HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Educational Sciences
Measurement and Evaluation in Education

INVESTIGATING THE VARIABLES AFFECTING STUDENTS' PERCEPTIONS OF 21ST CENTURY SKILLS AND PROBLEM-SOLVING SKILLS

Buket EREN JANSSEN

Ph.D. Dissertation

Ankara, (2022)

With leadership, research, innovation, high quality education and change,

To the leading edge... Toward being the best...

HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Educational Sciences
Measurement and Evaluation in Education

INVESTIGATING THE VARIABLES AFFECTING STUDENTS' PERCEPTIONS OF 21ST CENTURY SKILLS AND PROBLEM-SOLVING SKILLS

ÖĞRENCİLERİN 21. YÜZYIL BECERİ ALGILARI VE PROBLEM-ÇÖZME BECERİLERİNE ETKİ EDEN DEĞİŞKENLERİN İNCELENMESİ

Buket EREN JANSSEN

Ph.D. Dissertation

Ankara, (2022)

Abstract

The aim of this study is to investigate the effects of programme (IB vs non-IB) on students' perceptions of 21st century skills (PoTCS) and problem-solving skills (PSS) in relation with English comprehension, age, achievement, and gender. It was designed as a correlational study. Research questions were evaluated based on a theoretical model. Target group of the study was lower secondary school students who follow an IB programme and another programme in Netherlands. Measurement instruments (PoTCS scale, PSS test and Demographic information questionnaire) were developed by the researcher. PoTCS scale was formed with hypothetical short scenarios as a situational judgement test items with 21 items under four factors. There were 379 students attended to the pilot study of PoTCS scale and 304 different students attended to the main study to conduct CFA and SEM. Model showed a good fit after few modifications without disturbing the initial theory, x^2(29, N=287) = 34.755, p=.213, CFI=.988, RMSEA=.026. According to the findings, programme had a significant direct effect on two dimensions of PoTCS and English comprehension in favor of IB students, and significant effect on PSS in favor of non-IB students. English comprehension appeared as a significant mediator between programme and PoTCS, whereas an insignificant mediator between programme and PSS. Achievement did not strengthen the effect of PoTCS on PSS. Effects of variables affecting PoTCS or PSS did not differ by gender and only using information and technology effectively had a significant difference in terms of the duration in IB in favor of longer enrollment.

Keywords: 21st century skills, problem-solving, international baccalaureate, situational judgement tests, structural equation modeling

Öz

Bu çalışmanın amacı öğrencilerin takip ettikleri müfredatın (IB ve IB olmayan) 21. yüzyıl beceri algıları (YYBA) ve problem çözme becerileri (PÇB)'ne olan etkisini ve bu değişkenlerin İngilizce kavrama düzeyleri, yaş, başarı ve cinsiyet değişkenleriyle ilişkilerini incelemektir. Bu araştırma ilişkisel bir araştırma olarak desenlenmiştir. Araştırma soruları, kurulan kuramsal model çerçevesinde değerlendirilmiş ve yapısal eşitlik modellemesi ile test edilmiştir. Bu araştırmanın hedef kitlesini Hollanda'da 12-16 yaş aralığında IB programını takip eden ve etmeyen öğrencilerden oluşmaktadır. Kullanılan ölçme araçları (YYBA ölçeği, PÇB testi ve Kişisel bilgi anketi) araştırmacı tarafından geliştirilmiştir. YYBA ölçeği varsayımsal kısa senaryolar (durumsal yargı testi) şeklinde hazırlanmış, 21 madde ve dört faktörden oluşmuştur. Ölçek geliştirme sürecinde, açımlayıcı faktör analizi için veri toplanan pilot çalışmaya 379, doğrulayıcı faktör analizinde kullanılan ve modelin test edildiği ana çalışmaya ise pilot çalışmadan farklı 304 öğrenci katılmıştır. Araştırma sorularına yanıt aranacak olan kuramsal model test edilmiş ve yapılan birkaç modifikasyondan sonra model iyi uyum göstermiştir (x^2(29, N=287) = 34.755, p=.213, CFI=.988, RMSEA=.026). Araştırma bulguları, program türünün YYBA ölçeğinin iki boyutuna ve İngilizce kavrama düzeylerine IB öğrencileri lehine, PÇB'ne ise IB olmayan öğrencilerin lehine anlamlı etkisi olduğunu göstermektedir. İngilizce kavrama düzeyinin program ve YYBA arasında anlamlı bir aracıyken, program ve PÇB arasında anlamlı bir aracı olmadığı görülmüştür. Başarı YYBA'nın PÇB'ne etkisini artırmamıştır. YYBA ve PÇB'ne etki eden etmenlerin cinsiyete göre farklılaşmadığı ve IB programına devam etme süresinin sadece bilgi ve teknolojiyi etkin kullanma boyutunda ve uzun süre IB programına katılanların lehine farklılaştığı görülmüştür.

Anahtar sözcükler: yirmibirinci yüzyıl becerileri, problem çözme, Uluslararası Bakalorya, durumsal yargı testleri, yapısal eşitlik modellemesi

Acknowledgements

This dissertation has been a much more challenging journey for me than I could ever imagine. Luckily, I have had many companions and lovely people who have helped me through this difficult journey. I would like to thank you all in advance, in case I cannot mention each one of you here.

I feel myself very lucky to have my advisor Assoc. Prof. Dr. Burcu Atar by my side. Despite the 3000 km distance between us, she had always given me the hope to complete my study and motivation to pursue my dream. Especially, when I had a lot of struggles in collecting data and finding the strength to continue during Covid-19 epidemic. Thank you Dr. Atar for always being available and giving me all your support. I am extremely grateful to Prof. Dr. Şeref Tan and Prof. Dr. Selahattin Gelbal for their ideas and suggestions to improve my dissertation during our committee meetings. I also want to thank to Assoc. Prof. Dr. Dilara Bakan Kalaycıoğlu and Dr. Derya Çobanoğlu Aktan for their suggestions and beautiful words about my dissertation.

I would like to thank to all school managers who helped me during the data collection process in the Netherlands. Especially Ger Rombouts, I cannot thank you enough for your support and network that you helped me to reach.

I would like to thank to my family and particularly to my husband Jop Janssen for sharing me with my dissertation and sharing all the house chores as favoring me. I have to thank to my five year old daughter Ella and my seven month baby girl Mila for being there and smiling to me each day.

Besides all people who were there during not only hopeful but also desperate times, there was one person which made this dissertation possible to be completed... my mother, Aygün Eren. Mothers help their children all the time, but I feel myself very privileged to have my mom. She had stayed with us in the Netherlands, in a country which she does not know the language and has very limited social life for 90 days to look after my daughters. Therefore, thank you mom! You are an exclusively strong and special woman!

I would like to dedicate this dissertation to all strong women who made this dissertation possible… Especially my mother and my daughters Ella and Mila…

Table of Contents

List of Tables

List of Figures

Symbols and Abbreviations

ATC21S: The Assessment and Teaching of 21st-Century Skills

CFA: Confirmatory Factor Analysis

CAS: Creativity, Action, Service

CAT: Computer Adaptive Testing

CP: Career Programme

DP: Diploma Programme

EE: Extended Essay

EU: European Union

EFA: Exploratory Factor Analysis

IB: The International Baccalaureate

IBDP: The International Baccalaureate Diploma Programme

IBO: The International Baccalaureate Organisation

ICT: Information and Communications Technology

MoNEP: The Ministry of National Education Programme

MYP: Middle Years Programme

OECD: The Organisation for Economic Co-operation and Development

PAF: Principal Axis Factoring

PCA: Principal Component Analysis

P21: Partnership For 21st Century Learning

PoTCS: The Perception of 21st Century Skills

PYP: Primary Years Programme

SEM: Structural Equation Modeling

SJT: Situational Judgment Tests

TOK: Theory of Knowledge

Chapter 1
Introduction

In this section, statement of the problem, aim and significance of the study, research question, sub-research questions, assumptions, limitations and definitions are stated.

Statement of the Problem

Last decades, humanity is in a constant race to catch the improvements in technology and effects of them in daily life. Involved competencies for daily routine as well as expected skills for many professions have been evolved which has brought the requirement of change in how we perceive education. Improvements in societies and economies bring the need for educational systems to equip young people with new skills and capabilities. Future generations need to adapt to new forms of socialization and have an active part in economic development of a knowledge society (Ananiadou & Claro, 2009). As Serdar (2015) stated “Our educational system continues to teach our students to memorize information and sort facts, rather than how to learn” (p.8), and this is not only about how the curricula are structured but also about the assessment.

Why our students need to gain many more skills compared to only previous generation? Surely, there is no easy answer, whereas it can be summarized briefly as Care, Kim and Scoular (2017) stated that our students have to be prepared for an uncertain world which has many concerning problems and conflicts such as environmental issues, economic growth, a possible nuclear war (Dulun, 2018) or an unexpected pandemic like Covid-19 which we have been experienced since the end of 2019.

As educators, we cannot ignore the developments in the world around us and its expectations from teaching and learning processes. Students need to be prepared for a more challenging world. They are not expected only to be good at a specific topic. They are expected to be equipped with many skills such as being creative and quick problem-solvers. They need to transfer a knowledge or application in one case to another, analyze it, communicate it, reflect it, work collaboratively with others as using the proper technological support. Information is

not important by itself anymore, because information is already on most people's fingertips. The question of the new century is more about how to use that information in real-life problems or how to differentiate information critically. Some of these necessary skills needed for the new era can be considered as higher-order thinking skills (Haladyna, 1997).

Higher-order thinking has been already addressed by educators and researches for many years (e.g. Thorndike, 1931; Haladyna, 1997). Nowadays, as a common understanding, these higher-order thinking processes are included in the skills which are named as "21st century skills" as well. But there are more skills involved in the 21st century skills than only thinking skills, such as social skills and affective skills (Kang, Heo, Jo, Shin & Seo, 2010). There are some discussions about definitions of terms "skills" and "competence" (Ananiadou & Claro, 2009), though this is not going to be within the scope of this research and all behavioral and thinking expectations will be called as 'skills' through the current study.

There are many international organizations and groups from all over the world already designated their focus to define 21st century skills and improve ways to assess them since the early 2000s. However, the first thing to recognize is that there is not a consensus about the classification of 21st century skills. The most commonly emphasized skill sets can be listed according to Dede (2009)'s summary as critical thinking and problem solving, interactive communication, collaboration and team work, creativity and innovation, information and literacy, and technology operations.

A large-scale study conducted by Care, Kim and Scoular (2017) shows that 21st century skills like critical thinking and problem solving, social and affective skills or creativity have been already stated 76% of countries' national educational plans or their policy documents. On the other hand, findings suggested reflection of these skills on countries' curricula, teaching-learning or assessment processes fall behind (Care, Kim & Scoular, 2017). Therefore, we know pretty much what needs to be improved about our students, but how to put these necessary changes on action and how to assess the change are still the challenge we need to deal with.

Today's students are the ones who need to solve 21st century problems and this requires proper skills. Hence, as educators, we need to look for enriched curricula, teaching-learning processes and assessment techniques to prepare our students for future. Twenty first century skills are described as complex, cross-disciplinary, crucial for both school and life, much more demanding to teach and they are beyond rote memorization (Care, Kim & Scoular, 2017). For instance, in order to respond to complex problems, multidimensional skills would be needed and some factors might influence the performance in problem solving such as being familiar to the context of the problem (Bennett, Jenkins, Persky & Weiss, 2003).

Problem-solving, which is considered one of the 21st century skills by most of the organizations, expects individuals "to learn to think and solve problems like professionals in their field and to link theory with practice" (Edens, 2000, p.55). Problem-based learning (PBL) environments have been introduced in various age groups and for different programme types to prepare students for the current century's workplace. Yet, it still has some pitfalls especially because of the ill-structured nature of real-life problems and the challenge of integrating such tasks in classroom environments (Edens, 2000). Teaching other 21st century skills have the similar drawbacks. Integrating activities to support creativity or strengthen communication and collaboration skills across different subjects and measuring these skills validly and reliably are still areas to improve in many countries.

There have been a great number of studies and compilation of these studies conducted with the aim of defining and assessing the qualifications of 21st century skills during the last decade. Some examples of these studies can be listed as, Kozma (2009), Kang, Heo, Jo, Shin and Seo (2010), Kyllonen (2012) and Dulun (2018). Diversified theoretical frameworks have been formed as a result of these researches and based on these frameworks, various scales have been developed.

One of these frameworks classifies 21st century skills based on educational performance under three domains (cognitive, affective and sociocultural) each with four factors (Kang, Heo, Jo, Shin & Seo, 2010). Based on this framework, Kang, Heo, Jo, Shin and Seo, (2010) formed a 33 item and four-point perception scale. Items are formed as one sentence and expect respondents to choose out of four

options as considering which fits best to his or her perceived competencies. “When I study, I collect necessary data” (cognitive domain, information management subdomain) or “When I did something dishonest, I try to rectify it” (affective domain, self-value subdomain) or “I am usually nice to new students in the class” (sociocultural domain, social receptivity subdomain) (Kang et al., 2010, p.168) are some sample statements from the scale.

Another study was conducted by Kaya (2017) in order to determine the relation of high school students’ 21st century skills with their burnout and school engagement levels. Through this study, author had developed a 19 item seven-point Likert-type scale. Author limited the skills included into the research as critical thinking, problem solving, communication and collaboration, digital literacy, creativity and innovation as considering common factors from literature and general aims of the Turkish Education system. Similar with Kang et al. (2010)’s study, items are stated as one sentence, such as “I can analyze complex problems” or “I can solve possible conflicts between friends during group work” (Kaya, 2017, p.173).

First of all, these scales measure the perception of people about their 21st century skills, instead of 21st century skills, which is something generally missing in the titles. Secondly, types of items form these scales are easy to create and prompt to answer, but the question of how valid indicators they are for assessing the perception of 21st century skills need to be discussed. Twenty first century skills have been described as complex, difficult to teach, and most importantly very critical to measure competencies (e.g. Baker 2007; Care, Kim & Scoular, 2017). This complexity brings the struggle of understanding these skills well and so assessing people’s perceptions about these skills. Hence, one of the reference questions of the current study is “Can the perception of such a complex skill set which requires deeper learning as well as social and affective behaviors be measured as directing single, direct sentences?”.

Most of the scales developed or used in education and psychology are self-report. Self-report scales are much practical to apply and economic to obtain data compared to other methods such as observation or interview. Self-report data can be collected in days whereas observing a group of people can take months. On the other hand, self-report scales have been criticized in some ways especially in

terms of validity. One of the critics is even though self-reports are given anonymously, people often might be biased about their own experiences. This is originated mostly from social desirability concerns with or without awareness. Social desirability biases can cause ‘over reporting’ of socially desirable or ‘under reporting’ of socially undesirable behaviors and these can cause validity problems (Salters-Pedneault, 2018).

In this respect, it should be considered that when respondents perceive less about what is right or what is wrong, and when items are given more related with contextual experiences, less biased answers might be provided by the respondents. Based on this hypothesis, forming a scale with short-scenarios might measure people’s perception about the considered skills more validly and reliably. Such a format has existed for about last thirty years named as situational judgment tests (SJTs) (Weekley & Jones, 1999). Early initiatives of SJTs were about observing examinee’s actual behavior or reaction in a real circumstance and assessment centers used to implement those tests (Weekley & Jones, 1999). However, those were not the most cost efficient or convenient approaches.

Lately used situational judgment tests (SJTs) are formed by hypothetical small stories and multiple-choice answer options in terms of the reaction the respondent would give. SJTs differ in terms of the number of responses they offer to the respondents as single-response situational judgment tests and multiple-response situational judgment tests. Literature suggests single-response situational judgment tests having comparable internal consistency, convergent validity, and predictive validity estimates with multiple-response situational judgment tests (Martin-Raugh, et al., 2018). SJTs are the most commonly seen in use for predicting applicants’ job performance and many studies had been conducted to validate those tests in terms of construct and predictive validity (e.g. Bess, 2001; König, et al., 2007; Lievens & Sackett, 2012; Ron, 2019).

There has been validity and reliability issues about SJTs since they were introduced (Ron, 2019; Sorrel, 2016). Bess (2001) investigated the underlying constructs for situational judgment tests in order to bring an explanation for the validity problems of past studies about SJT. As a conclusion, SJTs’ multidimensional nature was recognized as the possible reason of previous inconsistent validity measures about SJTs (Sorrel et al., 2016).

Different forms of a questionnaire or scale is one way of measuring the perception of 21st century skills, but assessing the skills classified as 21st century skills need more time and planned observations. Integrating assessment processes of 21st century skills into classroom environment and assessment is the ultimate goal of today's educational systems. As Kutlu (2006) stated, higher order thinking skills, such as problem solving, reading comprehension, critical thinking, creativity, decision-making, need to be assessed with alternative assessment techniques. Kutlu, Doğan and Karakaya (2010) suggest three alternative assessment techniques. These are performance-based assessment, portfolio assessment and authentic assessment. Although measuring higher order thinking skills has been already a long-time discussion, it is still not easy to measure higher order thinking skills reliably and cost efficiently. In addition, assessment techniques such as these three mentioned assessments are hard to scale (Silva, 2009). Currently, one of the educational organizations, International Baccalaureate Organization (IBO), provides assessments which can be comparable with these alternative assessment techniques in its schools.

IBO is a non-profit worldwide community of schools who share the view of offering a challenging but balanced education since 1968. IBO designs its programmes with a deeper aim of supporting a global understanding, tolerance and respect for others. International Baccalaureate (IB) programmes claim to provide an education which enables students to understand "the complexities of the world around them, as well as equipping them with the skills and dispositions needed for taking responsible action for the future" (IBO, 2017, p.1) which has a particular emphasis on 21st century education. IB describes its learner profile with 10 characteristics with the motto: "We strive to be…" and these characteristics are listed as inquirers, knowledgeable, thinkers, communicators, principled, open-minded, caring, risk-takers, balanced and reflective. IB considered necessary skills for 21st century in order to form its learner profile. IB aims to fosters the improvement of skills for inquiry, communication, social commitment, intercultural empathy and respect in a global context which are fundamental qualities for young people who will lead but also be part of the future (IBO, 2014).

IB schools do not select their students via acceptance tests, yet they aim all of their students apart from their backgrounds to improve on IB learner profile,

which is a challenging task to achieve. Suldo, Shaunessy and Hardesty (2008) stated this issue together with its reward as IB programmes are certainly demanding and challenging programmes not only for students but also for its educators yet they have immense benefits. IB programmes intend to raise students who are equipped with 21st century skills (IB Africa, Europe & Middle East regional conference, 2013).

Education organizations and systems have been progressively placing importance on the ways of integrating teaching with assessment. This practice is stated on an IBO report as “Assessment of any kind should ultimately improve learning” (Toe et al., 2015, p. 9). Related with such an effort, IB follows the vision of, “Meaningful assessment supports curricular goals” which emphasizes curriculum and assessment need to be hand in hand and assessment is considered as “an ongoing, varied and integral to the curriculum” (IBO, 2017, p.5).

In IB schools’ assessment processes and instruments vary as curriculum do according to age groups. As considering the interest of current study, 11-16 years old students (attending MYP-Middle Years Programme) follows a criterion-based assessment which includes four different criteria to measure for each core subject. As an example, assessment in Mathematics is classified into four different categories as; criterion A knowing and understanding, criterion B investigating patterns, criterion C communicating and criterion D applying mathematics in real-life context. The ultimate subject grades/scores are defined as considering these four criteria together. Additionally, each criterion grade is not stating solely as a number but also shows what kinds of skill levels students could achieve for that criterion. For instance, for criterion A, students need to be able to solve complex problems related with the topic for familiar and unfamiliar situations and if students can only reach to solve complex problems in familiar situations, then they cannot reach to the highest levels which includes unfamiliar situations as well. The importance of solving problems in unfamiliar situations can be considered better when we think about the requirements of the 21st century.

In IB system, starting from the core philosophy, teaching-learning activities and assessment processes organized as considering necessary 21st century skills and intending to develop them. This fact brings the question to the current study

that if students from IB and non-IB schools differ from each other in terms of their problem-solving skills and perceptions of 21st century skills.

