBO9781107415324.004

Papert, S. (1996). An exploration in the space of mathematics educations. *International Journal of Computers for Mathematical Learning*, 95–123. https://doi.org/10.1038/537308a

Parkhomenko, M. (2017). UNT is not Rubicon anymore. Retrieved April 3, 2021, from Kazpravda website: http://www.m.kazpravda.kz/articles/view/ent-teper-ne-rubikon1/

Paxton, M. (2000). A linguistic perspective on multiple choice questioning. *Assessment and Evaluation in Higher Education*, *25*(2), 109–119. https://doi.org/10.1080/713611429

Peariso, J. F. (2006). *Multiple-Choice Versus Project-Based*. University of La Verne.

Pellas, N., & Vosinakis, S. (2017). How can a simulation game support the development of computational problem-solving strategies? *IEEE Global Engineering Education Conference, EDUCON*, 1129–1136. https://doi.org/10.1109/EDUCON.2017.7942991

Perkins, D., & Solomon, G. (1992). *Transfer of Learning*. Oxford: Pergamon Press.

Petre, M., & Price, B. (2004). Using Robotics to Motivate “Back Door” Learning. *Education and Information Technologies*, *9*(2), 147–158.

Peyton-Jones, S. (2014). *Understanding the new programmes of study for computing*. Retrieved from http://www.computingatschool.org.uk

Piaget, J. (1958). *The Growth of Logical Thinking from Childhood to Adolescence*. United States of America: Basic Books.

Polat, E., Hopcan, S., Kucuk, S., & Sisman, B. (2021). A comprehensive assessment of secondary school students’ computational thinking skills. *British Journal of Educational Technology*, (December 2020), 1–16. https://doi.org/10.1111/bjet.13092

Polya, G. (2004). *How to Solve It: A New Aspect of Mathematical Method*. https://doi.org/10.1201/9781003037149-1

Popham, W. . (1995). *Classroom Assessment, What Teachers Need To Know*. Allyn and Bacon.

Portelance, D. D. J., & Bers, M. U. M. (2015). Code and Tell: Assessing Young Children’s Learning of Computational Thinking Using Peer Video Interviews with ScratchJr. *Proceedings of the 14th International Conference on Interaction Design and Children*, 271–274.

https://doi.org/10.1145/2771839.2771894

Prottsman, C. L. L. (2011). *Computational thinking and women in computer science*. University of Oregon.

Punch, K., & Oancea, A. (2014). *Introduction to research methods in education* (2nd ed.). London: SAGE Publications Ltd.

Purgathofer, P. (2019a). Ways of Thinking in Informatics-Computational Thinking. Retrieved February 17, 2020, from TU Wien website: https://wot.pubpub.org/pub/b6x65c3m

Purgathofer, P. (2019b). Ways of Thinking in Informatics-Design Thinking. Retrieved February 17, 2020, from TU Wien website: https://wot.pubpub.org/pub/rjgbqa43

Qin, Y., Bothell, D., & Anderson, J. (2007). ACT-R Meets fMRI.pdf. *LNAI 4845*, 205–222. Berlin: Springer-Verlag.

Reynolds, C. R., Livingston, R. B., & Willson, V. (2009). *Measurement and assessment in education* (2nd ed.). New Jersey: Pearson.

Rizopoulos, D. (2018). *Latent Trait Models under IRT* (pp. 1–69). pp. 1–69. Retrieved from https://cran.r-project.org/web/packages/ltm/ltm.pdf

Roberts, T. S. (2006). The use of multiple choice tests for formative and summative assessment. *Conferences in Research and Practice in Information Technology Series*, *52*, 175–180.

Robson, C. (2002). Real World Research: A Resource for Social Scientists and Practitioner-Researchers. *Blackwell Publishing*, pp. 1–608. https://doi.org/10.1016/j.jclinepi.2010.08.001

Robson, C., & McCartan, K. (2016). *Real World Research* (4th ed.). Italy: Wiley & Sons Ltd.