Aim and Significance of the Study

The aim of this study is to investigate the effect of programme type (IB vs non-IB) on students' perceptions of 21st century skills and problem-solving skills in relation with students' English comprehension levels, age, Mathematics achievement, English achievement and gender. As investigating IB and non-IB schools, it is expected to reveal whether IB MYP-middle years programme and integrated assessment processes contributes positively to students' perceptions of 21st century skills. IBO has created its core, which includes the learner profile and assessment processes of its curricula considering 21st century skills. This is considered as a crucial contribution for the international educational parties to find out if schools and countries' national education teams can learn from IB system.

Problem-solving, but especially complex problem-solving is considered as one of the crucial skills expected from 21st century citizens. Solving a problem can be considered as very related and coexisting with other 21st century skills. For instance, problem-solving may require collaboration with others, communication with different sources, being self-directed, creative and innovative and so on. Hence, it is important to assess problem-solving skills separately and investigate the relation between students' level of problem-solving skills and their perception about their 21st century skills. Additionally, together with the results of problem-solving skills test, it is intended to get a more objective and concrete information to assess the effect of programme type (IB vs non-IB) in relation with the perception of 21st century skills.

Students might benefit differently from educational systems according to the different characteristics they already have or they develop in time. Additionally, some characteristics, such as academic achievement, may be resulted from or may lead to different skills. Therefore, the perception of 21st century skills and problem-solving skills will be considered in relations with English comprehension, age, Mathematics and English achievements and gender. English comprehension is considered as students' own perception about their English level and it was stated as not so good, good, very good and mother tongue. On the other hand,

English achievement is students' English grade, which is an academic achievement. These variables considered important to include into the current study. Better English comprehension might mean developing better 21st century skills as being able to reach more information, communicate and collaborate with more people, when it is considered that English is the most spoken and most used internet content language (Johnson, 2021; Szmigiera, 2021). Mathematics and English achievements are considered as the indicators of academic achievement for the current study. As considering, "Learning academic content is fundamental to education, and mastery of such content serves as the basis for higher-order thinking skills as well as the impetus for improved interpersonal and intrapersonal competencies.", (Soland, Hamilton & Stecher, 2013, p.4), Mathematics and English achievements are assessed as the moderator effects in the research model .

Throughout each year, many students begin to attend IB schools at different times of academic year because of expert immigrations. Some students transfer from other IB schools whereas others come from non-IB schools which follow the national curriculum of the country of departure. As the nature of each educational process, it might be expected that the longer students attend to an educational programme, the more likely they benefit from it. Hence, another aim is to investigate IB students' perceptions of 21st century skills in terms of the number of years that students attend to an IB programme.

As considering international and Dutch schools in the study group, it is aimed to bring an international view to the matter of assessing 21st century skills, as well as to diversify the study group of the current research. As Baker (2007) remarked "With collaboration of both the international community and our own communities, we can enable education to prepare our students far better for the future" (p.315).

Additionally, creating a valid and reliable scale to measure 12 to 16 years old students' 21st century skills perceptions can be considered as a sub-purpose of the current study. It is intended to form the scale based on real-life related hypothetical short-scenarios similar with situational judgment tests. It is experienced and as supported by Popham (2003, p.101), "the more that student's assessment task resembles the task to be performed by people in real life" the

more “it is an attempt to measure a student’s mastery of a high-level”. In other words, the more a scale is real life-related the more it can determine higher level skills and this can contribute for a valid measurement. Hence, with this study, it is aimed to make a contribution to the practice of assessment of the perception of 21st century skills of 12 to 16 years old students. The age group preference is based on the OECD age interval for “lower secondary school” (Ananiadou & Claro, 2009).

Although IB schools aim each IB learner to master 21st century skills, this goal is becoming an increasingly common goal for all the world schools. Thus, it is believed that creating a 21st century skills scale based on short-scenarios will provide a very useful tool for not only IB schools but also any school to track their students’ perception of their 21st century skills. Such a scale can be useful to monitor which skills or skill sets need to be improved by an individual or group of students over time. It is necessary to assess students’ 21st century skills and students’ perception about themselves in terms of these skills for observing the outcomes of curriculum and assessment reforms.

Research Questions

The main research question guides this study is “How are students’ perceptions of 21st century skills and students’ problem-solving skills affected by the programme type which students follow (IB vs non-IB) in relation with English comprehension, age, achievement (English and Mathematics scores) and gender?”

In the direction of the main research question, following sub-questions are examined:

Sub-research questions

1. a) How are the 21st century skill perceptions of students who follow an IB programme?
   b) How are the 21st century skill perceptions of students who do not follow an IB programme?
2. a) How are the problem-solving skills of students who follow an IB programme?

b) How are the problem-solving skills of students who do not follow an IB programme?

3. Is there a significant relation between the programme type students attend and their;
   a) perceptions of 21st century skills,
   b) problem-solving skills,
   c) English comprehension,
   d) perceptions of 21st century skills with the mediation effect of English comprehension,
   e) problem-solving skills with the mediation effect of English comprehension?
4. Is there a significant relation between students' problem-solving skills and their;
   a) age,
   b) English comprehension,
   c) perceptions of 21st century skills,
   d) perceptions of 21st century skills with the moderation effect of Mathematics achievement,
   e) perceptions of 21st century skills with the moderation effect of English achievement?
5. Do the effects of variables affecting students' perceptions of 21st century skills differ according to gender?
6. Do the effects of variables affecting students' problem-solving skills differ according to gender?
7. Are there significant differences between IB students' perception of 21st century skills according to the number of years that students attend to the IB programme?

Based on the sub-research questions (3, 4, 5 and 6) related with the perception of 21st century skills (PoTCS), problem solving skills (PSS), programme

type (IB vs non-IB), age, and English comprehension, following model shown in *Figure 1* will be described and tested. Gender is considered as the multigroup variable and Mathematics and English achievement variables are considered as moderators.

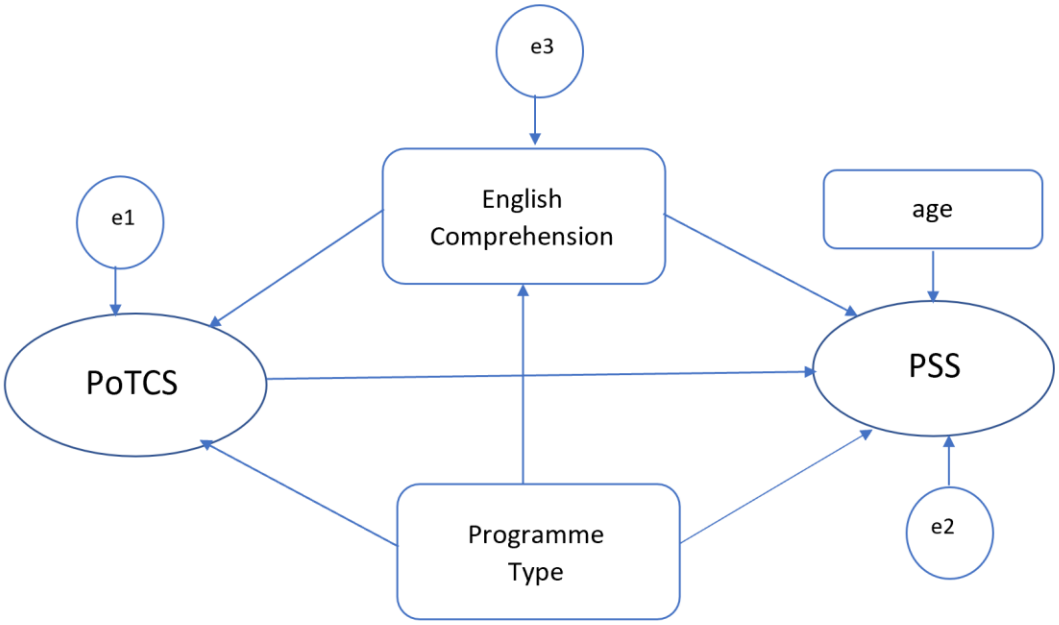


Figure 1. Theoretical model

Assumptions

Throughout this research, it is assumed that;

1. All students included into the study group have above a certain level of English language comprehension and they answer to all items properly.
2. When respondents read the options, they understand the same level of agreement from "Totally suits me, Somewhat suits me, Neutral, Somewhat not suits me, Totally not suits me".

Limitations

- Current study is limited with its data collection instruments.
- Study groups of the current research are limited in terms of the acceptance procedure of different school types. For instance, tto (tweetalig onderwijs-bilingual education) schools are more homogenous schools in terms of

students' academic achievements, whereas international schools are heterogenous.

- Data collection process are limited, because the data collection process of this research had been carried out during the Covid-19 epidemic. Data collection had taken much longer than planned and yet not all the expected data could be collected because of distance learning, limited lesson hours (20-40 minutes schedules instead of 50-60 minutes), high number of missing students in the lessons, uncertainty in educational institutes and etc.

Definitions

Some of the definitions related with the study are stated below.

21st century skills: New skills and dispositions, which are needed by people to correspond and contribute to the current century (Saavedra & Opfer, 2012)

Achievement: Level of success in school subjects. For the current study achievement is measured according to students' English and Mathematics scores stated on their last report.

IB Schools: Any school which offer a programme of International Baccalaureate and gain IB world school status (IBO, 2014).

Non-IB Schools: Schools which are not following IB curriculum and are not certified as IB school.

IB Middle Years Programme: IB education programme which is designed for 11 to 16 years old students.

IB Diploma Programme: IB education programme which is designed for 16 to 19 years old students.

Problem-Solving Skills: Competency to solve problems which are not clearly related to school curriculum and mostly require people to deal with unfamiliar situations when the solution is not obvious as thinking original ways for solution (OECD, 2004).

Chapter 2
Literature Review

21st Century Skills

The rapid change and transformation in technology have brought a massive shift in other areas as well. Social relations, ways of reaching information, types of problems people face with in a normal day, qualities to be required for getting a prosperous job, descriptions of professions are only some of the examples of these changes. The skills which can adapt people to the changing conditions and requirements related to digital age is called 21st century skills. Some (problem-solving, critical thinking) of these skills have been already well known and studied for the 20th century as well, whereas some others are new and hard to integrate into the education and even harder to assess (creativity, ICT literacy) (Voogt, Dede & Erstad, 2009). When the requirements of life have been evolving, education cannot insist to follow the ways of previous age. We need to educate our students to prepare them for a new century with new skill requirements. Since assessments have been a big impact on how the education being shaped, assessment processes need to be adapted as well.

Assessment is about the outcomes of the educational processes. Unfortunately, organizing the educational processes based on the assessments is still happening even for high achieving countries of internationally trend exams like PISA or TIMSS (Kang & Keinonen, 2016), whereas the other way around has to be intended. An example that can be given about this from Turkey might be the 'new generation' questions. Although this question type was introduced only since 2019 in Turkey, most of the educators have already forgotten to ask "Why we need new generation questions? What do they aim to measure?" The aim is still about teaching a way to solve these questions, whereas they should have only represented the elements of a measurement tool and measure and help assessing the outcome of a teaching-learning process. Therefore, these approaches clearly show that we have not understood the nature of the 21st century skills, why do we need them and why we cannot continue to arrange our school environments only depending on solving multiple choice tests anymore. It does not matter if they are called new generation or not.

One of the main questions Ananiadou and Claro (2009) asked with their study was “What types of assessment are appropriate for the monitoring and evaluation of 21st century skills and competencies? How can they be developed?” (p.17). Kyllonen (2012)’s review study shows that rating scales are the most commonly used assessment method to assess the 21st century skills, despite their limitations. On the other hand, there are also important innovations in assessment. For instance, *anchoring vignettes* and *situational judgment tests* are methods to assess skills like collaboration or creativity which are harder to assess, which can be considered more to assess 21st century skills.

Assessment can only work when the right characteristics or structures are measured, so in order to assess any skill correctly, it is substantial to have correct operational definitions (Griffin, McGaw & Care, 2012). Soland, et al. (2013) mentioned this as “It is important to have a precise definition for a given competency when trying to measure it,…” (p.4). In this sense, initially literature about defining 21st century skills are given for different frameworks and common skills of these frameworks.

Defining 21st Century Skills and Related Frameworks

In the present day, getting a fixed profession and work on it until the retirement or having one clear job description and stick on that description is not the case anymore. Today’s employees are expected to be adapting to the changing technology and World, and modify to new job titles and descriptions. In order to do this transfer of mindset, employees are expected to have some skills instead of a fix job title. The skills or competencies which are needed for current and future generations are called 21st century skills. As Kozma (2009) stated 21st century skills are easy to list but a lot difficult to put into operation. Many international and national organizations and partnerships have been working on defining, developing and assessing 21st century skills for the citizens of future. As considering literature to date, it has appeared that there is not a consensus on stating or defining 21st century skills. This gap creates trouble in assessing those skills as well.

In the context of 21st century skills, Care, Kim and Scoular (2017) mentioned that “…although just one process in a problem-solving scenario might

be relatively easy to master, developing proficiency in complex and interrelated skills for students across a wide range of ability, is challenging" (p.35). In other words, students who will be the citizens of future will be expected to be equipped with more skills and achieve a lot more compared to previous generations. Additionally, they will need to find themselves a position in a world where technology is increasingly replacing human power.

Education and business parties are all involved into this process of describing skills needed for future (Griffin et al., 2012). Baker (2007) had compiled 21st century skills as: "adaptive problem solving, assessing and responding to risk, managing distraction and giving mindful, rotating attention to tasks, working alone with self-management and playing changeable roles in real or virtual teams and groups" (p. 313).

Kang, Heo, Jo, Shin and Seo (2010) formed one of the teams who aim to develop a conceptual framework of 21st century educational performance and develop a scale to measure 21st century educational performance in a valid and reliable way. Their study reveals that for future learners' educational performance requires competencies in cognitive, affective and sociocultural areas. The cognitive domain includes information management ability, knowledge construction ability, knowledge utilization ability and problem-solving ability. The affective domain includes self-identity, self-value, self-directedness, self-accountability. The sociocultural domain includes social membership, social receptivity, socializing ability and social fulfillment.

Kozma (2009) analyzed the range of 21st century skills which have been proposed by different organizations such as 21st Century Partnership, Lisbon Commission, ISTE etc. and summarized a set of core skills needed for 21st century as, "Creativity and innovation, critical thinking, problem solving, communication, collaboration, information fluency, technological literacy embedded in school subjects" (p.18).

Another framework is formed by Trilling and Fadel (2009) under three main categories as follows:

Learning and innovation skills:

- Critical thinking and problem solving
- Communication and collaboration
- Creativity and innovation

Digital literacy skills:

- Information literacy
- Media literacy
- Information and communication technologies (ICT) literacy

Career and life skills:

- Flexibility and adaptability
- Initiative and self-direction
- Social and cross-cultural interaction
- Productivity and accountability
- Leadership and responsibility (pp. xxvi)

Günüç, Odabaşı and Kuzu (2013) aimed to figure out how teacher candidates describe the characteristics of 21st century students. The study group was formed by students of Computer Education and Instructional Technologies department at Faculty of Education at a university in Turkey. They collected the data via Twitter as starting a topic as “student of the future” and running content analysis on tweets of 39 out of 92 students who had participated to the activity actively. Participation had been voluntarily; data collection process had been limited to five days and number of tweets participants could send was not limited. Participants had been interviewed by the help of semi-structured interview form as well in addition to the tweets they had posted. According to the findings, 21st century skills of student of the future are classified into four domains which are personal skills, research and acquisition of knowledge skills, creativity, innovation and career skills and technology skills.

Ananiadou and Claro (2009) aim to create a list of skills and competencies of 21st century skills as broad and comprehensive as possible. In order to reach this goal, they sent a questionnaire which is developed by CERI secretariat with the support of external experts and the Flemish Ministry of Education to all OECD

member countries. From the answers of 17 countries, one of the most prominent results was the general lack of understanding of concepts about 21st century skills and competencies. Additionally, Ananiadou and Claro (2009) described a framework for 21st century skills and competencies, which are categorized into three dimensions as information, communication, and ethics and social impact. These dimensions are related to several sub-dimensions and related skills are given in *Table 1* (Ananiadou & Claro, 2009, pp. 9-11).

Table 1

Summary of 21st Century Skills and Competencies

Dimension	Sub-dimension	Skills and Competencies
Information	Information as a source	• Information literacy • Media literacy • Research and inquiry
	Information as a product	• Creativity and innovation • Problem solving • Decision making
Communication	Effective communication	• Information literacy • Media literacy • Critical thinking • Communication
	Collaboration and virtual interaction	• Collaboration/team working • Flexibility and adaptability
Ethics and social impact	Social responsibility	• Critical thinking • Responsibility • Decision making
	Social impact	• Digital citizenship

Table 1 displays that some skills are repeating under different dimensions. The reason behind it is during Ananiadou and Claro (2009)'s study, one-to-one mapping was not the intention because of the inter and intra relations of the skills needed for lifelong learning.

The Global Cities Educational Network (GCEN) categorized the 21st century competencies (skills) into three main competencies as cognitive, interpersonal and intra personal (Soland, Hamilton & Stecher, 2013). Cognitive competencies were described with three sub-categories as academic mastery, critical thinking and

creativity. Interpersonal competencies were also defined under three sub-categories as communication and collaboration, leadership and global awareness. Intrapersonal competencies were considered as four sub-categories: growth mindset, learning how to learn, intrinsic motivation, and grit (p.4).

Another project is Partnership for 21st Century Learning (P21). P21 is a partnership which was created in 2002 in order to define needed skills for work, life and citizenship in 21st century with the input from teachers, education experts as well as business leaders ("Partnership for 21st Century Learning", 2019). According to P21 authentic assessment of 21st century skills is certainly the idealistic way, whereas in reality and in school context it is not easy to implement such assessments. One way of assessing 21st century skills is integrating these skills into the core subject assessments which some school systems (such as IBO schools) have been improved a lot on that although still there is not a standard even in those schools. Another way of measuring 21st century skills is applying questionnaires or scales aiming to measure those skills individually, whereas it is against infusing 21st century skills into core subjects according to P21 (Dede, 2009).

Another study aims to define the assessment of 21st century skills in education, to define the most important skills and suggest different ways to measure those skills (Kyllonen, 2012). After reviewing the frameworks from multiple studies (e.g. Binkley et al., 2012; ATC21S) he adopted a summary framework from the National Research Council (NRC) which describes three main areas for 21st century skills as; "*cognitive skills*-critical thinking, problem solving, creativity; *interpersonal skills*-communication skills, social skills, teamwork, cultural sensitivity, dealing with adversity; *intrapersonal skills*-self management, self-regulation, time management, self-development (lifelong learning), adaptability, executive functioning" (p.8). According to one of Kyllonen (2012)'s conclusions, learning and teaching outcomes can be built around 21st century skills.

Dede (2009) compares the frameworks for 21st century skills which have been defined by different organizations, partnerships and researchers. Briefly, major skills for learning and thinking that are stated in separate but complementary frameworks discussed at Dede (2009)'s paper. Six skills which are listed with a higher frequency by different organizations such as, P21 "Partnership for 21st

Century Skills", ISTE "International Society for Technology in Education ICT Skills", EnGauge Framework from Metiri/NCREL, Organization for Economic Cooperation and Development are critical thinking and problem-solving, interactive communication, collaboration and teamwork, creativity and innovation, information and media literacy, and technology and operations. There are other important skills for 21st century, other than these six. For instance, multitasking, digital citizenship, inquiry (research and information fluency), acting autonomous/ self-directedness, organization skills can be listed as some of the missing skills future work places ask from the citizens of future.

There are also countries or unions of countries, such as European Union (EU), OECD, USA, Japan, Australia, Scotland, England, etc., have been leading projects about defining 21st century skills and had already published documents on this issue (Binkley et al., 2012). One of the common and main aims of these organizations, partnerships or projects is to transform measurement in order to reveal 21st century skills in a valid and reliable manner (Binkley et al., 2012). Some other organizations which have been working to create a framework for 21st century skills can be listed as Metiri Group and NCREL (EnGauge), American Association of College and Universities (AACU), International Society for Technology in Education (ISTE) and Educational Testing Service (ETS) (Dede, 2009). Therefore, there are already many organizations or groups which try to define 21st century skills and find ways to assess and improve these skills for the future society. One of these projects have been done in between 2009 and 2012 called Assessment and Teaching of the 21st Century Skills (ATC21S).

Assessment and teaching of the 21st century skills (ATC21S) project. Although the frameworks for 21st century skills are clustered in different ways, it is remarkable that all frameworks are similar to each other and cover comparable characteristics. For this study, the literature is investigated in terms of the KSAVE (Knowledge, Skills, Attitudes, Values and Ethics) framework, which is not included in Dede (2009)'s compilation. This model is developed under project "Assessment and Teaching of the 21st Century Skills (ATC21S)". ATC21S put assessment into the foundation of the project and its purpose is to define 21st century skills and improve ways to assess them. The project aims to associate assessment with teaching 21st century skills (Binkley et al., 2012). ATC21S has the view that

curricula need to go beyond only introducing subjects like language related topics, mathematics or science but improve abilities such as digital-literacy, collaboration and problem solving.

Griffin et al. (2012) explains how project ATC21S define 21st century skills and which processes the project focusing on. ATC21S focuses on how to use standardized tasks to give feedback to teachers and students in order to develop the classroom applications and how assessment data can be used for instructional purposes. Additionally, these assessments can also give feedback to the schools and policy makers. Binkley et al. (2012) stated the learning-based quality assessment system in *Figure 2*.

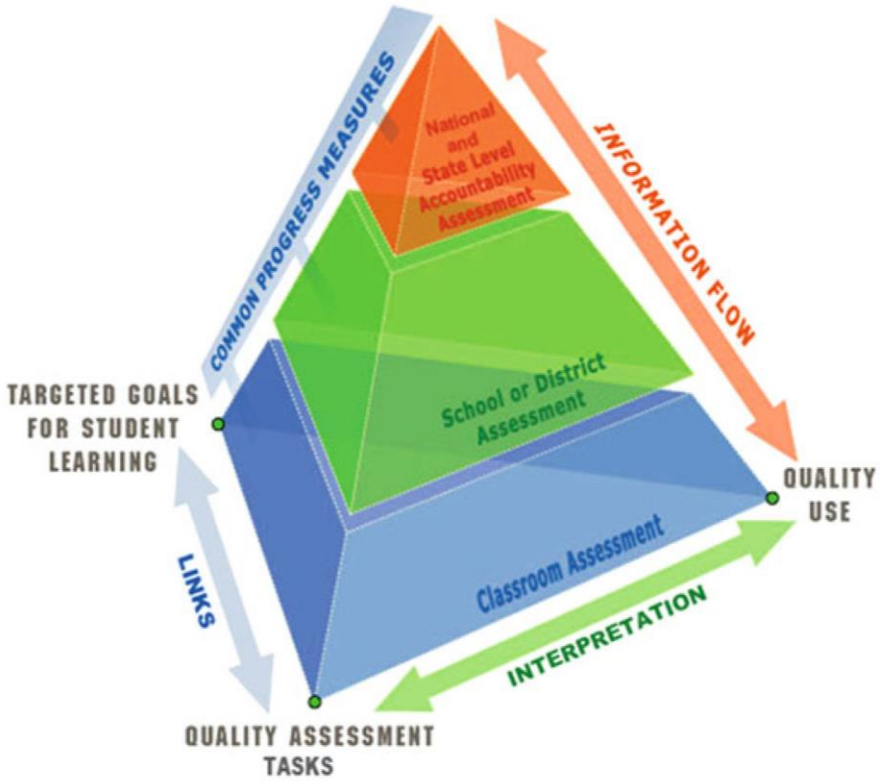


Figure 2. Integrated assessment system

The pyramid shows three dimensions and the relationship between these three dimensions. The starting point for a quality assessment is stated as specifying the goals for student learning clearly. Then the link between the targeted student learning goals and the assessment tasks needs to be established well. These assessment tasks need to capture the important dimensions of intended understanding and skills well. Lastly, these assessment findings need to

be interpreted well and used for a quality improvement with an information flow starting from classroom to national level.

With the need of improvement in 21st century skills of our students, such a system needs to be aligned with the 21st century skills. Identifying the skills is the first step to set the targeted goals correctly for the system. KSAVE model identifies 21st century skills as follows (Binkley et al., 2012, p.36):

Ways of Thinking:

1. Creativity and innovation
2. Critical thinking, problem solving, decision making
3. Learning to learn, Metacognition

Ways of Working:

4. Communication
5. Collaboration

Tools for Working:

6. Information literacy
7. ICT literacy

Living in the world:

8. Citizenship- local and global
9. Life and career
10. Personal and social responsibility-including cultural awareness and competence.

Gathering information about skill sets of related student population has critical importance in order to assess skills which are needed for 21st century. As Binkley et al., (2012) stated that “Teachers need the data to make decisions about appropriate intervention, and they need the skills to interpret the implications of data if they aim to assist students to develop expertise in twenty-first century skills” (p.10). In other words, ATC21S project supports teachers to assess 21st century skills and interpret the outcomes. ATC21S gives curricula recommendations and innovative assessments together with classroom applications. ATC21S states that such applications will shape the future of economic and social development of countries (“Assessment & Teaching of 21st Century Skills”, 2009-2012).

Although the model identifies ten different skills under four main titles, ATC21S project had applications focusing on collaborative problem-solving and Information Communication Technology (ICT) literacy. So, during the ATC21S project, problem solving which is a skill under the category of 'Ways of thinking' is used together with collaboration which is under the category of 'Ways of working'. This is actually typical expectation of the use of 21st century skills; combining more than one skill together.

On the other hand, it is hard to identify, describe and measure 21st century skills, as well as integrating them into the curricula, which cause some challenges for the ATC21S project as well. One of the key challenges for ATC21S was developing original educational and psychological assessment tools. Other ones were the challenge of following students' thinking processes (as one of the biggest challenges of any educational programme), integrating new types of communication methods into the processes and providing the validity standards as the base of the assessments (Assessment & teaching of 21st century skills (ATC21S), 2010). The skills mentioned in the ATC21S framework are defined through the following sections.

Creativity and innovation. Although creativity and innovation are both about generating original ideas Binkley et al. (2012) explains the difference as "Creativity is often the concern of cognitive psychologists. Innovation, on the other hand, is more closely related to economics where the goal is to improve, advance, and implement new products and ideas." (pp. 37-38). Both creativity and innovation are common skills described in many different frameworks by different researches and organizations and they gained increasing focus in educational programs globally (Soland, et al., 2013).

Creativity and innovation can be summarized as, developing new ideas and products as using the past and current knowledge together with the expectations of future as valuing different cultures and boundaries of the real-world. Additionally, creativity and innovation require a systematic thinking as analyzing the parts of a whole, identifying the relations, synthesizing and making new connections between different information (Binkley, at al., 2012) to form new and original ideas and products and evaluate them in terms of usefulness and

originality. New technologies, especially digital cameras and varying software support and make it easy to assess creativity and innovation skills of students.

Critical thinking, problem-solving, decision making. Contrary to creativity and innovation, critical thinking, problem-solving and decision making have been mentioned in educational processes and large-scale assessments such as PISA for long time and critical thinking is often seen as a part of reading comprehension, mathematics and science assessments (Binkley et al., 2012). Facione (2000) describes critical thinking as "a self-adjusting process of judging what to believe or what to do in a given context" (p.65). Soland, et al. (2013) stated the need for the critical thinkers in the context of a company which compete in a global economy and mentioned that companies need people who think about a continuous improvement of the products and so the company and this depends on asking the right questions.

Describing the ideal critical thinker or being aware of the need of the critical thinking is not just happened in the 21st century. In 1987 American Psychological Association (APA) required a panel formed by experts on critical thinking and inquiry to articulate critical thinking and the ideal critical thinker.

> The ideal critical thinker is habitually inquisitive, well-informed, trustful of reason, open-minded, flexible, fair-minded in evaluation, honest in facing personal biases, prudent in making judgements, willing to reconsider, clear about issues, orderly in complex matters, diligent in seeking relevant information, reasonable in the selection of criteria, focused in inquiry, and persistent in seeking results which are precise as the subject and the circumferences of inquiry permit (Facione,1990, p.6).

According to this consensus statement of experts, there were two dimensions stated: critical thinking as a cognitive skill and as an affective disposition. Core critical thinking skills were listed as analysis, interpretation, self-regulation, evaluation, explanation and inference. On the other hand, the affective disposition dimension is about the intrinsic motivation to use the critical thinking skills (Facione, 1990). According to Facione (1990), we need 'our students to be both willing and able to engage in CT [Critical Thinking]' (p.81), to achieve this we need to include critical thinking in to our curricula, instructional assignments and assessment processes, not only show the how to use critical thinking skills but also motivate them to do so. When critical thinking requires to ask correct

questions (Horvathova, 2019), problem-solving skills help to solve these questions which are the product of critical thinking.

İncebacak and Ersoy (2016) describes problem-solving as “offering a solution, an idea to a problematic situation” (p.276). On the other hand, Binkley, et al. (2012) highlighted the sub-skills that problem-solving assessments focus on as measuring “how well students can evaluate evidence, arguments, claims, and warrants; synthesize and make connections between information and arguments; and analyze and evaluate alternative points of view” (p.41).

Some problem-solving examples suggested by Binkley, et al. (2012) can be summarized as authentic open-ended tasks which can be machine scored (Primum), lively, interactive, 5-10 minute long, complex problems for students to solve in a context of an on-screen test (World Class Tests), and virtual performance assessment using technology for inquiry (The VPA Project). All these projects focus on assessing problem-solving skills in a non-routine and unpredictable environment different that the current testing. Solving non-routine problems is an essential part of today’s work requirements, whereas studies show that students struggle the most with non-routine and unfamiliar problems. For instance, İncebacak and Ersoy (2016) reported that majority of the secondary school students attended to their study had difficulty in solving non-routine problems, whereas they were successful at solving familiar problems.

PISA 2003 and another study done with high achiever 4th graders show that students in the Netherlands are tended to give up to solve a problem when it takes too much time or when students need to take notes (Doorman, Drijvers, Dekker, van den Heuvel-Panhuizen, Lange & Wijers, 2007), which supports similar structure as critical thinking is also valid for problem-solving skills. Developing the problem-solving skills does not necessarily guarantee that students would use these skills. They need to be motivated to use these skills and show resilience to continue for searching a solution method.

Decision making can be considered together with critical thinking and problem-solving. After critical reflections on solutions of problems through a system of thinking (as analyzing, synthesizing, interpreting different thoughts, conflicts and solutions), these reflections need to be incorporated into the decision-making processes (Binkley, et al., 2012).

Learning to learn, metacognition. Learning to learn is firstly about being self-aware about the current status of the individual's learning. Learning to learn requires showing the ability of self-managvement of learning as dedicating and organizing the necessary time, showing autonomy, discipline and information management, concentrating in varying time intervals depends on the requirements of the task, reflecting critically about the learning process and being responsible for reaching to the crucial information sources with self-initiative (Binkley, et al., 2012).

Learning to learn (or metacognition or meta-learning) concerns the process about "reflecting on and adjusting one's learning" (Horvathova, 2019, p.48). Learning to learn can be considered as one of the latest occurring skills in the set of 21st century skills, but one of the most required skills that employers are looking for. As it was mentioned earlier, we live in a constantly changing world which expects us and our students to adapt, which is possible with a lifelong learning as providing our students the ability to control and be aware of their own learning.

Communication. Communication and collaboration are considered jointly in some researches, but each of them is a broad concept itself (Soland, et al., 2013). Communication is most commonly performed and assessed in speaking, reading, writing and listening forms. However, especially conducting most of the large-scale tests only in written forms shows that we do not have taken the full range of possibilities of communication into account in education. On the other hand, communication is stated as one of the vital 21st century competencies bu many organizations and researches (Horvathova, 2019).

Society has already started to use many different ways of communication primarily via social media channels such as Facebook, Youtube, Instagram, Twitter. Politicians share their messages with their electors via Twitter or companies spread their advertisements via various web pages, Youtubers or Instagram social phenomena, because the effect of different communication channels is obvious in societies. Additionally, as Soland, et al. (2013) mentioned "effectiveness with clients [in companies] often hinges on effective communication…" (p.6). Hence, work force requirements of the new era bring the necessity of different ways of communication to be integrated into the teaching-learning and assessment processes.

The definition of effective communication needs to be clear. An effective communication can be summarized as understanding and making others understand various messages or arguments which are given in different forms (such as written or oral) with various purposes in variety of situations (Binkley, 2012).

Collaboration. Soland, et al. (2013) emphasizes the importance of cooperation with the following words: "…teamwork is necessary to produce a superior product." (p.6). As Horvathova (2019) beautifully stated "the most radical breakthroughs such as television, airplane, e-mail emerged from a collaborative network of people" (p.44) and today's innovations expect collaboration in even more diverse teams, which requires collaboration skills. For gaining the proficiency of collaboration skills in other words for working efficiently in a team, developing communication skills can be considered as a prerequisite, because individuals need to interact effectively with others first of all to be able to work collaboratively. Speaking clearly and listening carefully when others explain their view are the key points to interact effectively with others. Quality of work can be enriched with social and cultural differences, as well as the strengths of others. These differences need to be used to work towards the common aim of the team as demonstrating selflessness, integrity and ethical behavior for developing a collaborative environment (Binkley, et al., 2012).

Information literacy. The introduction of personal computers, the invention of the Worldwide Internet, the development of interoperable computers and software through internet and transmission protocols that allows the collaboration between computers provide a huge information spread across country borders. Binkley, et al. (2012) stated the extreme amount of information being transferred between people the words: "…,consider that it is estimated that a week's worth of the New York Times contains more information than a person was likely to come across in a lifetime in the eighteenth century" (p.49). This excessive increase in knowledge has advantages as well as disadvantages and some necessities.

The citizens of future must have some skills to reach and assess new information, eliminate the irrelevant information and organize what is necessary. These skills are named as information literacy skills. Information literacy skills can be considered in two main skills as accessing and evaluating information (i.e.

ability to search, collect, and process information) and using and managing information (ability to organize knowledge in a systematic way and use information to support various ways of thinking) (Binkley, et al., 2012).

ICT literacy. In terms of the use of skills, the difference between the 21st century and the previous century is primarily based on the advanced information and communication technologies (Dede, 2009). As Allen and van der Velden (2013) stated the importance of the information, communication technologies have increased a lot which brought the changes not only in our daily lives but also the organization of education systems.

ICT literacy which is also called as digital literacy is one of the ways to deal with the growing amount of information. Based on the International ICT Literacy Panel (2002), Ainley et al. (2005) described the ICT literacy as "the ability of individuals to use ICT appropriately to access, manage and evaluate information, develop new understandings, and communicate with others in order to participate effectively in society" (as cited in Binkley, et al., 2012, p.51). Similar to creativity and innovation skills, ICT literacy is one of the newest skills which is changing our view on what is being assessed. The required sub-skills considered in ICT literacy can be summarized as "accessing and evaluating ICT, use and manage information, create media products, apply technology effectively" (Binkley, et al., 2012, p.52). When using, managing and creating information, technology and media tools, the ethical and legal issues need to be considered and applied properly for ICT literacy as well.

Using information and technology is a requirement for most of us and our students have faced with it more and more. For the last two decades, all areas in a person's life encourages them to use information and technology sources, because of the technological change and all its effects on society. Hilbert (2020) mentioned that in 1980s less than 1% of the world's information was stored digitally, whereas the digitally stored information amount reached to 99% by 2012, and now each 2.5 to 3 years, more information than the beginning of the civilization can be added on the existing information.

Today's lower secondary students (aged between 12 to 16), who are called Gen Z (born in 1990s and raised in 2000s), are born and raised with social web

and technology is part of their identity (Singh & Dangmei, 2016). Using information and technology effectively is unsurprisingly a common characteristic for all today's lower secondary school students, especially in developed countries, where laptop is one of the school materials nowadays, more than a notebook. Therefore, our students needed to adapt into an effective use of information and technology age by nature regardless of the programme type. The development in information and technology brought the social web to the new generation as mentioned, which requires more variety in the ways of communication and collaboration. Hence, today's lower secondary school students are more aware of the outer world than the previous generation and they develop communication and collaboration skills accordingly.

Citizenship-local and global. Local and global citizenship states "how we use knowledge to act on our community and the world around us" (Binkley, et al., 2012. p.56). These skills represent the ability to participate the activities and actions, taking responsibility to solve problems affecting the local or wider communities, respecting the values of others and taking the chances provided by the local and international programs.

The increasing international demand for products, easy marketing via internet, and decreasing costs of transport and communication motivate retailers and consumers to buy on a global market (Allen & van der Velden, 2013). As Binkley, et al., (2012) stated the more we move in the 21st century's competitive and collaborative world, the more we need to understand different aspects of citizenship outside of our own country's borders. We need to educate and assess our students in terms of their understanding of global and international fundamentals and issues.

Life and career. Globalization and the change in technology and economy have brought change in life and career plan of individuals. Most people have and will need to change their career plan in a continuous base year by year. Even those who continue in the same company or job title will be required to adapt and improve on new tasks. Furthermore, employees will need to cooperate with colleagues in another city or in another company more often and they will need to accept differences and be flexible for the changing roles more than ever (Allen & van der Velden, 2013). Binkley, et al. (2012) listed the necessary life and career

skills as: “Adapt to change as operating in varying roles, be flexible, manage goals and time, work independently but interact effectively with others as well as in diverse teams, manage projects and guide and lead others when necessary” (p.57).

Personal and social responsibility. Personal and social responsibility is one of the skills which is hard to assess, whereas one of the crucial skills that today’s society need to show to be able to leave a livable world for the future generations. Binkley, et al. (2012)’s operational definition for this skill can be summarized as: the ability to constructively communicate and negotiate, create confidence with others and empathy with others’ backgrounds, and ability to maintain a separation between professional and personal lives.

Education and Assessment in the Light of 21st Century Skills

Voogt, Erstad, Dede and Mishra (2013) mentioned that “agreement about what 21st century competencies are and how they can be learned does not guarantee the uptake of 21st century competencies and learning in schools” (p.406). Hence, it is valuable to review the ideas formed and studies conducted about the reflections of 21st century skills on education and assessment with its challenges.

There are some obstacles in implementing 21st century skills into the current educational systems. One of them is the need for restructuring of the curriculum (Kozma, 2009; Voogt et al., 2013). It is not a change about the content only, like change the 20th century content with 21st century content. It is a change about the core of the curriculum according to the learning requirements of 21st century skills and their implications for schools (Voogt et al., 2013).

Hopfenbeck (2018) stated that we need to consider our aim as supporting students’ possibilities and life chances. With this perspective, types of assessment used in decision-making processes for students future need to be examined critically. We cannot insist on assessing solely knowing and understanding of our students while current time ask them to inquire, communicate, take risks, consider and reflect the world and their own ideas. They are required to investigate, they are required to work collaboratively, they are required to lead their own learnings and relate everything they learn to real-life applications.

Kozma (2009) mentions a systematic reform in education not only in terms of curriculum but also in terms of pedagogy, teacher training and organization of schools. The reform in assessment is stated as a particular need by Kozma. As it is very clear with all the worldwide discussions and studies, 21st century skills are not about repetitive tasks or knowledge which based on memorization, because these actions can be easily done by computers and can be automized. On the contrary, 21st century's economic and social environments require the ability to adapt complex problems, create solutions, work collaboratively with others, manage information, use technology effectively and produce new knowledge (Kozma, 2009). Hence, in order to integrate these requirements into education, changes in teacher training, in pedagogy, in organizations of schools and in curriculum are required, which is not easy to achieve.

For instance, inquiry-based learning is one of the requirements of especially today's science education and inquiry-based learning can play an important role in developing many 21st century skills, such as scientific problem-solving skills, scientific communication ability, lifelong learning and etc. (Kang & Keinonen, 2016). On the other hand, applying and assessing inquiry-based learning are a big challenge, because of "low confidence and competence in using inquiry instructions of teachers, lack of time and resources, tight curricula, inadequate professional development, large class sizes" (Kang & Keinonen, 2016, p.32), and these obstacles are also standing in front of teaching-learning and assessing the 21st century skills.

As Kuramoto and Koizumi (2018) stated conflicts between the education system and principles of measurement needs to be eliminated. It is crucial to give high priority to educational research for understanding the nature of 21st century skills, reviewing teaching-learning processes and modifying the assessment accordingly as well. Assessment needs to be aligned with the shift in other areas of education (Care, Kim & Scoular, 2017). It cannot work the other way around like first change the testing system without understanding the objectives of the current century and then expect the teaching learning processes will adapt accordingly.