Rodrigues, R. S., Andrade, W. L., & Campos, L. M. R. S. (2016). Can Computational Thinking help me? A quantitative study of its effects on education. *Proceedings - Frontiers in Education Conference, FIE*, *2016-Novem*. https://doi.org/10.1109/FIE.2016.7757409

Rodriguez, M. C. (2003). Construct Equivalence of Multiple-Choice and Constructed-Response Items : A Random Effects Synthesis of Correlations. *Journal of Educational Measurement*, *40*(2), 163–184. https://doi.org/10.1111/j.1745-3984.2003.tb01102.x

Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues & Practice*, *24*(2), 3–13. https://doi.org/10.1111/j.1745-3992.2005.00006.x

Roediger, H. L., & Marsh, E. J. (2005). The Positive and Negative Consequences of Multiple-Choice Testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 1155–1159. https://doi.org/10.1037/0278-7393.31.5.1155

Rogers, W. T., & Harley, D. (1999). An Empirical Comparison of Three-and Four-Choice Items and Tests: Susceptibility to Testwiseness and Internal Consistency Reliability. *Educational and Psychological Measurement*, *59*(2), 234–247. https://doi.org/10.1177/00131649921969820

Román-González, M., Moreno-León, J., Robles, G., Román-gonzález, M., Moreno-león, J., Juan Carlos, R., & Rey Juan Carlos, U. (2017). *Complementary Tools for Computational Thinking Assessment*. Retrieved from https://www.researchgate.net/publication/318469859

Román-González, M., Pérez-González, J.-C., & Jiménez-Fernández, C. (2016). Which cognitive abilities underlie computational thinking? Criterion validity of the Computational Thinking Test. *Computers in Human Behavior*. https://doi.org/10.1016/j.chb.2016.08.047

Román-González, M., Pérez-González, J.-C., Moreno-Leon, J., & Robles, G. (2016). Does computational thinking correlate with personality? *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality - TEEM '16*, 51–58. New York: ACM.

Román-González, M., Pérez-González, J.-C., Moreno-Leon, J., & Robles, G. (2017). Extending the nomological network of computational thinking with non-cognitive factors. *Computers in Human Behavior*. https://doi.org/10.1016/j.chb.2017.09.030

Román-González, M., Pérez-González, J. C., Moreno-León, J., & Robles, G. (2018). Extending the nomological network of computational thinking with non-cognitive factors. *Computers in Human Behavior*, *80*, 441–459. https://doi.org/10.1016/j.chb.2017.09.030

RStudio. (2021). About RStudio. Retrieved from https://www.rstudio.com/about/

Sabitzer, B., Demarle-Meusel, H., & Jarnig, M. (2018). Computational thinking through modeling in language lessons. *IEEE Global Engineering Education Conference, EDUCON*, *2018-April*(2), 1913–1919. https://doi.org/10.1109/EDUCON.2018.8363469

Sanford, J. F., & Naidu, J. T. (2016). Computational Thinking Concepts for Grade School. *Contemporary Issues in Education Research (CIER)*, *9*(1), 23–32. https://doi.org/10.19030/cier.v9i1.9547

Saxena, A., & Basnet, R. B. (2017). Algorithmic thinking, cooperativity, creativity, critical thinking, and problem solving: exploring the relationship between computational thinking skills and academic performance. *Journal of Computers in Education*. https://doi.org/10.1007/s40692-017-0090-9

Schuwirth, L. W., van der Vleuten, C. P., & Donkers, H. H. (1996). A closer look at cueing effects in multiple-choice questions. *Medical Education*, *30*(1), 44–49. https://doi.org/10.1111/j.1365-2923.1996.tb00716.x

Scott, S., Scott, D., & Webber, C. (2016). *Assessment in Education* (Vol. 2). Switzerland: Springer International Publishing.

Seiter, L. (2015). Using SOLO to Classify the Programming Responses of Primary Grade Students. *Proceedings of the 46th ACM Technical Symposium on Computer Science Education - SIGCSE '15*, 540–545. https://doi.org/10.1145/2676723.2677244

Seiter, L., & Foreman, B. (2013). Modeling the learning progressions of computational thinking of primary grade students. *Proceedings of the Ninth Annual International ACM Conference on International Computing Education Research*, 59–66. https://doi.org/10.1145/2493394.2493403

Selby, C. (2012). Promoting computational thinking with programming. *Proceedings of the 7th Workshop in Primary and Secondary Computing Education on - WiPSCE '12*, 74. https://doi.org/10.1145/2481449.2481466

Selby, C. (2013). Computational Thinking : The Developing Definition. *ITiCSE Conference 2013*, 5–8.

Selby, C. (2014). *How can the teaching of programming be used to enhance computational thinking skills?* University of Southampton.