From this perspective, as ATC21S project emphasizes, we need to acknowledge "the symbiotic relationship between assessment and teaching and learning" (Mountain, Care, Scoular & Woods, 2009-2012, p.5). According to the

project reports of ATC21S, assessments of 21st century skills need to be embedded into the teaching and learning processes. Assessment cannot be thought as a separate process which happens after teaching and learning are completed, but it should be spread over the teaching and learning processes. This is one of the facts which makes the objective assessment of 21st century skills really challenging.

In the documents of Project ATC21S, Care, Griffin, Woods and Mountain (2009-2012) explains the compromise between the ideal and the measurable side of 21st century skills in educational environments through collaborative problem-solving skills. Ideal situation is to be able to measure students' reactions to a complex real-life problem which is ambiguous and ill-defined. They acknowledge the trouble in making comparisons between different tasks or different problem solvers with ill-defined problems. Therefore, even for such a big range project, the problem-solving questions are given as well-defined for the respondents to provide easy comparison between tasks and respondents. Some examples of these problems can be seen on Module 3 of the project reports (Scoular & Awwal, 2009-2012). Despite this limitation, ATC21S contributes considerably to the literature of the assessment of 21st century skills as showing the example problems which combines the technology use (ICT literacy), collaboration and unfamiliar problem-solving. Despite forming the framework for 21st century skills ATC21S has not formed an assessment tool as including all the skills which appear in their framework.

As Griffin et al. (2012) brought up the first and maybe the most significant issue about assessing 21st century skills that these skills are not well understood by all parties but especially by teachers yet. Assessment in order to modify teaching in classrooms and assessments to give feedback to the system provide different natures of data. ATC21S project stands out to provide information for both, but they emphasize the importance of different parties' cooperation, especially teachers. Teachers' understanding of 21st century skills is crucial and the system needs to help teachers, instead of implicitly encourage teachers to improve scores but not skills. In order to achieve this goal, higher thinking skills and assessment for higher thinking needs to be emphasized.

As considering 21st century skills are higher order thinking skills, challenges in assessing 21st century skills can be also explained by the challenges in assessing higher order thinking skills. Higher order thinking skills are actually introduced with Bloom taxonomy in 1950s, but reflections of it is a topic still being struggled. One of the reasons is understanding of them have still not been completed in education environments and so by teachers. When a good definition and adequate training of teachers about these skills are not actualized, teaching and assessing of these skills would be elusive (Haladyna, 1997). Kutlu, Doğan and Karakaya (2010) stated the higher order thinking skills as synthesis of more than one knowledge and ability. They also remarked that in order to use higher order skills, a person needs to combine more than one ability and unite this combination to his/her own capacity. Briefly, defining higher order thinking skills is a hard issue, and so measuring and assessing them.

About assessing 21st century skills, Information and Communications Technology (ICT) is mentioned quite often in literature (Kozma, 2009). It is not because all assessment procedures require ICT, but because of the advantages of ICT to produce and apply assessment materials and evaluate their results. ICT reduces the costs and time in all processes of assessment, provide opportunity for computer adaptive testing (CAT) which can arrange the level of the tasks according to the level of the respondent and provide feedback to all parties related with the assessment process. On the other hand, there are both technological and methodological challenges for ICT-based assessment.

Some of the technological challenges are considered by Kozma (2009) as serious beginning costs for technological assessment systems, security problems especially for the online applications, scoring problems for open-ended and different symbol system questions, the need to keep up with rapid change both in technology and skills. On the other hand, some of the methodological challenges can be exemplified as the need to determine the age level relevancy of 21st century skills, the need to improve ways of assessing and scoring compound skills without overlooking the existence of the skill partially, or the need to differentiate individual contributions on a task which are completed as a team (Kozma, 2009).

Specifically for problem-solving skills, one of the main challenges mentioned in the literature is 'the design of good problem-solving tasks that are original, nonroutine and new to the students' (Doorman et al., p.405).

21st century skills need to be assessed with real life tasks which brings us another challenge for assessing 21st century skills. As Kozma (2009) stated the vital difference between standardized tests which education society has been used to for decades is a problem for assessment for new era. Comparison by Kozma (2009) can be summarized as follows:

- standardized assessment methods are designed as subject based whereas real life tasks are interdisciplinary,
- standardized assessment methods are based on rote learning and apply simple problem-solving methods for pre-structured, well-defined problems whereas real life tasks are complex and ill-structured given in real life contexts,
- standardized assessments based on individual work whereas real life tasks expect people to work both individually and as a team member,
- standardized assessment methods are applied without access to notes or any source whereas real life tasks expect people to reach and use many technological tools and information, and as sorting out the useful ones analyze complex problem, develop solutions and come up with products.
- standardized assessment methods are related with the expectations of teacher and school system, whereas real life tasks based on official requirements as well as requirements of an audience group or associates.

Turkish Education and 21st Century Skills

The controlling body for education in Turkey is the Ministry of National Education (MoNE). All schools are subject to the MoNE until the end of secondary school. In another words, schools or educational institutes in Turkey have to abide by the rules and regulations of MoNE and must obtain an approval from MoNE in

order to modify their curriculum. Therefore, new reforms in Turkish education such as adapting the 21st century skills into education process should be evaluated in the country level with MoNE's decisions.

There have been some actions taken in terms of describing and developing the 21st century profile in Turkey since 2001. Research and Development unit of the Ministry of National Education (EARGED) had conducted a study in 2001 in order to define the characteristics of the 21st century teacher (Geban, 2001). A quite comprehensive report was published related to this study, and the knowledge, skills, characteristics that a 21st century teacher should possess were discussed.

Another similar study was conducted again by Research and Development unit of the Ministry of National Education (EARGED) in 2011, this time to reveal the student profile for 21st century. The research question was "What is the current profile of secondary school students and what is expected to be? If this expected profile does not exist, what should be done to achieve this goal and what measures should be taken?". The research was titled as "The characteristics of a 21st century student" (MEB, 2011). As an output of the research, 21st century student profile was summarized with 38 skills. Some of these skills can be listed as; communication, collaboration, problem-solving, conducting rational and scientific research, critical thinking, accessing and managing information, using technology effectively and efficiently, be selective in technology use, self-renewal, caring others and be democratic, be sensitive to the global problems and contribute to the global peace, life-long learning, learn how to learn, etc. These characteristics are overlap a lot with those offered by many international organizations.

Although comprehensive researches had been conducted about 21st century skills in Turkey together with their ongoing problems to reach the target profiles, there was not supportive case studies which support these big scale studies (Hamarat, 2019). On the other hand, the awareness of 21st century skills have caused a philosophy change in Turkish education from behaviorist to constructivist which brings the curriculum changes since 2005 (Hamarat, 2019). The completed and ongoing studies about 21st century skills can be considered as

the proof that MoNE places importance on 21st century skills parallel with the rest of the world.

Dutch Education and 21st Century Skills

Dutch schools have a high degree of autonomy which means any legal person can found a school, and hire whoever they want as a teacher and receive the public funding. Schools can determine the content and the methods of teaching, and central government determines the learning objectives and quality standards. Learning objectives and the quality standards are inspected by the Inspectorate of Education. Additionally, “The Netherlands has a strong tradition of well-developed evaluation and assessment systems” (Nusche, Braun, Halász & Santiago, 2014, p.6). Evaluation and assessment systems are organized mainly between the Ministry, Inspectorate and schools, but there are also educational research institutions, private educational service providers and organizations involved in the process (Nusche, Braun, Halász & Santiago, 2014).

The quality of education in the Netherlands is generally accepted as high, and this causes a strong ideological nature for educational reforms which does not involve teachers and classroom feedbacks in the design of the reforms. Van Veen, Bloemert and Wolthuis (2020) claimed that not involving teachers into the process is the main cause of failing reforms. For the 21st century skills, the story is different! All the teachers from primary to secondary schools are aware of the 21st century skills. This gives hope for the future of the educational system in the Netherlands in terms of integrating the 21st century skills in education. “Major curriculum modifications have consequences at the classroom level” (Nieveen & Plomp, 2017) p.1), so there should be a feedback mechanism between national and classroom levels mutually.

Nieveen and Plomp (2017) proposed five principles for dealing with the challenges in integrating the 21st century skills in education. These suggestions are the implementation of the 21st century skills should be a learning experience for all parties involved, the old should not be totally forgotten, but the new should be encouraged as considering the old as well. Implementation should firstly happen in small scale instead of changing everything at the same time and the

necessary time should be provided. Implementation of 21st century skills need freedom as well as clear borders.

Despite such proposals have been done by different researchers and organizations, the autonomy of Dutch schools "the absence of a common vision for schooling in the Netherlands" (Nusche, Braun, Halász & Santiago, 2014, p.7) creates an obstacle in front of a coherent plan, especially in terms of evaluation and assessment of the goals for the 21st century. According to Nusche, Braun, Halász and Santiago (2014, p.7): "Defining such learning goals [creative thinking, problem-solving and collaboration] for the 21st century would allow key stakeholders to engage in reflection and dialogue on how evaluation and assessment should evolve in order to support a future-oriented education system".

International Baccalaureate (IB) Programmes and 21st Century Skills

International Baccalaureate (IB) programmes aim to educate today's learners for intercultural, competitive and collaborative work environments. All IB programmes except Primary Years Programme offer both IB-validated assessments and criterion-based formative and summative classroom assessments together. IB-validated assessments aim to ensure validity and reliability of assessments in IB schools as well as balance the quality between IB schools. In order to create valid and reliable exams, IB examiners started to work two years prior than candidates take the exams. First the committee which aims to prepare the exam should be selected. They discuss assessment standards as aiming each level is represented and content of the course and objectives are covered (content validity). They also criticize and evaluate each question in order to hinder any possible bias or ambiguity. Additionally, they examine mark schemes meticulously in order to ensure rater reliability, in terms of being clear, comprehensive and detailed (IBO, 2019b).

Harlen & Johnson (2014) conducted a project in order to review Primary years programme (PYP) curriculum of IB in the light of assessment which they find quite related with 21st century education. Harlen & Johnson (2014) describes the general approach of assessment in PYP as "a well-designed learning experience will provide data on students' knowledge, skills and conceptual understanding…" (pp.4-5). According to the key findings of review projects show the importance

given to formative assessment through PYP as well as feedback, self and peer reflection and using any assessment data to improve learning processes.

IB Diploma Programme (DP) is one of IB programmes which takes two years with also an option of three years. IB DP aims to ensure students a balanced education with an international understanding. IB DP is offered for students aged 16 to 19 (IBO, 2018a). Besides IB-validated challenging exams, programme aims to support students' educational qualifications with three core long term assessment experiences. Extended Essay (EE), Creativity, Action, Service (CAS) and Theory of Knowledge (TOK) are three core elements/assessments of the programme which provide students to learn through and from assessment. Three core elements can be explained briefly as follows (IBO, 2019a);

Extended Essay (EE) is an individual, self-directed research which investigates a subject of student's interest. Through EE process, students are expected to develop skills, such as generating a proper research question, engaging into a research process, communicating ideas written and verbal so on (IBO, 2018a).

Theory of Knowledge (TOK) is a mandatory course for all DP students and collectively with EE can add up to 3 points in which minimum 24 points mean diploma is awarded. TOK course is assessed via two tasks, one is an oral presentation which assess students' skills to apply theory of knowledge to real-life situations, and other is a more conceptual 1600 words essay (IBO, 2018a).

Creativity, Action, Service (CAS) is a process of following range of activities which enables students to use initiative and show dedication about the projects they decide to pursue, as well as developing the skills mentioned in IB learner profile. CAS expects students to be involved in different experiences which are classified in three strands of CAS as creativity, action and service. Creativity strand includes all activities which involve creative thinking. Activity strand support activities which support a healthy individual and society. Service strand support students to help others unpaid with a learning exchange. All chosen experiences need to be volunteer and support self-development with learning outcomes and contribute to the society. CAS experiences are not formally assessed, but students

need to complete their experiences as providing proofs and reflections about stated learning outcomes in order to complete one out of three cores of the programme. CAS is not a graded assessment, but all responsibilities need to be completed and confirmed by advisor in order to be awarded for diploma (IBO, 2018a).

When DP core is examined, it would be clear how various assessment types have been used in IB schools in order to be awarded for a diploma. Findings of a study reveals that DP prepares its students better for college. Additionally, students who completed four or more DP classes during secondary school are less like to drop college, more able to deal with high workload and time pressure of college life. Another finding of the same study is that DP students appreciated three core elements, especially Extended Essay, as very useful to feel prepared for college (Conley, McGaughy, Davis-Molin, Farkas & Fukuda, 2014).

International Baccalaureate Organization (IBO) offers a different programme for the students aged 11 to 16, which is called Middle Years Programme. MYP aims to develop student-centered, international-minded, inquiry-based and real-life related learning environments (IBO, 2018b). In MYP, assessment is indispensable in all teaching and learning processes. MYP use a criterion-based assessment which will be explained in detail.

Assessment processes are given integral with curriculum in IB schools. Assessment is organized according to four different criteria, and a report grade from any subject is formed as taking the average of these four scores. Hence, IB MYP utilized from criterion-referenced assessments which compare student's performance with the aimed criterion instead of other students. Each criterion based on a different skill. For instance, Criterion A is assessed mostly by classroom tests which have questions structured according to level. There is not a total point graded but a level from one to eighth. If there are multiple scores for Criterion A, for instance, teacher uses his/her view for that criterion and it is called 'best fit approach'. This is valid for each criterion. Other three criterion can be a short- or long-term project or a classroom application and each criterion has different lists of skills in order to decide students' scores.

Silva (2009) explains how assessment of 21st century skills come true in diploma programme of IB with the words, "…nonprofit International Baccalaureate Organization, serves as evidence that the assessment of core content and advanced skills, aligned with a program of standards and curriculum, can happen at a large scale, even international scale" (p.633).

Öztermiyeci (2019) aims to explore the 21st century skills of students who follow the national curriculum of Turkey and IB Diploma Programme, and investigate the differences between these two groups of students with their Master thesis. The perception of 21st century skills are questioned in terms of three domains (cognitive, affective and sociocultural) of Kang et al. (2010)'s study. The study group is formed by students from four different schools which have both IBDP and national programme students. According to the results, IBDP students' averages are higher and statistically significant for all domains compared to the students who follow the national curriculum. The highest difference appears for the cognitive domain in favor of IBDP students and it is concluded with the aforementioned study that IBDP has a positive effect on improving students' perceptions about their 21st century skills.

Dulun (2018) focused in her study to investigate how IBDP advances international mindedness which promotes students' 21st century skills. It is designed as a comparative case study and compare students' perceptions of how they were qualified for the IBDP (during grades 11 and 12) when they follow three different curricula during grades 9 and 10, which are Turkish national curriculum, the International General Certificate of Secondary Education (IGCSE) and IBO's Middle Years Programme (IBMYP). Findings states that students from all three different preparation processes had positive perceptions about their learning experiences during grades 9 and 10. Students from the Turkish national curriculum and IGCSE report their confidence on preparation for national university examinations whereas students from IBMYP report their advantage on improving 21st century skills, such as research skills, critical thinking skills and communication skills.

Bayülgen (2012) aims to determine the achievement levels of students who attend to International Baccalaureate (IB) Turkish A1 program and Turkish language national program in terms of the target skills stated in MEB (Ministry of

Turkish Education) Turkish language program. The study had been conducted in five different schools in Turkey with 340 students and 30 teachers. Findings show a significant difference between General program and International Baccalaureate program students' target skills on behalf of IB program. This study shows that IB students reach to a higher level than general program students in the skill averages. On the other hand, except "using information technologies" there is significant difference between male and female students on behalf of female students. According to teachers' views, teachers find IB program more useful to enable students to develop target skills.

Related Studies

Some of the thesis and dissertations consider the 21st century skills as the multidimensional skills set (e.g. Alpaslan, 2021; Benek, 2019; Karakaş, 2015), whereas some of the studies focus on one of the domains, and this is mostly the learning and innovation domain (e.g. Bircan, 2019; Erdoğan, 2019; Karademir, 2020).

Related studies in Turkey. There have been quite a number of thesis and dissertations completed in Turkey on 21st century skills in education since 2015. These thesis and dissertations can be examined under four main purposes; 1) Theses which aim to investigate the effect of a method or approach to improve students' 21st century skills (e.g. Alpaslan, 2021; Benek, 2019; Bircan, 2019; Dinç Bilgin, 2021; Külegel, 2020; Murat, 2018; Yavaş, 2021), 2) Theses which aim to evaluate educational curricula or materials in terms of 21st century skills (e.g. Akçay, 2019; Doğan, 2020), 3) Theses which aim to investigate the relation between 21st century skills and other variables (e.g. Alkış, 2020; Aydın, 2021; Erkılıç, 2020; Karademir, 2020; Bircan, 2019; Karakaş, 2015; Kaya, 2020), 4) Theses which aim to investigate the perspectives or perceptions of student teachers, teacher candidates, teachers or managers about 21st century skills (e.g. Atakişi, 2019; Aydın, 2019; Çınar, 2019; Erbek, 2021).

The common purpose of all these theses and dissertations can be summarized as "improving the 21st century skills of future's citizens". We want to improve the 21st century skills of our students, so we are looking for the effect of our methods or approaches that we use in our lessons. We want to improve the

21st century skills of our students, so we are evaluating our curricula or lesson materials to observe how 21st century skills are taken into consideration in those curricula or lesson materials. We want to improve the 21st century skills of our students, so we investigate if having some characteristics gives any advantages to these individuals or groups of people in terms of 21st century skills. We want to improve the 21st century skills of our students, so we want to investigate if people who are responsible to deliver the education have these skills themselves.

Theses which aim to investigate the effect of a method or approach to improve 21st century skills can be summarized with two findings; the ones which show a positive effect on improving the perception of 21st century skills with their method and the ones which could not find an effect.

Alpaslan (2021) focused on interdisciplinary science activities in his master's thesis and its effects on students' perception of 21st century skills and creative problem-solving skills. During the research, he followed the explanatory sequential pattern with one group pre- and post-test applications. He used the creative problem-solving properties inventory and the perception of 21st century skills scale as pre- and post-test with 50 fifth grade students and conducted interviews with 8 of the students after the application of interdisciplinary teaching approach. Findings showed that application of interdisciplinary teaching approach had a positive impact on students' perception of 21st century skills and their perception about their creative problem solving skills. These findings were supportted by the interviews as well.

Yavaş (2021) aimed to investigate the effect of cooperative multiple intelligence cycle (CMIC) model on 6th and 7th grade students' perception of 21st century skills, scientific creativity, critical thinking, entrepreneurship based on science and attitude towards science and technology with her master's thesis. The study was conducted with 36 students. Measurement tools were applied one time before the 6 weeks of CMIC model application and one time after the application. According to the findings, CMIC model affected both 6th and 7th grade students' 21st century skills positively. Additionally, the model affected 6th grade students' attitude towards science and 7th grade students' critical thinking and entrepreneurship positively. Combined findings of quantitative and qualitative data collection and analysis were interpreted as CMIC model improved middle school

students' some characteristics needed for 21st century and such models should be used in classroom environments. Dinç Bilgin (2021) reached to a similar conclusion in her master's thesis which focuses on the effects of 2D and 3D supported modeling on students' academic achievement and 21st century skills. Dinç Bilgin (2021) conducted an experiement-control group design with 43 students (22 in experiment group, 21 in control group) and according to her findings, experiment group improved more in terms of both the perception of 21st century skills and academic acievement of the stated science unit after the 7 weeks of 2D and 3D supported learning.

Karademir (2020) investigated the effect of digital storytelling activities on perception of 21st century learning and innovation skills. The study group was formed by 20 gifted 2nd grade students. The 21st century skills scale used for collecting data was formed of three factors; creativity and innovation, critical thinking and problem-solving, and collaboration and communication. According to the quantitative findings of the research students' perception about creativity and innovation, and critical thinking and problem-solving were improved positively, whereas their perception about collaboration and communication skills were not changed. These findings support Binkley et al. (2012) KSAVE model in one way, because of the different findings for the skills of ways of thinking (creativity and innovation, critical thinking and problem-solving) and the skills of ways of working (collaboration and communication).