Selby, C. (2015). Relationships: computational thinking, pedagogy of programming, and Bloom's Taxonomy. *WiPSCE '15*. https://doi.org/10.1145/2818314.2818315

Selby, C., Dorling, M., & Woollard, J. (2015). Evidence of assessing computational thinking. In *IFIP TC3 Working Conference: a New Culture of Learning: Computing and Next Generations*. Lithuania.

Selby, C., & Woollard, J. (2014). Refining an Understanding of Computational Thinking. *Technology, Pedagogy and Education*, (2013), 1–23.

Sengupta, P., Kinnebrew, J. S., Basu, S., Biswas, G., & Clark, D. (2013). Integrating computational thinking with K-12 science education using agent-based computation: A theoretical framework. *Education and Information Technologies*, *18*(2), 351–380.

https://doi.org/10.1007/s10639-012-9240-x

Shaftel, J., & Shaftel, T. L. (2007). Educational Assessment and the AACSB. *Issues in Accounting Education*, *22*(2), 215–232. https://doi.org/10.2308/iace.2007.22.2.215

Shaniyev, Y., Gesen, I., Aidarbayev, N., Akhmetov, N., & Yerzhanov, E. (2017). *Informatics - A bilingual textbook - Grade 8* (1st ed.). Astana: Astana Kitap.

Sheil, B. a. (1981). The Psychological Study of Programming. *ACM Computing Surveys*, *13*(1), 101–120. https://doi.org/10.1145/356835.356840

Sherman, M., & Martin, F. (2015). The assessment of mobile computational thinking. *Journal of Computing Sciences in Colleges*, *30*(6), 53–59.

Simkin, M. G., & Kuechler, W. L. (2005). Multiple-Choice Tests and Student Understanding: What Is the Connection? *Decision Sciences Journal of Innovative Education*, *3*(1), 73–98. https://doi.org/10.1111/j.1540-4609.2005.00053.x

Smith, M. A., & Karpicke, J. D. (2014). Retrieval practice with short-answer, multiple-choice, and hybrid tests. *Memory*, *22*(7), 784–802. https://doi.org/10.1080/09658211.2013.831454

Snow, R. E. (1993). (1993). Construct validity and constructed-response tests. In *E. B. Randy & C. W. William (Eds.) Construction versus choice in cognitive measurement*. Hillsdale, NJ: Lawrence Erlbaum.

Sobota, J., Pisl, R., Balda, P., & Schlegel, M. (2013). Using online lectures to support student learning of control engineering. *IFAC Proceedings Volumes (IFAC-PapersOnline)*, *10*(PART 1), 132–137. https://doi.org/10.3182/20130828-3-UK-2039.00003

Spielman, R., Jenkins, W., & Lovett, M. (2020). *Psychology 2e* (2nd ed.). Houston: OpenStax.

Tarrant, M., Knierim, A., Hayes, S. K., & Ware, J. (2006). The frequency of item writing flaws in multiple-choice questions used in high stakes nursing assessments. *Nurse Education in Practice*, *6*(6), 354–363. https://doi.org/10.1016/j.nepr.2006.07.002

Tedre, M., & Denning, P. J. (2016). The long quest for computational thinking. *Proceedings of the 16th Koli Calling International Conference on Computing Education Research - Koli Calling '16*, 120–129. https://doi.org/10.1145/2999541.2999542

Telyukanov, Z. (2021). *Bilim Innovation Lyceums achievement report*. Nur-Sultan.

The College Board. (2017). *AP Computer Science Principles Including the Curriculum Framework*. Retrieved from https://apcentral.collegeboard.org/pdf/ap-computer-science-principles-course-and-exam-description.pdf

The Constitution of the Republic of Kazakhstan. (n.d.). Retrieved January 12, 2021, from https://www.akorda.kz/en/official_documents/constitution

The Ministry o Education and Science of the Republic of Kazakhstan-UNESCO Cluster Office. (2004). *The Background Report of Kazakhstan The Status of Preschool Education in the Republic of Kazakhstan*.

The Royal Society. (2017). *After the reboot: computing education in UK schools*. London.

Thissen, D., & Wainer, H. (2001). *Test Scoring*. New Jersey: Lawrence Erlbaum Associates.

Tseng, C., Doll, J., & Varma, K. (2019). Exploring Evidence that Board Games can Support Computational Thinking. *Proceedings of International Conference on Computational Thinking Education 2019*, 61–64. Hong Kong: The Education University of Hong Kong.