Bircan (2019) investigated in his doctoral dissertation the effect of STEM (Science, Technology, engineering & Mathematics) education activities on fourth grade primary school students' attitudes towards STEM, 21st century skills and mathematics achievements. The study was conducted with 34 fourth grade students and quantitative and qualitative methods used together. For quantitative part of the study, a time-series quasi-experimental design was used and data was collected via the STEM attitude scale, 21st century learning and innovation skills scale, scratch achievement test and mathematics achievement test developed by the researcher. For the qualitative part, semi structured interviews were used in order to determine the views of the six participants of the study. Bircan (2019) found that, STEM education did not make a difference in fourth grade students' attitudes towards 21st century skills, whereas STEM education caused a

statistically significant difference between the pre- and post-test results of 21st century learning and innovation skills scale. Additionally, students stated during the interviews that the activities conducted during the STEM education develop their creativity, problem-solving, collaboration-communication and critical thinking skills. On the other hand, STEM education did not make a difference on mathematics grades.

Benek (2019) carried out his doctoral dissertation research on the effects of socio-scientific STEM's activities. He investigated the effects of designing these activities on students' attitudes towards STEM, their perceptions of 21st century skills and opinions regarding STEM. The research was conducted with 16 seventh grade students for 24 weeks and quantitative and qualitative methods were used together. Benek (2019) concluded according to the quantitative results that designing products as using the engineering design steps affected students' attitudes towards STEM and perception of 21st century skills positively. The qualitative findings supported the quantitative ones. There was no difference in terms of gender.

After all these researches which found positive effects of different models and approaches, there are also some researches in the literature which could not find out any effect of the used approach or method. Korkmaz (2019) investigated the effect of argumentation-based science learning approach on 7th grade students' perception of 21st century skills with an experimental design and he could not find any difference on the perception of 21st century skills between the experiment (n=40) and control (n=40) groups. Similarly, Murat (2018) could not find any effect of flipped classroom model and Özgün (2019) could not find any effect of creative drama instructions on fifth grade students' level of using 21st century skills.

For the purpose of the current study, it is important to summarize especially the studies which investigate 21st century skills in relation with other student variables. Alkış (2020) conducted her master's thesis to investigate whether the 21st century qualifications of university students differ according to their gender, age, parents' income or types of university entrance exam scores (e.g. numeric, verbal, etc.). Data was collected from 572 university students with a four domain, 17 factor scale called "21st century skills scale test" developed by Yılmaz and Alkış

(2019). These four domains were knowledge, ability, character and meta-learning. Alkış (2020) also questioned if the knowledge of university students can interpret the outcome of the other three domains of the 21st century skills scale. Findings were varied according to the variables included into the study and the factors of the 21st century skills scale test. The remarkable findings can be summarized as follows; There are significant differences according to the gender in communication and creativity factors in the advantage of boys, whereas there was no difference found for critical thinking or collaborative learning between boys and girls. There was not any significant difference according to the age for the aforementioned factors. There was a significant difference in creativity of the students in the advantage of the highest income group.

Aydın (2021) investigated the relation between the perception of 21st century skills, achievement-oriented motivation and motivation for learning English. The target group of the study was the students of English preparatory class of a public university in Turkey. According to the findings the perception of 21st century skills of university language preparation class students differed according to the department and programme types whereas did not differ according to gender or the level of English.

Erkılıç (2020) investigated the 21st century skills of undergraduate students who completed physics course and the relation between these students' perceptions of 21st century skills, achievements in physics, attitudes towards physics course, and perceptions about physics teaching and learning processes. The study was conducted with 329 undergraduate students. According to the findings, there were no significant difference in students' perception of 21st century skills in terms of gender or the types of secondary school that they were graduated. On the other hand, there were significant relation in students' perceptions of 21st century skills and their attitudes towards physics course, and significant relation in students' perceptions of 21st century skills and perceptions about physics teaching and learning processes. Findings showed that students' perceptions of their 21st century skills explained 35% and 36% of their attitudes towards physics course and perceptions about physics teaching and learning processes, respectively.

Kaya (2020) aimed in her master's thesis to determine perception of 21st century skills and entrepreneurship of pre-service primary teachers, and investigate the relationship of these variables together with other student information, such as gender, grade, achievement etc. Kaya (2020) conducted her study with 275 pre-service primary teachers. Findings showed that pre-service primary teachers have an above average level of perception of 21st century skills and entrepreneurship. There was also a positive and above average relation between these two variables and entrepreneurship explained 52.8% of pre-service teachers' perception of 21st century skills. There was no difference between pre-service teachers' perception of 21st century skills in terms of achievement levels or work experience for any factor, whereas there were some differences between gender (career awareness factor) and grade (social responsibility and leadership).

There are Turkish researchers who developed (e.g. Anagün, Atalay, Kılıç, & Yaşar, 2016; Kaya, 2017; Yılmaz & Alkış, 2019) or adapted (e.g. Karakaş, 2015) scales to measure the 21st century skills. Despite some of the scales emphasizes the concept "perception" in the name of the scale, there is a general misuse about what the scales actually measure. In all the thesis reviewed from the National Thesis Centre (https://tez.yok.gov.tr/UlusalTezMerkezi/) with the key work "21. yüzyıl becerileri" (21st century skills), if there is a scale or questionnaire used for measuring the 21st century skills, they actually measure the perception of the target group in terms of 21st century skills. Therefore, even though findings are mentioned as "21st century skills" in some of these studies, these findings from the literature are stated as "perception of 21st century skills" in the current study. The scale development or adapting studies in Turkey can be summarized chronologically as follows.

Karakaş (2015) aimed to determine the level of perceptions of middle year students' 21st century skills and investigate if these perceptions differ according to their gender. Karakaş (2015) adapted a scale from English to Turkish during this study in order to collect data about eighth grade (middle school) students' perceptions. A 32 item 5 point Likert scale was formed after the adaptation process and the scale was used by more researchers (e.g. Alpaslan, 2021; Dinç & Bilgin, 2021; Korkmaz, 2019) in Turkey. According to the findings, the perception of 21st century skills of the eight grade students were high which was supported

by semi-structured interviews as well and there were differences in terms of gender in the favor of girls.

Anagün, Atalay, Kılıç and Yaşar (2016) aimed to present the validity and reliability findings of a scale which aims to measure the pre-service teachers' perceptions of 21st century skills. After the factor analyses, the scale was formed as 42 items under three factors as learning and innovation skills, Life and career skills and Information, Media and Technology skills. The scale was used by other researchers (e.g. Çınar, 2019; Erbek, 2021).

Kaya (2017) developed a scale during her doctoral dissertation aiming to determine the differences in the perception of 21st century skills, burnout and school engagement levels in terms of some demographic variables such as age, gender, school type, the education and socio-economic status of parents. She also tried to observe if the 21st century skills are a significant predictor of high school students' burnout and engagement. In order to conduct the aforementioned study, Kaya (2017) developed a 19 items 7-point Likert scale as considering the 21st century skills as critical thinking, problem-solving, communication and collaboration, digital literacy and creativity and innovation. The 19 items of the scale appeared under three factors as; using the information, access to information and querying the information. Findings showed that high school students have a high level of perception of 21st century skills. There was a difference in terms of gender in the favor of girls and there was no difference in terms of age or grade. Additionally, there was a negative moderate relationship between perception of 21st century skills and burnout, and a positive moderate relationship between perception of 21st century skills and school engagement.

Çevik and Şentürk (2019) aimed to develop a 21st century skills scale for 15–25-year group people. They reported the validity and reliability study of a five-factor scale formed by 41 five-point Likert scale. The factors were called as; information and technology literacy skills, critical thinking and problem-solving skills, entrepreneurship and innovation, social responsibility and leadership skills and career awareness. Some item samples of this scale can be listed as; "I like to listen new and different ideas" or "I do not want to be friends with people who do not think like me". So, one sentence and self-report items form this

multidimensional scale similar to the other 21st century skills scale developed in Turkey. The scale was used by other researchers (e.g. Aydın, 2021; Altun, 2021).

Yılmaz and Alkış (2019) developed a 21st century skills scale formed by four sections which can be used separately. These sections are called 'Knowledge' which is formed by 7 factors and 27 items, 'Skill' which is formed by four factors and 19 items, 'Personality' which is formed by four factors and 19 items and meta-cognition which is formed by two factors and 16 items. Hence, the scale is formed by 80 five-point Likert scale items. The validity and reliability study of the scale was conducted with 560 university students of three different universities in Turkey. Both EFA and CFA were performed with the same group and researchers reported that the scale had sufficient psychometric characteristics. The scale was used by other researchers (e.g. Alkış, 2020).

Other related studies. Serdar (2015) carried out his doctoral dissertation research on analysis of traits, attides and characteristics of teachers who show innovation, creativity and 21st century skills. As doing this research, he aimed to gain a better insight about schools and give ideas to improve them. The researcher formed the study group from classroom teachers who were identified by their colleagues as being innovative and creative. Teachers of the grades three to six from four different school in the same district filled a questionnaire to identify their creative and innovative collegues and they pointed out five female classroom teachers. Data collection was done with teacher interviews, student interviews, reseracher observations and artifact (lesson plans, influential books and notes) analysis. The researcher intended to find common themes as analyzing those input collected by interviews and observations. Reseracher found seven themes as the most remarkable attributes of the innovative and creative teachers who attended into his research; "see the teachers as lead learner, encourage learner reflection, foster class community and relationships, give students choice to instill ownerhip, employ project/problem-based learning, make connections tor eal life, encourage teacher and student collaboration" (p.77)

As a product of their nationwide survey, Kang et al. (2010) constructed a 33 items scale sourced from three domains (cognitive, affective and sociocultural) with four factors in each domain. Items are formed as four-point Likert scale. One of the core findings of this study was in order to promote a learning environment

which supports the competencies for 21st century, activity and process-centered teaching and learning processes need to be carried out. Only then, we can help transferring abstract knowledge or skills to needed performances.

Summary

Briefly, literature and the conceptual framework behind the study are supported by various researches. Twenty-first century expects new skills from current students in order them to be skillful citizens of the future. Assessing these skills needs to be integrated with curriculum as well as teaching and learning processes. Traditional measurement methods might be still useful for some skills but for many more skills expect to be assessed. Alternative assessment processes need to be created which do not focus on the end result but the process, not to the scores but to the skills and according to the researches and reports published, IB programmes can be considered as a sample of future assessment for education.

Chapter 3
Methodology

In this chapter; setting and participants, data collection, instruments and data analysis are presented. The purpose of this study is to investigate the effect of programme type (IB vs non-IB) on students' 21st century skills perceptions and students' problem-solving skills in relation with some variables. To achieve this goal, this research is designed as a correlational study. Correlational studies intend to determine the relationships between variables or use these relationships to make predictions for possible outcomes (Frankel, Wallen & Hyun, 2012). Correlational studies provide a numerical estimate of how the variables are related. Higher this numerical estimation is, the higher the relation or prediction power is accepted (Gay, Mills & Airasian, 2006).

Setting and Participants

The target study group of the current research is formed of students aged 12 to 16 years old with English as their language of instruction. Process of collecting data had been completed in the Netherlands, considering the researcher had lived in the Netherlands during the years of the research, and research questions were formed based on the researcher's work experiences in the Netherlands. Additionally, there is a previous study which explored the 21st century skills of students who followed the national curriculum of Turkey and IB Diploma Programme, and investigated the differences between these two groups of students (Öztermiyeci, 2019). According to the results of this aforementioned study, IBDP students' averages are higher and statistically significant for all domains (cognitive, affective and sociocultural) compared to the students who follow the national curriculum. The highest difference appears for the cognitive domain in favor of IBDP students and it is concluded that IBDP has a positive effect on improving students' perceptions about their 21st century skills in Turkey. Therefore, it is interesting to investigate the perceptions of 21st century skills in terms of the programme types in the Netherlands. Based on the language of education in IB schools in the Netherlands, language of the measurement instruments was decided to be in English. This brought the need for the non-IB

schools that the language of instruction to be English as well to be able to collect data in English.

Target group of this study (12 to 16 years old) was the lower secondary school students according to the definition of OECD (Ananiadou & Claro, 2009). Age boundaries were chosen based on the reading comprehension of students and the compulsory education age of Netherlands. By age 12 students' reading comprehension is expected to be developed enough for understanding the scales and tests used to collect data for the current research. Additionally, all children in the Netherlands have to attend to school until 16 years old (Nuffic, n.d.-a).

Correlational analyses are less reliable if they are estimated from small samples (Tabachnick & Fidell, 2001). Hence, researchers generally aim to reach as many respondents as possible for quantitative studies. Additionally, as Kline (1993) mentioned that the bigger the sample, the lower the standard deviation of correlation. Hence, both to test the construct validity of the scale (PoTCS) and for model analyses, it is aimed to reach at least 300 students for each analysis.

For the first stage, the perception of 21st century skills (PoTCS) scale was applied to 379 students. This study was conducted in an International School as considering the language of the scale and the diversity of the students in terms of their school background and nationalities. Data was collected from whole classes, so there were students in those classes who were younger than 12 or older than 16 years old in the data set. Distribution of the participant students in terms of age is stated in *Table 2*.

Table 2

Distribution of Students Who Answered the PoTCS Scale for EFA

Age	Frequency	Percentage (%)
11	39	10.3
12	62	16.4
13	78	20.6
14	62	16.4
15	68	17.9
16	51	13.5
17	18	4.7
18	1	0.3
Total	379	100.0

Target age group of the study was 12 to 16, so students aged 11 (39 students), 17 (18 students), and 18 (one student), in total 58 students were eliminated from the data set. Hence, the study group remained with 321 participants for the first phase analysis of PoTCS scale and distribution of the students by age can be seen in *Figure 3*.

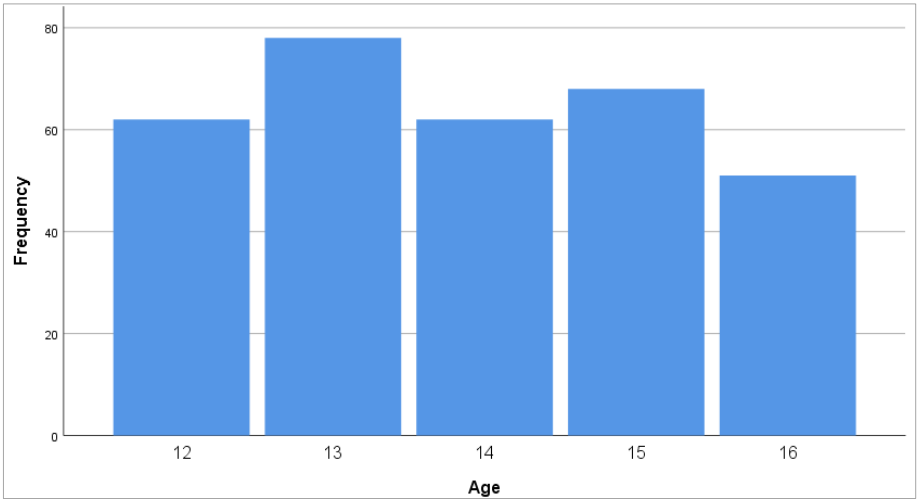


Figure 3. Distribution of participants by age for the pilot study.

For investigating the effect of the programme type (IB vs non-IB) on students' perception of 21st century skills and problem-solving skills in relation with English comprehension, age, achievement (Mathematics and English scores) and gender, students from both IB and non-IB programme schools were included in the study group. In 2020-21, when this study was planned and conducted, there were 26 IB world schools in the Netherlands, eight of them were offering IB PYP, 14 were offering IB MYP and 19 were offering IB DP and three were offering IB CP (IBO, 2020). Distribution of IB world schools and programmes offered in Netherlands are given in *Table 3*.

Table 3

Number of IB World Schools and Offered Programmes in the Netherlands

IB Programmes	Number of Schools
PYP	8
MYP	14
DP	19
CP	3

Based on the target age group, IB schools which offer middle years programme (MYP) formed the target IB schools for the current study. Before the data collection researcher attempted to contact with all 14 school which were offering MYP via email or IBO website. Four of the schools were positive about attending to the study, whereas one of them could not attend the last minute because of the Covid-19 lockdown and the decreased lesson hours. For the three schools attended, there were also limitations about the number of students who could answer the scales based on the lesson cancellations or quarantine procedures of some classes.

Dutch secondary schools which use English as language of instruction (other than the international schools) are classified as bilingual schools (tto-tweetalig onderwijs) and in 2020-2021, there were 130 tto schools in the Netherlands (Nuffic., n.d.-b). Tto schools formed the target group who follow a non-IB programme and have English as their language of comprehension to collect data. In order to investigate the perception of 21st century skills of students for each group (IB and non-IB Netherlands), it is aimed to reach enough and similar number of students. Therefore, five tto schools were contacted. These schools were chosen to collect data because they offer tto for each age group from 12 to 16 years old students. Four of those schools replied positive for the data collection of the research, whereas again based on Covid limitations only two of them could attended to the data collection. Distribution of the participant students for the second phase of data collection is stated in *Table 4* according to students' programme type, age and gender.

There were 304 students answered all scales. These students are different students from the 379 students mentioned earlier. Two students eliminated because of missing data in demographic information questionnaire. The distribution of 302 students can be seen in *Table 4* in terms of gender, age and programme type.

Table 4

Distribution of Students According to their Programme Type, Age and Gender

Programme			Age					Total
			12	13	14	15	16	
Non-IB	gender	Male	16 19.5%	18 22%	22 26.8%	22 26.8%	4 4.9%	82 100%
		Female	20 25.6%	24 30.8%	16 20.5%	15 19.2%	3 3.8%	78 100%
	Total		36 22.5%	42 26.3%	38 23.8%	37 23.1%	7 4.4%	160 100%
IB		Male	16 24.6%	18 27.7%	18 27.7%	12 15.6%	4 5.2%	77 100%
		Female	27 35.1%	16 20.8	18 23.4%	12 15.6%	4 5.2%	77 100%
	Total		43 30.3%	34 23.9%	36 25.4%	22 15.5%	7 4.9%	142 100%
Total		Male	32 21.8%	36 24.5%	40 27.2%	32 21.8%	7 4.8%	147 100%
		Female	47 30.3%	40 25.8%	34 21.9%	27 17.4%	7 4.5%	155 100%
	Total		79 26.2%	76 25.2%	74 24.5%	59 19.5%	14 4.6%	302 100%

As seen in *Table 4,* the study group was formed by 147 (48.7%) male and 155 (51.3%) female students. There were 142 (47.0%) students who were attending to an IB school and 160 (53.0%) students who were attending to a non-IB school. Test-retest reliability was conducted later with different people than previous two groups which were used for EFA and SEM. Data collected for test-retest reliability of PoTCS scale and SEM analysis were used together for CFA of PoTCS scale.

Data Collection

Data collection was completed over a two-year process (2019-2021) in different steps. Data collection and analysis were done in an interlocking manner. The summary of the data collection and analysis process can be seen in *Figure 4*.

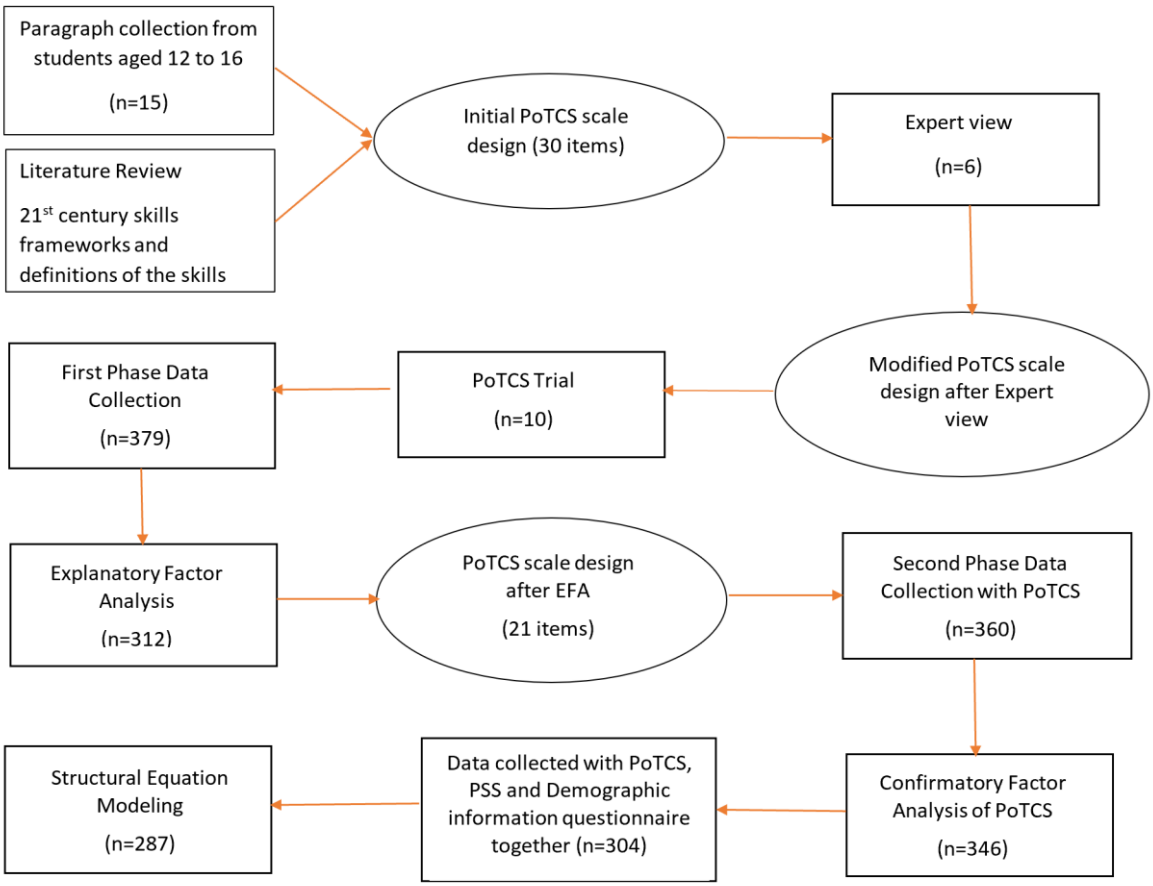


Figure 4. Data collection and analysis process.

Preparation of the instruments and data collection had been implemented as follows;

1. Forming the initial item pool for PoTCS scale.
2. Making the necessary modifications and prepare the PoTCS scale for the initial data collection, after the expert view and trial application.
3. Collecting data with the initial form of PoTCS scale. The allocated time was 15 minutes for this application.
4. Conducting EFA and completing necessary modifications and preparing PoTCS scale for second phase data collection for model test.
5. Preparing the demographic information questionnaire and problem-solving skills test.

6. Collecting data with PoTCS scale, demographic information questionnaire and problem-solving skills test. The allocated time was 50 minutes for this phase.
7. Conducting CFA and deciding the final form of PoTCS.
8. Application of PoTCS scale in 2 weeks for test-retest reliability purpose to a smaller group.
9. Testing the model with SEM to investigate the effect of programme type (IB vs non-IB) on the perception of students' 21st century skills and problem-solving skills in relation with English comprehension, age, achievement (Mathematics and English) and gender.

For the second phase data collection, a demographic information questionnaire and problem-solving skills test were added to the final PoTCS scale (as seen on Step 6). Related variables for the research questions are stated in *Table 5*. English comprehension was recoded as increasing when English comprehension become better (1: Not so good till 4: Mother tongue).

Table 5

Variables Related to the Sub-Research Questions

Variables	Types of data Continuous vs. Discrete	Categories of variable and codes (if coded as discrete)
PoTCS scores	Continuous	-
Problem- solving skills test scores	Discrete	Possible points for each question: 0,1,2,3
Programme type	Discrete	IB (1), Non-IB (0) school
Gender	Discrete	Female (1), Male (0)
Age	Discrete	12,13,14,15,16
English Comprehension	Discrete	Not so good (4), Good (3), Very good (2), Mother tongue (1) [*recoded*]
Mathematics achievement	Continuous	-
English achievement	Continuous	-
Participation duration in IB programme	Discrete	1 year or less than 1 year (1), 2 years (2), 3 years (3), 4 years (4), 5 years or more (5)

In the Netherlands, conducting research in a school is based on a permission given by the school leaders. Each school is autonomous to decide if they are interested in any kind of research application. Hence, the research purpose and procedures are communicated with head of schools via email. In the Netherlands, for any age level of school, most schools are very interested in improving and assessing their students' 21st century skills and specifically problem-solving skills in order to prepare their students for future. Based on this, some school leaders were very interested in attending the current study. Schools have responsibility to inform the parents about any research conducted in their schools in terms of the EU law GDPR (General Data Protection Law). For this purpose, a Parent Information Letter (see Appendix A) were sent out to parents for the voluntary schools.

Covid-19 epidemic had made a big impact on the data collection for the current study as mentioned in 'Settings and Participants' section as well. In the middle of the data collection process, the 'intelligent lockdown' in the Netherlands began, meaning all level of schools started to have distant education. This lockdown had caused half a year delay in completing the data collection process. Besides, even very interested schools had struggle to find suitable time for the application. Therefore, the number of questionnaires collected were much less than expected.

Instruments

In this study, initial PoTCS scale, final PoTCS scale, demographic information questionnaire and problem-solving skills test were used to collect data. The initial form of PoTCS scale was used to investigate the construct validity (EFA) and scale reliability. Final PoTCS scale was firstly used for CFA and scale reliability, and then used together with demographic information questionnaire and problem-solving skills test to analyze the relations between the related variables, problem-solving skills test scores and PoTCS scale scores.

Demographic information scale includes questions for; gender, grade/year group, age, English level, programme type (IB vs non-IB). Additionally, the questionnaire consists of questions for the overall grade average, English language score and Mathematics score on student's last report to get information

about the achievement levels of students. Demographic information questionnaire can be seen on Appendix E.

Calculating the overall grade average was not easy for most of the IB school students and so they skipped the overall grade average question. Because of too many missing data, the general grade average variable was not included into the data set. Achievement was considered in terms of English language score and Mathematics score for the current study. These two subjects are two of the mandatory subjects for both IB and non-IB school students. Study group consists of students who follow different types of educational programmes and have different grading systems. For example, in a school which follows the Dutch system, grading is done out of 10 or 100 whereas in an IB school, grading is done out of 8 for each criterion and out of 7 for overall average of four criteria. In order to standardize the achievement indicators, students' scores were converted to a value out of 100. In order to convert the score of a student who attended to a Dutch school, the score out of 10 is multiplied by 10. On the other hand, to convert the grade/score of a student who attended to an IB school, conversions in *Table 6* and *Table 7* were used. These levels for the total of four criteria are taken from the assessment policy of an International School in the Netherlands.

Table 6

Grade Conversion for out of Eight

Levels	Total for four Criteria	Midpoint	%
1	0-4	2	6
2	5-8	6	19
3	9-12	10	31
4	13-16	14	44
5	17-20	18	56
6	21-24	22	69
7	25-28	26	81
8	29-32	30	94

Table 7

Grade Conversion for out of Seven

Levels	Total for four Criteria	Midpoint	%
1	0-5	2.5	8
2	6-9	7.5	23
3	10-14	12	38
4	15-18	16.5	52
5	19-23	21	66
6	24-27	25.5	80
7	28-32	30	94

MYP uses four criteria to decide the final score for Mathematics and English language. Each criterion is graded with a level out of eight. The total can take values between zero and 32. According to the application period of the questionnaire, some students had a grade out of 8 and some students had a grade out of 7. Therefore, two different scaling were used. To calculate the percentage, the midpoints of each interval were taken, which was divided by 32 and multiplied by 100. Of course, these conversions are not as accurate as an overall grade average which is written out of a 100, but grading out of 7 or 8 also has some accuracy issues in itself. Hence, these conversions were accepted valid enough with their limitations for the current study purposes.

The summary of data collection instruments related with the research questions can be seen in *Table 8*. The development processes of PoTCS scale and problem-solving skills test are stated on the following sections.

Table 8

Data Collection Instruments

Research Questions	Data Collection Instrument
Question 1	Final PoTCS scale and Demographic information questionnaire
Question 2	Problem-solving skills test and Demographic information questionnaire
Question 3	Final PoTCS scale, Demographic information questionnaire and Problem-solving skills test
Question 4	Final PoTCS scale, Demographic information questionnaire and Problem-solving skills test
Question 5	Final PoTCS scale and Demographic information questionnaire
Question 6	Problem-solving skills test and Demographic information questionnaire
Question 7	Final PoTCS scale and Demographic information questionnaire

The perception of 21st century skills scale (PoTCS scale). In order to develop this scale, generally Cohen and Swerdlik (2010)'s five stage process of "test conceptualization, test construction, test tryout, item analysis and test revision" (p.233) is followed. In more detail, following steps (Crocker & Algina, 2008) were followed for the development process of PoTCS scale:

1. The conceptual framework was defined and indicators representing the construct were identified. The conceptual framework of this study is based on the framework stated in Binkley et al. (2012) as a reference. The framework based on ten skills under four categories are as follows (p.36).

    Ways of Thinking (*WoT)*

    1. Creativity and innovation
    2. Critical thinking, problem solving, decision making
    3. Learning to learn, metacognition

    Ways of Working (*WoW)*

    4. Communication
    5. Collaboration

    Tools for Working (*TfW)*

    6. Information literacy
    7. ICT literacy

Living in the World (*LiW)*

8. Citizenship-local and global
9. Life and career
10. Personal and social responsibility

The framework formed by Binkley et al. (2012) was chosen to guide development processes of 21st century skills scale due to its comprehensive structure and sufficient broadness.

2. A set of short-scenarios (items) based on each skill in the framework were prepared. For this step 15 volunteer students (aged 12 to 16) wrote stories about the competencies in 21st century skills framework. These students became volunteer out of 90 students because of their interest in creative writing.

3. An initial item pool was constructed with three short scenarios for each of 10 skills (under four dimensions), in total 30 items. The scale was created as a 5-point Likert-type scale, scaled from 5: Totally suits me until 1: Totally not suits me, where higher option indicates a more favorable answer. According to several researches, 5-point Likert-type scales get higher response rate and quality data, and they cause less frustrations for respondents (Babakuş & Mangold, 1992). Additionally, it is an odd number rating scale, so it provides higher reliability and validity values for a scale (Şencan, 2005).

4. Three native speaker, English Language and Literature teachers reviewed the items in terms of language and items were edited according to experts' feedbacks.

5. Three measurement and evaluation experts with PhD degrees reviewed the scales in terms of content validity. It is considered that evaluation of the items by both language and measurement perspectives have significant importance to establish the content validity of the scale (Klein, 2011). Expert view request document can be seen on Appendix B which consists of questions about whether the item was proper to measure the given competency, if yes in which degree (1-3) and if the item was not good enough, what kinds of suggestions could the experts give. Items were edited and four items were rewritten according to a consensus between the expert feedbacks. Summary of the

expert views about the content validity can be seen on Appendix C. Only four experts' views are given on the summary, because two language experts only edit the items in terms of its English.

6. After the editing based on expert feedbacks, items were tried out with 10 students to see if there was any misunderstanding and to determine the time needed for completing the scale. At the end of this process, it was determined that the items were clear for the target age groups.
7. Field-test (pilot study) of 30 items had been done with 379 students. The initial form of PoTCS scale can be seen on Appendix D. The scale was applied via google forms together with question of student's age in order to limit the age interval between 12 to 16 (lower secondary school according to OECD*). Although the scale was given via google forms, students had answered the survey during mentor/year tutor hours at each class in 15 minutes. Researcher sent a brief instruction sheet to the mentor teachers (see Appendix F). Data collection process took three weeks for the initial form of PoTCS scale.
8. After checking the assumptions of EFA for the collected data, EFA was conducted and four factor 21 items PoTCS scale was formed.
9. To confirm the structure of the PoTCS scale explained with EFA, the final form of PoTCS scale was given to a second group of students in the same age interval (12-16) and 360 students answered the final version of PoTCS scale. The final scale can be seen on Appendix G. Final PoTCS scale is scored as 'sum scores by factor'. According to this scoring, for each student, raw scores for all items loading on a factor would be added and to provide a scale and use factors with different numbers of items together, divided by the number of items in each factor (DiStefano, Zhu & Mindrila, 2009). In other word, average point of each factor would be calculated.
10. The internal reliability (stratified alpha coefficient) and test-retest reliability of the final PoTCS scale were calculated.

Study groups included in the development of PoTCS scale can be summarized with *Table 9*.

Table 9

Study groups for developing the PoTCS scale

Intention	Classification of Participants
Developing the PoTCS scale	• Fifteen students (aged 12 to 16 years old) • Three Experts (Native English) in Language • Three Experts in Measurement and Evaluation • Ten students (aged 12 to 16 years old) for trial • Exploratory Factor Analysis Group (N=379) • Confirmatory Factor Analysis Group (N=360) • Test-retest Group (N=34)*

*60 students attended to test-retest applications, but 34 students responded in both applications.

Construct validity analyses for PoTCS scale. In psychology and education, in order to develop objective tests for measuring attitudes, intelligence, skills and the like, factor analysis has a wide range of applications. Factor analysis has been used quite often during scale development processes. One of the reasons behind this is the convenience of interpretation of the results for few factors instead of many single items. Another reason is reliability of scores on factors are more than individual items (Tabachnick & Fidell, 2001). As Kline (1993) stated, factor analysis has helped to develop many good personality scales as well as many important variables in psychology.

There are two main most common use of factor analysis: Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). These two-factor analysis were used to reduce the observed variables to a smaller number of factors and testing the structure of the PoTCS scale.

Exploratory factor analysis (EFA). EFA is the most commonly used factor analysis method, which aims to simplify a big set of variables (items) in order to determine the most important and prominent variables (factors) (Kline, 1993; Tabachnick & Fidell, 2001). Nunnally and Bernstein (1994) and Kline (1993) mention factors are basically the linear combinations of variables and these variables can be included into this combination with different proportions. Because of this different proportionality of variables in the factors, there are many different possibilities to decide the number of factors and structure of these factors. SPSS 23.0 was used for the EFA.

Before conducting EFA, the data set need to be controlled in terms of factorability as well as the assumptions of the analysis such as missing data, outliers, sample size, normality, linearity, multicollinearity and singularity, and necessary measures need to be taken if needed (Kline, 1993; Tabachnick & Fidell, 2001; Klein, 2011; Çokluk, Şekercioğlu & Büyüköztürk, 2012).

Assessing the factorability and assumptions for EFA. For the first phase of data collection, there were 321 students, aged 12 to 16 answered the PoTCS scale. The 30-item scale was given to students as an online form, and each item was required to be answered in order to complete the questionnaire. Therefore, for 321 students, there was no missing data.

Statistical calculations are sensitive to the outliers because generally statistics are dependent on squared deviations of values from the mean of the distribution (Mertler & Vannatta, 2005). During the current study, univariate outliers were identified as considering the standard z scores. z scores which are less than -3.29 and more than 3.29 are considered as outliers (Tabachnick & Fidell, 2001). In order to evaluate the univariate outliers, all scores had been converted to standardized z scores. All standardized scores except few (less than ten) provided the two-tailed 3.29 condition. According to Tabachnick and Fidell (2001: 67-68), "Cases with standardized scores in excess of 3.29 ($p<.001$, two-tailed test) are potential outliers. However, the extremeness of a standardized score depends on the sample; with a very large N, a few standardized score in excess of 3.29 are expected…". In addition, according to Mertler and Vannatta (2005), if few cases excess the ±3 standard scores interval and sample size is N>100, then the interval for z scores can be widen as ±4. Therefore, because all the data were in the ±4 interval and except few all the data were in ±3 interval, it was accepted that there is not a univariate outlier problem in the data set.

Mahalanobis distances were calculated in order to evaluate the presence of multivariate outliers. Mahalanobis distance is a calculation which gives the distance of a case from the means of all variables (centroid) is an easy way to identify the multivariate outliers (Tabachnick & Fidell, 2001; Mertler & Vannatta, 2005). Multivariate outliers are identified as comparing observed Mahalanobis distance with chi-square (x^2) critical value ($p<.001$) according to degrees of freedom of considered variables (Tabachnick & Fidell, 2001). Hence, p values for

the Mahalanobis distances were calculated and nine out of 321 respondents produced scores smaller than .001 which made them considered as outliers. These cases were eliminated from the data set. Therefore, data set for EFA remained with 312 participants.

After the elimination of outlier cases, in order to check if size of the study group is suitable for the factorization, Kaiser-Meyer-Olkin measure of sampling adequacy (KMO) was calculated and this value was seen as .852 as given in *Table 10*. KMO value is evaluated as good when it is between .80 to .90 (Şencan, 2005; Çokluk et al., 2012). Hence, the data group size was considered as suitable for the factorization in terms of KMO value. Furthermore, Tabachnick and Fidell (2001) stated that required sample size is also about the ratio of sample size to estimated variables and if the estimated variables (sub-tests) have a high reliability than the ratio of 16 to 1 is adequate. As considering the theoretical framework of the scale consists of 10 skills, it is expected to measure maximum 10 sub-tests. Çokluk et al. (2012) emphasized the importance of providing at least two of the suggested criteria stated in the literature. Hence, it can be recognized that 312 participants ensure at least two of the criteria mentioned in the literature for the sample size for EFA. Additionally, Bartlett's test of sphericity had been performed and as shown in *Table 10*, the test result provided a significant value ($X^2_{210} = 1529.573$, $p < .001$) which was considered as the proof of factorability.<.001

Table 10

KMO and Bartlett's Test Results

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		.863
Bartlett's Test of Sphericity	Approx. Chi-Square	2326.878
	Df	435
	Sig.	.000

After the consideration of factorability, univariate and multivariate normality had been questioned. Because the items formed PoTCS scale are Likert type and not continuous, the univariate normality could not be provided, although the skewness and kurtosis values were in the range of normal distribution. Tabachnick and Fidell (2001) mentioned that there is not one straight forward method to check

multivariate normality. According to Mecklin and Mundfrom's (2005) literature review, there are four approaches to test multivariate normality. Two of these approaches are graphical approaches, and skewness and kurtosis approaches. Multivariate normality had been assessed via scatter graphs (as using SPSS) and Mardia's multivariate coefficient (as using AMOS) which is based on multivariate kurtosis values before the EFA. Both of these analyses were performed based on the theoretical model under four-factor structure. The scatter plots were generally not elliptical as seen in *Figure 5*, but based on the suggestion of Mertler and Vannatta (2005), scatter plots were assessed after a natural logarithmic transformation as well. Yet, the scatters did not become elliptical.

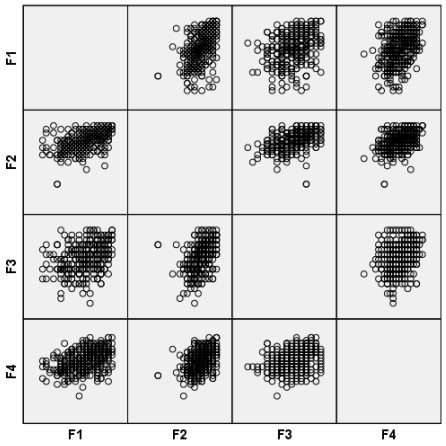


Figure 5. Scatterplot matrix of factors without transformation.

Additionally, Mardia's (1970) test of multivariate kurtosis was assessed under four factors in AMOS. For a multivariate data set the critical value is expected to be in between ±1.96 ($p<.05$). On the other hand, Mardia's test is sensitive to sample size and so there are different suggestions given in the literature for the threshold of the critical value. One is given by Bentler (2006) as the critical value larger than 5 could be considered as an indication of multivariate normality issues, and another suggestion is given by West et al. to accept kurtosis value larger than 7 as an indication of a departure from normality (cited in Byrne, 2010). Yet, the critical value for the initial data set (N=312) was found as 15.420. Therefore, the existence of multivariate normality could not be approved.

According to Fuller and Hemmerle (1966)'s Monte-Carlo simulation study with 200 observations on five different distributions and six different model their 'limited' results showed strong indication for "the maximum-likelihood estimation

procedure for factor analysis is relatively insensitive to departure from normality of distribution of the factors for large numbers" (p.266). Therefore, for the current study, the extraction methods which are not limited by distributional assumptions needed to be used (Koğar & Yılmaz Koğar, 2016; Zygmont & Smith, 2014). Unweighted Least Squares (ULS), which is based on "the minimization of eigenvalues of the reduced correlation matrix" (Zygmont & Smith, 2014, p.45) was considered as a proper method for the possible EFA. Li (2016) stated the advantages of ULS and Diagonally Weighted Least Squares (DWLS) against Maximum Likelihood (ML) and robust-ML, when the variables are ordinal and data is not symmetrical. His simulation study showed that ULS and DWLS compensated for the bias more effectively than robust ML.

Multicollinearity means some of the test items or variables are highly correlated with each other (r_{xy}>.90). On the other hand, singularity means two variables are perfectly correlated (r_{xy}=1.00) (Şencan, 2005). Especially for analyses which require to take the inverse of correlation matrix (such as confirmatory factor analysis) multicollinearity and singularity cause problems. Recognizing the individual effect of each variable would be hard. It is suggested to delete the variable which has multicollinearity or singularity (Tabachnick & Fidell, 2001).