Utting, I., Cooper, S., & Kölling, M. (2010). Alice, Greenfoot, and Scratch-A Discussion. *ACM Transactions on Computing*, *10*(4), 1–11. https://doi.org/10.1145/1868358.1868364

Vahldick, A., Mendes, A. J., Marcelino, M. J., & Roberto, P. (2016). Computational thinking Practiced with a Casual Serious Game in Higher Education. *CSBC 2016 - XXIV WEI*, (July).

Vallance, M., & Towndrow, P. A. (2016). Pedagogic transformation, student-directed design and computational thinking. *Pedagogies*, *11*(3), 218–234. https://doi.org/10.1080/1554480X.2016.1182437

van der Linden, W. J. (2015). *Handbook of Item Response Theory Modeling Volume One*. https://doi.org/10.4324/9781315736013

Violato, C. (1991). Item difficulty and discrimination as a function of stem completeness. *Psychological Reports*, *69*(3 Pt 1), 739–743. https://doi.org/10.2466/pr0.1991.69.3.739

Vygotsky, L. (1986). Thought and language. Translation newly revised and edited by Alex Kozulin. In *Behavioural Neurology* (Vol. 5). https://doi.org/10.3233/BEN-1992-5106

Weber, G. (2012). ACT - Adaptive Control of Thought. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 60–62). https://doi.org/10.1007/978-1-4419-1428-6_533

Weese, J., & Feldhausen, R. (2017). STEM Outreach: Assessing Computational Thinking and Problem Solving. *Proceedings of the 2017 American Society for Engineering Education Annual Conference & Exposition*. https://doi.org/10.18260/1-2--28845

Weese, J. L. (2016). Mixed Methods for the Assessment and Incorporation of Computational Thinking in K-12 and Higher Education. *Icer*, 279–280.

Weinberg, G. M. (1998). *The psychology of computer programming* (Silver Ann). New York: Dorset House Publishing Co., Inc.

Weintrop, D., Beheshti, E., Horn, M., Orton, K., Jona, K., Trouille, L., & Wilensky, U. (2016). Defining Computational Thinking for Mathematics and Science Classrooms. *Journal of Science Education and Technology*, *25*(1), 127–147. https://doi.org/10.1007/s10956-015-9581-5

Weintrop, D., Orton, K., Horn, M., Beheshti, E., Trouille, L., Jona, K., & Wilensky, U. (2015). *Computational Thinking in the Science Classroom: Preliminary Findings from a Blended Curriculum*. 5. Retrieved from https://ccl.northwestern.edu/papers/2014/Weintrop_et_al-NARST_2015.pdf

Weintrop, D., & Wilensky, U. (2013). Robobuilder: a computational thinking game. *Sigcse*, 736. https://doi.org/10.1145/2445196.2445430

Werner, L., Denner, J., & Campe, S. (2015). Children Programming Games: A Strategy for Measuring Computational Learning. *Trans. Comput. Educ.*, *14*(4), 24:1--24:22. https://doi.org/10.1145/2677091

Werner, L., Denner, J., Campe, S., & Kawamoto, D. C. (2012). The Fairy Performance Assessment : Measuring Computational Thinking in Middle School. *Proceedings of the 43rd ACM Technical Symposium on Computer Science Education - SIGCSE '12*, 215–220. https://doi.org/10.1145/2157136.2157200

Wesolowsky, G. O. (2000). Detecting excessive similarity in answers on multiple choice exams. *Journal of Applied Statistics*, *27*(7), 909–921. https://doi.org/10.1080/02664760050120588

Wiebe, E., Mott, B. W., London, J., Boyer, K. E., Aksit, O., & Lester, J. C. (2019). Development of a lean computational thinking abilities assessment for middle grades students. *SIGCSE 2019 - Proceedings of the 50th ACM Technical Symposium on Computer Science Education*, 456–461. https://doi.org/10.1145/3287324.3287390

Wilensky, U., Brady, C. E., & Horn, M. S. (2014). Fostering computational literacy in science classrooms. *Communications of the ACM*, *57*(8), 24–28. https://doi.org/10.1145/2633031

Wiliam, D. (2014). *Principled assessment design*. London: SSAT Ltd.

Williamson, B. (2015). *Coding/learning: Software and digital data in education. A report from the*

ESRC Code Acts in Education Seminar Series.