The most used threshold for multicollinearity can be considered as VIF and TV values. Hence for the current study VIF and TV values were calculated and assessed to check multicollinearity. Some authors suggest stricter threshold than mentioned above for VIF and TV values in the literature. Based on there is no consensus about how strict researchers need to be about evaluating VIF and TV, Jongh, Jongh, Pienaar, Gordon-Grant, Oberholzer and Santana (2015) questioned whether the threshold value of VIF as low as 2.5 is too strict or not. They compared the effects of VIF values 2.5, 5 and 10 for different sample sizes. The conclusion was thresholds of 2.5 and 5 did not make much difference, and for larger data sets (N≥1000) the thresholds can be more lenient. For the current study, in order to calculate VIF and TV, each 30 items included into the initial scale was set as dependent variable in separate linear regression analyses and the rest of the items as independent variables and analyzed. In another words, the analysis for the multicollinearity were repeated 30 times. For each item, VIF values

were observed less than 2 (all TVs were observed greater than .50, which was expected, because VIF and TV are inverse values according to multiplication). Hence, it is concluded that multicollinearity was not a problem for any of the items even for stricter threshold of 2.5 for VIF. Therefore, after all the pre-requisite analysis had been considered, data were analyzed using SPSS to run EFA with ULS extraction and if needed with Promax rotation (N=312).

Findings of EFA. EFA converts a subjective selection procedure of variables into an objective statistical basis (Kline, 1993). On the other hand, factor analytical procedures are multiplex processes and identifying the number of factors and deciding the variables included into these factors still have subjective decisions in them. Therefore, experimenting with the number of factors and evaluate the effect of different factor rotation techniques can be valuable to obtain the simplest structure for the PoTCS scale.

In order to decide the number of factors to extract, initial EFA was run with ULS as extraction method and showed eight components with eigenvalues bigger than 1. According to Mertler and Vannatta (2005), the eigenvalue criteria can be used reliable when the number of items is less than 30 and all communalities are more than .70. These, both conditions were not provided for the current data set, so, besides the eigenvalues, the scree plot shown in *Figure 6* was evaluated.

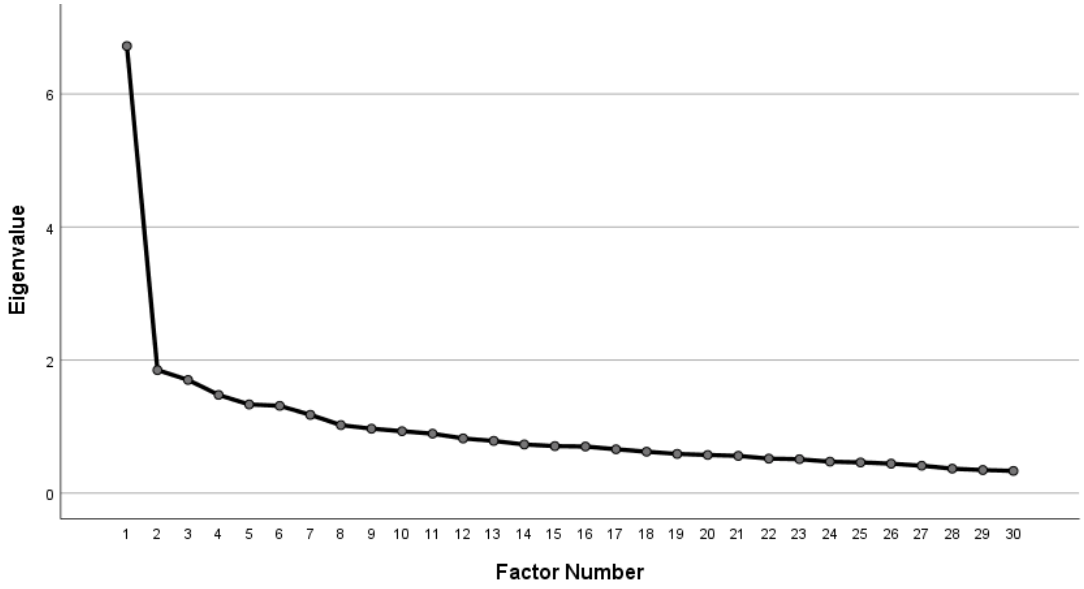


Figure 6. Scree plot for the initial form of PoTCS scale

It is clear that the differences between the eigenvalues in another words how much each factor adds on the explained variance get lower after the first

factor. On the other hand, after the fourth factor, there is not any decrease happening, after the fifth factor decrease continues smoothly until the eighth factor and then plot forms a horizontal trend. Thus, deciding the number of factors based on the eigenvalues and the scree plot was not easy and again based on Mertler and Vannattta (2005), scree plot is mostly reliable when the number of respondents are more than 250 and communalities are bigger than .30. The number of individuals is more than 250 for the current study, whereas there were some items with communalities smaller than .30.

Therefore, as an alternative for the number of factor decision, parallel analysis was run as using a tool developed by Vivek et al. (2017) based on Horn (1965)'s test. Horn (1965)'s test to decide the number of factors is based on the comparison between the eigenvalues extracted from the researcher's dataset and the eigenvalues from a randomly generated correlation matrix. In the engine, the number of variables in the current study's data set, sample size and type of analysis were set. The number of factors were decided as the number of eigenvalues generated from the current dataset that were larger than the corresponding random eigenvalues, which was a four-factor structure. Despite this outcome of the parallel analysis and the conceptual four-factor framework (KSAVE Model) that PoTCS scale created based on, EFA was repeated for two, three, four and five factors in order to observe the simplest structure which explains the most variance. With five factor structure, the fifth factor had less than three items after the low loaded and overlapping variables were eliminated, so five factor model was not considered for the decision of number of factors.

The possible rotation options are orthogonal and oblique rotations for a factor analysis. Orthogonal rotations are limited to 90° rotations only, whereas the axes for oblique rotations can take any positions, because of the less than 90° possibility (Kline, 1993). Additionally, in orthogonal rotation, the factors are expected to be uncorrelated, whereas in reality it is not likely to get factors which are uncorrelated. Oblique factors are accepted as correlated (Kline, 1993), so oblique rotations were preferred for the current study. Direct Oblimin and Promax are two of the oblique rotation methods. According to some comparison studies that Kline (1993) summarized, Direct Oblimin method is the most efficient method to get the simplest structure, but Promax can provide good structure as well. Both

of the methods were analyzed for the current study and findings came out very similar. Therefore, it is decided to present the findings of the Unweighted Least Squares (ULS) extraction method, with four factors after Promax rotation (Kappa was taken as 4) (Nunnally & Bernstein, 1994). The summary of initial explained total variances for 30 items, number of items and explained variances after the necessary item eliminations and explained total variances after the item eliminations with ULS and Promax rotation are given in *Table 11*.

Table 11

Explained Variance Summary of Different Number of Factor Trials

Number of Factors	Initial Explained Total Variance	Number of Items After Necessary Eliminations	Initial Explained Total Variance After Item Eliminations*	Explained Total Variance with ULS After Item Eliminations*
2	28.565 %	18	36.147 %	28.516 %
3	34.238 %	15	47.881 %	35.358 %
4	39.158 %	21	46.988 %	34.749 %

**After Promax Rotation*

As seen in *Table 11*, the maximum explained variance was observed for three factor structure. On the other hand, explained variances were very close for three and four factor structures, so the decision for the number of factors needed to be done in between three and four factors. As considering the theoretical framework, eigenvalues, scree plot, parallel analysis and experimenting with different numbers of factors, it is decided to form PoTCS scale with 21 items and four factors. The eigenvalues and explained variances for PoTCS scale before eliminating any items due to the insufficient factor loading or loading on multiple factors are presented in *Table 12*.

Table 12

The Eigenvalues and Proportions of Variances for PoTCS Scale (n=30)

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.720	22.401	22.401	6.072	20.241	20.241
2	1.849	6.164	28.565	1.199	3.996	24.237
3	1.702	5.673	34.238	1.074	3.579	27.816
4	1.476	4.920	39.158	.817	2.724	30.540

Table 12 shows the eigenvalues and the proportion of the explained variance by each of four factors. Before eliminating any items due to the insufficient factor loading or loading on multiple factors, four factors explained 30.540% of the total variance. This value would be 39.158 % if PCA would be used instead of ULS.

Factor loadings represent “a unique relationship between the factor and the variable” (Tabachnick & Fidell, 2001, p.625). The factor loadings need to be evaluated in terms of the insufficient factor loading or loading on multiple factors. For the current research, to decide the loading on multiple factors, the difference between the factor loadings were considered and interpreted as loaded on multiple factors if the difference was equal or smaller than .10 (Çokluk et al., 2012). Items were eliminated one by one as starting from the item which shows smallest difference between factors. To decide the insufficiency of the factor loadings, .32 is considered as the threshold value (Tabachnick & Fidell, 2001). The items with factor loading of less than .32 were eliminated one by one as starting from item with the lowest factor load. Eliminations started from the items which had both insufficient factor-loading and loading on multiple factors problems. Elimination of an item changes the factor loadings of the remaining items on factors, so the order of elimination needed to be considered very carefully. When there was a dilemma between two items to eliminate, elimination was repeated for both items repeatedly as keeping one of the items and eliminating the other and vice versa. Eventually, it is decided to eliminate the items 1, 6, 8, 11, 14, 21, 25, 27 and 28 from the PoTCS scale, and final form of PoTCS scale remained with 21 items.

After EFA was completed as using ULS for extraction and Promax as an oblique rotation method, PoTCS scale was formed by four factors. These factors were named as; Factor 1: *Being a Responsible Global Citizen*, Factor 2: *Using Information and Technology Effectively*, Factor 3: *Learning to Learn*, Factor 4: *Communication and Collaboration*. Factor loadings of the final form of PoTCS scale are presented in *Table 13*. Item numbers were reorganized for the 21-item version of PoTCS scale after the EFA and item eliminations. Item numbers are stated as ‘*old item number*’ for the initial form and ‘*new item number*’ for the final form of PoTCS scale.

Table 13

Factor Loadings of PoTCS Scale Under Four Factors

Old item number	New item number	Factor* Loadings (1)	Factor Loadings (2)	Factor Loadings (3)	Factor Loadings (4)
4	17	.494	-.121	.312	-.039
22	16	.713	.141	-.154	-.079
23	3	.527	.180	-.068	-.023
24	8	.554	-.192	.204	.177
26	19	.377	.222	-.072	.147
29	15	.622	.144	-.110	-.055
30	21	.474	-.088	.101	.039
16	18	-.090	.626	.168	.078
17	9	.077	.501	.246	.018
19	14	.109	.645	.037	-.062
20	20	.082	.519	.036	.091
3	5	-.180	.216	.352	.000
7	2	.000	.045	.659	-.040
9	6	.053	.335	.462	-.111
18	13	.145	.194	.347	.009
2	1	.098	-.084	.092	.448
5	4	-.050	-.128	.287	.400
10	7	-.056	.240	-.104	.454
12	11	-.198	.227	.042	.483
13	12	.127	-.014	-.119	.455
15	10	.196	.177	-.129	.343

**All factors are extracted with ULS and rotated with Promax.*

As seen on *Table 13*, the first factor which was named as *Being a Responsible Global Citizen (GC)*, is formed by the items 4, 22, 23, 24, 26, 29 and 30. The factor loadings of these items vary between .377 and .713. The second factor which was named as *Using Information and Technology Effectively (ICT)*, is formed by the items 16, 17, 19 and 20. The factor loadings of these items vary between .501 and .645. The third factor which was named as *Learning to Learn (LtoL)*, is formed by the items 3, 7, 9 and 18. The factor loadings of these items vary between .347 and .659. The fourth factor which was named as *Communication and Collaboration (ComCol)*, is formed by the items 2, 5, 10, 12, 13 and 15. Factor loadings of these items vary between .343 and .483. Item-total and item-factor correlations of the final form of PoTCS scale are presented in *Table 14*.

Table 14

Item-Total and Item Factor Correlations of PoTCS Scale

Old item number	New item number	Item Correlation with factor 1	Item Correlation with factor 2	Item Correlation with factor 3	Item Correlation with factor 4	Corrected Item-Total Correlation
4	17	.640*	.228*	.294*	.211*	.500*
22	16	.721*	.359*	.131*	.230*	.527*
23	3	.646*	.341*	.193*	.245*	.514*
24	8	.675*	.266*	.293*	.337*	.570*
26	19	.596*	.390*	.204*	.350*	.546*
29	15	.674*	.345*	.175*	.236*	.518*
30	21	.599*	.230*	.170*	.205*	.444*
16	18	.292*	.764*	.343*	.345*	.552*
17	9	.381*	.736*	.395*	.340*	.596*
19	14	.379*	.769*	.299*	.264*	.548*
20	20	.364*	.746*	.266*	.333*	.551*
3	5	.056	.213*	.614*	.133*	.310*
7	2	.201*	.224*	.758*	.216*	.452*
9	6	.311*	.402*	.672*	.221*	.518*
18	13	.316*	.351*	.628*	.274*	.514*
2	1	.279*	.209*	.196*	.539*	.429*
5	4	.170*	.136*	.265*	.568*	.394*
10	7	.240*	.323*	.130*	.621*	.452*
12	11	.160*	.314*	.197*	.640*	.442*
13	12	.242*	.204*	.111	.587*	.404*
15	10	.343*	.325*	.196*	.575*	.500*

**p<0.01*

As *Table 14* presents, the corrected item-total correlation coefficients which show the correlations between students' answers for each item of PoTCS scale and the whole scale vary between .310 and .596 and they are significant at 0.01. This finding can be interpreted as the skills that each item measures and the skills that the whole scale measures are consistent. When the correlations between each item and factor in which item is included are examined, it is seen that the correlations vary between .539 and .769 and they are significant at 0.01. This finding can be interpreted as the skills that each item measures and the factor that item belongs to aim to measure the same skill. The proportion of explained variance for each factor are presented in *Table 15*.

Table 15

The Eigenvalues and Proportions of Variances for PoTCS Scale

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.223	24.870	24.870	4.598	21.897	21.897
2	1.725	8.215	33.085	1.094	5.209	27.106
3	1.522	7.246	40.331	.901	4.292	31.398
4	1.398	6.657	46.988	.703	3.350	34.747

Note. *n* = 21.

According to the findings, 21 items under four factors explain 34.747% of the total variance. This value would be 47 % if PCA would be used instead of ULS. In this explained variation percentage, factor 1, *Being a Responsible Global Citizen,* explains 21.897%, factor 2, *Using Information and Technology Effectively*, explains 5.209%, factor 3, *Learning to Learn*, explains 4.292% and factor 4, *Communication and Collaboration*, explains 3.350%.

Explained variation percentages presented in *Table 15* are not high as considering the multi-factor structure of the scale. On the other hand, as considering the complexity of the structure that PoTCS scale try to measure, it can be interpreted that there are other factors which explain the PoTCS other than these four factors. Additionally, item types used in the scale were designed as situational judgement tests (SJTs), which were expected to generate low factor loadings and reliability compared to straightforward single sentence items structures (Ron, 2019; Sorrel et al., 2016). Therefore, the explained variation proportion is considered as acceptable for this study.

For assessing the discriminant validity, loadings on multiple factors for each item were examined and differences between loadings were expected to be higher than .10. Additionally, correlations between factors expected to be less than .70 to provide discriminant validity (Gaskin, 2013). Correlation coefficients between factors formed PoTCS scale are given in *Table 16*.

Table 16

Correlation Coefficients of Factors of PoTCS Scale

Factors	GC	ICT	LtoL	ComCol
GC	1.00	.469*	.323*	.395*
ICT		1.00	.433*	.426*
LtoL			1.00	.311*
Co&Col				1.00

Note. *$p < 0.01$.

As considering *Table 16*, correlations between the factors of PoTCS scale vary between .311 and .469 (p<.01). Based on McDonald's (1985, p.220; as cited in Rönkkö & Cho, 2020) explanation, discriminant validity can be accepted as provided if "the common factors are correlated, but the correlations are low enough for the factors to be regarded as distinct 'constructs'". Therefore, the correlation coefficients in *Table 16* can be considered as evidence for discriminant validity. In order to assess the construct validity, four-factor model of PoTCS scale were assessed with confirmatory factor analysis (CFA).

Confirmatory factor analysis. CFA is another method of factor analysis, that aims to confirm hypotheses which are formed based on the literature and EFA findings (Kline, 1993). For the second phase of data collection, in total 360 students had answered the PoTCS scale (12 - 16 years old). CFA could test if the structure of four-factor and 21-items scale would be confirmed with the data provided by 360 students. Initially, the assumptions of CFA were assessed for 360 cases. The assumptions of CFA are adequate sample size, missing data, absence of univariate and multivariate outliers, providing univariate and multivariate normality and linearity for some estimation methods, absence of multicollinearity and singularity.

Assessing the assumptions for CFA. First of all, descriptive statistics were analyzed in order to check for missing data and there were no missing data for 21 items and 360 cases. As referring the sample size adequacy stated for EFA, sample size of 360 was accepted as adequate. Univariate outliers were identified considering the standard z scores. All standardized scores except seven cases had provided the two-tailed 3.29 condition and these seven cases were eliminated from the data set and so data set remained with 353 cases. Mahalanobis

distances were calculated to evaluate the presence of multivariate outliers. Seven of remaining 353 cases produced p values smaller than .001 which made them considered as outliers. These cases were eliminated from the data set and data set to conduct CFA remained with 346 participants.

After the outlier consideration and eliminating outlier respondents, univariate and multivariate normality had been evaluated. Because the items are Likert type and not continuous, the univariate normality could not be provided, although the skewness and kurtosis values were in the normal distribution thresholds given in the literature. For multivariate normality Mardia's (1970) test of multivariate kurtosis was assessed under four factors in AMOS. For the data set to use in CFA of PoTCS the critical value for Mardia's (1970) multivariate test of kurtosis was found as 8.750. As considering the thresholds mentioned for EFA [±1.95, (Mardia, 1970); ≤5 (Bentler, 2006) and ≤ 7 (West et al. as cited in Byrne, 2010], the existence of multivariate normality could not be approved. Hence, Unweighted Least Squares (ULS) and Diagonally Weighted Least Squares (DWLS) were considered as proper estimators for CFA.

The existence of multicollinearity issues was assessed with VIF values. Each of 21 items formed the four-factor structure at the end of EFA was set as dependent variable in separate linear regression analyses and the rest of the items as independent variables and analyzed. For each item, VIF values were observed less than 2. Hence, it was concluded that multicollinearity was not a problem for any of the items even for stricter threshold of 2.5 for VIF. Therefore, data were analyzed for CFA using R version 4.1.2 with Lavaan (latent variable analysis) 0.6-9 (Rosseel, 2021). Lavaan package uses DWLS as the estimator (which uses polychoric correlation matrix) when the variables are specified as ordered (for instance, for Likert-type scales). As mentioned before Li (2016) concluded with his study that DWLS and ULS outperformed ML in providing more accurate factor loadings, inter-factor correlations and structural coefficient estimates in asymmetric conditions, also outperformed robust ML in controlling type 1 error for small samples (N≤1000).

Findings of CFA. When the initial model, 21 items distributed on four-factors structure was analyzed for CFA in R (lavaan) with DWLS as estimator, fit indices were found for as; $x^2/df = 2.437 \quad (p = .000)$; $CFI = .907$; $TLI =$

$.893, RMSEA = .065$ and $SRMR = .070$. Fit indices and the decision they indicated according to thresholds are given in *Table 17*.