Williamson, B. (2016). Political computational thinking: policy networks, digital governance and 'learning to code.' *Critical Policy Studies*. https://doi.org/10.1080/19460171.2015.1052003

Willingham, D. T. (2012). Why Don't Students Like School?: A Cognitive Scientist Answers Questions About How the Mind Works and What It Means for the Classroom. In *Why Don't Students Like School?: A Cognitive Scientist Answers Questions About How the Mind Works and What It Means for the Classroom*. https://doi.org/10.1002/9781118269527

Wing, J. (2006). Computational Thinking. *Communications of the Association for Computing Machinery (ACM)*, *49*(3), 33–35. https://doi.org/https://www.cs.cmu.edu/~15110-s13/Wing06-ct.pdf

Wing, J. (2010). Progress in CT and Expanding the HPC Community. *COMMUNICATIONS OF THE ACM*, *59*(7), 10–11. https://doi.org/DOI:10.1145/2933410

Wing, J. (2011). Research notebook: Computational thinking—What and why? *The Link Magazine*.

Wolz, U., Stone, M., Pearson, K., Pulimood, S. M., & Switzer, M. (2011). Computational thinking and expository writing in the middle school. *ACM Transactions on Computing Education*, *11*(2). https://doi.org/10.1145/1993069.1993073

Wood, R. (1987). *Measurement and Assessment in Education and Psychology*. Lewes: Falmer Press.

Woodford, K., & Bancroft, P. (2005). Multiple choice questions not considered harmful. *Seventh Australasian Computing Education Conference (ACE '05)*, *42*, 109–116.

Xu, X., Kauer, S., & Tupy, S. (2016). Multiple-Choice Questions : Tips for Optimizing Assessment In-Seat and Online. *Scholarship of Teaching and Learning in Psychology*, *2*(2), 147–158. https://doi.org///dx.doi.org/10.1037/stl0000062

Yadav, A., Hong, H., & Stephenson, C. (2016a). Computational Thinking for All: Pedagogical Approaches to Embedding 21st Century Problem Solving in K-12 Classrooms. *TechTrends*, 1–4. https://doi.org/10.1007/s11528-016-0087-7

Yadav, A., Hong, H., & Stephenson, C. (2016b). Computational Thinking for All: Pedagogical Approaches to Embedding 21st Century Problem Solving in K-12 Classrooms. *Tech Trends*, 1–4. https://doi.org/10.1007/s11528-016-0087-7

Yadav, A., Zhou, N., Mayfield, C., Hambrusch, S., & Korb, J. T. (2011). Introducing computational thinking in education courses. *Educational Studies*, (2), 465–470. https://doi.org/10.1145/1953163.1953297

Yakavets, N. (2013). Reforming society through education for gifted children: the case of Kazakhstan. *Research Papers in Education*, *29*(5), 513–533. https://doi.org/10.1080/02671522.2013.825311

Yildiz, H. D., Yilmaz, F. G. K., & Yilmaz, R. (2017). Examining the Relationship between Digital Game Preferences and Computational Thinking Skills. *Contemporary Educational Technology*, *8*(3), 359–369.

Zapata-Caceres, M., Martin-Barroso, E., & Roman-Gonzalez, M. (2020). Computational thinking test for beginners: Design and content validation. *IEEE Global Engineering Education Conference, EDUCON*, *2020-April*, 1905–1914. https://doi.org/10.1109/EDUCON45650.2020.9125368

Zeidner, M. (1987). Essay versus Multiple-Choice Type Classroom Exams: The Student's Perspective. *Journal of Educational Research*, *80*(6), 352–358. https://doi.org/10.1080/00220671.1987.10885782

Zhong, B., Wang, Q., Chen, J., & Li, Y. (2015). An Exploration of Three-Dimensional Integrated Assessment for Computational Thinking. *Journal of Educational Computing Research*, *53*(122), 0735633115608444. https://doi.org/10.1177/0735633115608444

Zhussipbek, T. (2019). *Students' and teachers' experiences and perceptions of studying in "Bilim-Innovation" lyceum-boarding schools in Kazakhstan*. The University of Sheffield.

Zimmerman, D. W., & Williams, R. H. (2003). A New Look at the Influence of Guessing on the Reliability of Multiple-Choice Tests. *Applied Psychological Measurement*, *27*(5), 357–371. https://doi.org/10.1177/0146621603254799

Zur-Bargury, I., Pârv, B., & Lanzberg, D. (2013). A nationwide exam as a tool for improving a new curriculum. *Proceedings of the 18th ACM Conference on Innovation and Technology in Computer Science Education - ITiCSE '13*, 267. https://doi.org/10.1145/2462476.2462479