Table 17

Fit Indices of the Initial Model (21 items) and Thresholds

Fit indices	Good fit criteria	Fair fit Criteria	Calculated fit indices	Result
x^2/df	$0 \leq x^2/\text{df} \leq 2$	$2 < x^2/df \leq 3$	2.437	Fair fit
NFI	$.95 < \text{NFI} \leq 1.00$	$.90 \leq \text{NFI} \leq 0.95$	.896	Poor fit
CFI	$.95 < \text{CFI} \leq 1.00$	$.90 \leq \text{CFI} \leq 0.95$	.907	Fair fit
TLI	$.95 < \text{TLI} \leq 1.00$	$.90 \leq \text{TLI} \leq 0.95$	.893	Poor fit
GFI	$.95 < \text{GFI} \leq 1.00$	$.90 \leq \text{GFI} \leq .95$	.952	Good fit
AGFI	$.90 < \text{AGFI} \leq 1.00$	$.85 \leq \text{AGFI} \leq .90$	.939	Good fit
PGFI	$.95 < \text{PGFI} \leq 1.00$	$.50 \leq \text{PGFI} \leq .95$	.754	Fair fit
RMSEA	$.00 \leq \text{RMSEA} \leq .05$	$.05 < \text{RMSEA} \leq .08$	.065	Fair fit
SRMR	$.00 \leq \text{SRMR} \leq .05$	$.05 < \text{SRMR} \leq .08$	.070	Fair fit
RMR	$0 \leq \text{RMR} \leq .05$	$.05 < \text{RMR} \leq .08$	.062	Fair fit

Note. $x^2 = 445.941, df = 183$.

Tabachnick and Fidell (2001) stated that "with large samples, trivial differences between sample and estimated population covariance matrices are often significant because the minimum of function is multiplied by N-1." (p.698), which makes the analysis dependent on the size of the sample. In another words, considering a significant x^2 value as a poor fitting model might lead a wrong model decision because of the sample sizes. Bentler and Bonett (1980) suggested that the higher the sample size is the higher the probability of rejecting any model. Therefore, it is suggested to use the ratio of x^2 to the degrees of freedom (x^2/df) to be smaller than 2 as an indication for a good-fitting model (Tabachnick & Fidell, 2001), which could not be provided by the initial model, but rather an acceptable x^2/df value is provided as 2.437. Cutoff values for a good and acceptable fit in Table 17 and Table 18 are stated as a summary from some resources in literature. (Atalay, 2015; Bentler & Bonett, 1980; CFA in R with Lavaan, n.d; Çokluk et al.,

2012; Hu & Bentler, 1999; Li, 2016; Schermelleh-Engel, Moosbrugger & Müllers, 2003; Tabachnick & Fidell, 2001).

Tabachnick and Fidell (2001) classified the fit indices under five categories; 1) Comparative Fit Indices (NFI, NNFI, IFI, CFI, RMSEA), 2) Absolute Fit Index (MFI), 3) Indices of Proportion of Variance Accounted (GFI, AGFI), 4) Degree of Parsimony Fit Indices (PGFI, AIC, CAIC), Residual-Based Fit Indices (RMR, SRMR). Model was assessed with all the fit indices, except the absolute fit index. Based on the fit indices presented in Table 17, GFI and AGFI show good model fit and x^2/df, CFI, RMSEA, PGFI, SRMR and RMR showed fair (acceptable) fit, whereas NFI and TLI indicated poor fit. Although there was not any modification suggestion for the model, Q2 (item 2) showed a low factor loading (standardized estimate) (.31) and the highest standard error (.90), so model was assessed again after deleting the Q2 with R-lavaan (DWLS estimator). CFA for the model after deleting Q2 can be seen in Figure 7.

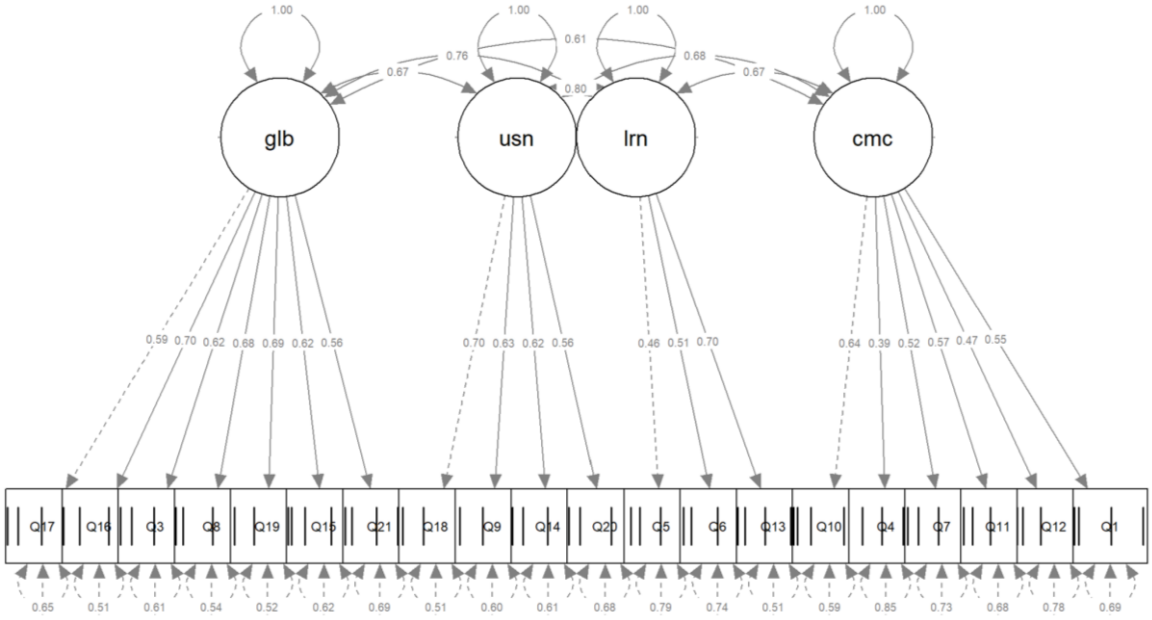


Figure 7. Four-factor 20 items CFA with standardized estimates on R

After deleting Q2, fit indices improved compared to the indices of the initial model of 21 items. Indices in Table 18 show good or fair fit indices, which can be interpreted as 20 item four-factor model represents a sufficient/fair fit model.

Table 18

Fit Indices of the Model (20 items) and Thresholds

Goodness-of-fit indices	Good fit criteria	Fair fit Criteria	Calculated fit indices	Result
x^2/df	$0 \leq x^2/\mathrm{df} \leq 2$	$2 < x^2/df \leq 3$	2.195	Fair fit
NFI	$.95 < \mathrm{NFI} \leq 1.00$	$.90 \leq \mathrm{NFI} \leq 0.95$	.924	Fair fit
CFI	$.95 < \mathrm{CFI} \leq 1.00$	$.90 \leq \mathrm{CFI} \leq 0.95$	.928	Fair fit
TLI	$.95 < \mathrm{TLI} \leq 1.00$	$.90 \leq \mathrm{TLI} \leq 0.95$	.917	Fair fit
GFI	$.95 < \mathrm{GFI} \leq 1.00$	$.90 \leq \mathrm{GFI} \leq .95$	.964	Good fit
AGFI	$.90 < \mathrm{AGFI} \leq 1.00$	$.85 \leq \mathrm{AGFI} \leq .90$	.954	Good fit
PGFI	$.95 < \mathrm{PGFI} \leq 1.00$	$.50 \leq \mathrm{PGFI} \leq .95$	.753	Fair fit
RMSEA	$.00 \leq \mathrm{RMSEA} \leq .05$	$.05 < \mathrm{RMSEA} \leq .08$	.059	Fair fit
SRMR	$.00 \leq \mathrm{SRMR} \leq .05$	$.05 < \mathrm{SRMR} \leq .08$	.065	Fair fit
RMR	$.00 \leq \mathrm{RMR} \leq .05$	$.05 < \mathrm{RMR} \leq .08$	.054	Fair fit

Note. $x^2 = 360.002, df = 164$.

The first index, x^2/df was calculated as 2.195, which indicated a fair fit. Normed fit index (NFI) is calculated as .924. Bentler and Bonett (1980) developed NFI to assess model fit as being independent of sample size and statistical significance test information. They stated .90 and above as the required NFI value for a good fitting model. Similarly, comparative fit index (CFI) and Tucker-Lewis index (TLI) show fair fit. Root mean square error of approximation (RMSEA) represents a comparative fit index as well, but it estimates the probability of lack of fit in a model compared to a saturated model contrary to aforementioned three indices. For the 20 items, four factor model RMSEA was calculated as .059, which shows around 6% probability of a not fitting model, which is an acceptable index. Under the category of 'indices of proportion of variance accounted' both goodness-of-fit index (GFI) and adjusted goodness-of-fit index (AGFI) were obtained as .964 and .954 respectively. AGFI is preferred against GFI when there are more parameters to estimate in the model (Tabachnick & Fidell, 2001). For the current model and data set, both indices showed good fit. Under the category of 'degree of parsimony fit indices' which takes the degree of parsimony into account in the model, an adjusted version of GFI, parsimony goodness of fit index (PGFI) was calculated as .753. PGFI appears fairly smaller compared to other indices. If the number of parameters estimated is smaller than the data points, than PGFI can

occur comparable to other indices like GFI (Tabachnick & Fidell, 2001). For the current study, PGFI shows an acceptable fit. Under the category of 'residual-based fit indices', the root mean square residual (RMR) was attained as .054 and standardized root mean square residual (SRMR) was obtained as .065. SRMR and RMR have a range between 0 and 1, and small values represent a good-fitting model, which shows a sufficient fit for the current study for RMR and SRMR with values .054 and .065 respectively. Therefore, it can be interpreted as 20 item four-factor model showed a sufficient/fair fit.

When the model shows an acceptable fit, parameter estimates would be examined and the ratio of each estimate to its standard error (C.R. on AMOS, z values on R-lavaan) are expected to have values higher than 1.96 or lower than -1.96 at 0.05 significance level, and higher than 2.58 or lower than -2.58 at 0.01 significance level (Hoyle, 1995; as cited in Suhr, 2006.). If this is not the case, items which do not have significant values could be considered to be eliminated from the scale/model. *Table 19* presents the standard parameter estimates, standard errors (SE) and z-values for 20 items under 4 factor structure.

When *Table 19* inspected below, it is seen that all z-values were found higher than 2.58, which show all items have significant z-values at .01 significance level (Hoyle, 1995; as cited in Suhr, 2006; Kline, 2011, as cited in Atalay, 2015), and so elimination was not needed for any item from the model. Additionally, Q17 for Being a Responsible Global Citizen, Q18 for Using Information and Technology Effectively factor, Q5 for Learning to Learn factor and Q1 for Communication and Collaboration factor did not give z-values, because parameter values of these items were fixed as 1 before the CFA was run.

Table 19

PoTCS Scale Estimates, SEs and z-values

Item No and Related Factor	Estimates	SEs	z-values
Q17 ←Globalcitizen	1.000		
Q16 ←Globalcitizen	1.187	0.110	10.8
Q3 ←Globalcitizen	1.056	0.110	9.57
Q8 ←Globalcitizen	1.155	0.098	11.8
Q19 ←Globalcitizen	1.175	0.110	10.7

Table 19 (continued)

PoTCS Scale Estimates, SEs and z-values

Item No and Related Factor	Estimates	SEs	z-values
Q15 ←Globalcitizen	1.053	0.109	9.64
Q21 ←Globalcitizen	0.948	0.102	9.30
Q18 ←UsingICT	1.000		
Q9 ←UsingICT	0.903	0.082	11.0
Q14 ←UsingICT	0.889	0.086	10.3
Q20 ←UsingICT	0.806	0.085	9.52
Q5 ←LearningtoLearn	1.000		
Q6 ←LearningtoLearn	1.106	0.157	7.04
Q13 ←LearningtoLearn	1.534	0.197	7.79
Q1 ←ComCol	1.000		
Q4 ←ComCol	0.704	0.114	6.16
Q7 ←ComCol	0.945	0.132	7.15
Q11 ←ComCol	1.027	0.126	8.16
Q12 ←ComCol	0.853	0.118	7.24
Q10 ←ComCol	1.160	0.149	7.81

For the factor *Being a Responsible Global Citizen* z-values are between 9.301 and 11.830, for the factor *Using Information and Technology Effectively* z-values are between 9.516 and 11.017, for the factor *Learning to Learn* z-values are 7.040 and 7.788, and for the factor *Communication and Collaboration* z-values are between 6.164 and 8.162.

Reliability of the perception of 21st century skills scale. Internal consistency of PoTCS scale was calculated with stratified alpha coefficient. Although Cronbach alpha is the most popular internal reliability coefficient, according to findings of some researches, Cronbach alpha coefficient do not take the factorial structure into account (Cortina, 1993) and underestimates the internal consistency value compared to stratified alpha coefficient (Kamata, Turhan & Darandari, 2003). Because PoTCS scale was formed by more than one factor and stratified alpha coefficient is an internal consistency measure especially for composite scores (Kamata et al., 2003), stratified alpha coefficient was calculated for the whole scale. Stratified alpha coefficients for the whole scale and Cronbach's alpha coefficients for each factor were calculated after both EFA and

CFA, and also for the total data as matching the common items. Reliability analyses were done with SPSS 23.0 and MS Office – Excel 2019.

After the EFA, Cronbach's Alpha coefficient for each factor of PoTCS scale were calculated as .771, .747, .587 and .618 respectively. According to George and Mallery (2010) Cronbach's alpha is generally accepted as good if it is above .7, acceptable if it is between .6 and .7, and needs to be questioned for lower values. The PoTCS scale was designed to measure the perceptions of complex skills with situational judgment test items, which brought the problem of low inter-item correlations, and test-retest reliability is suggested in the literature for situational judgement tests instead of composite reliability measures (König, et al., 2007). Additionally, because Cronbach's "alpha is very much a function of the number of items in a scale…" (Cortina, 1993, p.102), the small number of items in the sub-scales (for LtoL n=3) could be considered the reason for low values (König, et al., 2007). Yet, Cronbach's Alpha coefficient for factor 3 (Learning to Learn) and even maybe factor 4 (Communication and Collaboration) needed to be questioned. When the item-factor and item-total correlations are evaluated, item 3 (in factor 3) and item 13 (in factor 4) showed statistically insignificant very low correlations with the factors 1 (.056) and 3 (.111), respectively. Additionally, these two items had lower correlations with the total test. Therefore, eliminating these items were considered. Eliminating item 3 would increase the Cronbach's alpha coefficient from .833 to .837 for the whole scale and .587 to .592 for the Factor 3 only. These were not big differences and when the content that the items assumed to measure were considered, it was decided to keep both items in the scale. Additionally, according to Pallant (2013)'s suggestion for the scales formed by more than 10 items Cronbach's alpha values larger than .70 and for the scales formed by less than 10 items Cronbach's alpha values larger than .50 indicates reliable scales.

Stratified alpha coefficient, proposed by Cronbach, Shonenman, and McKie (as cited in Kamata, Turhan & Darandari, 2003) were calculated for PoTCS scale. Calculation of stratified alpha coefficient for PoTCS scale was done as using the formula:

$$Stratified\ \alpha = 1 - \frac{\sum_{i=1}^{k} \sigma_i^2 (1 - \alpha_i)}{\sigma_x^2}, \quad (1)$$

Where the symbol σ represents variance for each factor (i) and the total test variance (x) for 21 item PoTCS scale and α represents the Cronbach's alpha coefficient for each factor (i). Therefore, stratified alpha coefficient for PoTCS scale is obtained as 0.853, which is considered as a good reliability value. Cronbach's alpha coefficient and test-retest correlation coefficients of each factor can be seen together with stratified alpha coefficient and test-retest correlation coefficient of PoTCS scale in *Table 20*.

Table 20

Summary of Reliability Calculations of PoTCS Scale

	Cronbach's α				Stratified α
Factors	GC	ICT	LtoL	ComCol	
After EFA (N=312)	.771	.747	.587	.618	.853
After CFA (N=360)	.779	.647	.532	.636	.852
Total Study Group (N=672)	.766	.693	.555	.621	.849
Test-Retest Reliability (N=34)	r=.746	r=.831	r=.595	r=.769	r=.834

Test-retest reliability of PoTCS scale was aimed to be assessed as applying the scale to a group of students (12-16 years old) two times in between two weeks in order to assess the stability of the scale over time. Correlation coefficient was calculated between the first and second application scores. Unfortunately, there had been a three weeks Covid-19 lockdown on December 2021-January 2022 which was in between two applications. Time interval in between two applications had to be longer than the planned two weeks, and 26 scales did not have a matching because of missing students in the class due to the epidemic related reasons. Correlation coefficients showed similar values with the internal reliability coefficients and all r values were significant ($p<.01$).

Problem-solving skills test. One of the most mentioned 21st century skills can be stated as the problem-solving. Problem-solving is not a new skill only mentioned in 21st century but had been recognized and studied a lot in 20th century as well (Voogt, Dede & Erstad, 2009). PoTCS scale aims to assess students' perceptions about their 21st century skills, especially because assessing all the included skills in the frameworks would require a lot more time, people force and money investment. On the other hand, analyzing the possible relations between students' perceptions of their 21st century skills and their problem-solving skills would be a valuable process to answer how much students' perceptions about their 21st century skills would have a relation with their problem-solving skills test score which is also accepted as one of the 21st century skills. Because of the application time effectivity, the test was formed with four questions under three different intentions. The intentions of the questions are based on PISA problem solving test; 1) decision making, 2) system analysis and design, 3) troubleshooting (National Centre for Education Statistics, n.d.).

Initially, three stories were created by the researcher based on the PISA problem-solving intentions. The story for the decision-making was called 'Flying to Turkey', and it had two questions, about the total travel cost and the total travel time. The first question was designed easier than the second one, as the second question required to use more skills to reach the full correct solution. The story for the system analysis and design was called 'Pet Sitter', and it had one question about placing the pets into proper rooms with proper caretakers as considering the given rules. This question is inspired from the Children's Camp question stated on National Centre for Education Statistics (n.d.). The last story which had an intent for trouble shooting was about finding the problem in a car, but this question was too technical. After the expert feedback, another story for trouble shooting was written as 'Cookie Production Line' and reviewed with one of the experts again. The expert view request document for problem-solving skills test can be seen on Appendix H. Five experts gave feedback for problem-solving skills test, three measurement and assessment, and two language experts. After the expert feedback, some modifications were made in language and structure of the stories and questions.

After the expert view, problem-solving skills test was given to 18 students aged 15 to 16 with the purpose of recording the time to complete the test. Additionally, students were asked to note any part that they did not understand on the test. Time to complete the test was recorded between 16 to 29 minutes. It is expected that the time needed to understand and complete the questions vary according to students' age group. Time allocated for the problem-solving skills test was decided as 30 minutes. The final form of the problem-solving skills test can be seen on Appendix I. The problem-solving skills test rubric (see Appendix J) was prepared by the researcher. The alternative correct solutions given during the pilot application were added to the rubric as well. For the first three questions, there were three points allocated for each question and six points for the fourth and the last question. In total, problem-solving skills test were assessed out of 15 marks. Problem-solving skills test was decided to be graded as percentages out of the maximum possible score per question.

Item analysis were conducted in order to investigate the validity of the problem-solving skills test with the answers of 304 students. Item difficulty and discrimination indices were calculated for polytomous items according to the classical test theory as follows (Nitko, 2004, as cited in Kilmen, 2014):

$$\text{Item difficulty } (p)=\frac{\text{item point average}}{\text{item point range}} \tag{2}$$

$$\text{Item discrimination index } (D)=\frac{\text{upper group}'s\text{ point average}-\text{ lower group}'s\text{ point average}}{\text{item point range}} \tag{3}$$

Item point average was calculated as adding 304 respondents' points for that item and dividing by the item point average. Item point range for the first two questions based on the context of 'Flying to Turkey' were substituted as three for each, because the minimum possible mark was zero and maximum mark was three. Item point range for the third question based on the context of 'Pet Sitter' was substituted also as three and item point range for the fourth question based on the context of 'Cookie Production Line' was substituted as six.

To calculate the item discrimination index, there are different percentages accepted for upper and lower groups in different sources. In the current study, based on Kelley (1939)'s suggestion, upper 27% of the respondents (N_{upper}=82) were considered as the upper group and lower 27% of the respondents (N_{lower}=82) were considered as the lower group (as cited in Crocker & Algina, 2008).

Table 21

Item Statistics for Problem-Solving Skills Test

Item	Item difficulty	Upper group's mean	Lower group's mean	Item discrimination index
Flying to Turkey 1 (DMQ1)	**0.684**	2.72	1.32	**0.468**
Flying to Turkey 2 (DMQ2)	**0.501**	2.30	0.620	**0.562**
Pet Sitter 1 (SAQ1)	**0.575**	2.71	0.190	**0.839**
Cookie Production Line 1 (TSQ1)	**0.419**	4.46	0.987	**0.579**

As shown in *Table 21,* the first question about decision making (Flying to Turkey 1) was the easiest question, whereas the trouble shooting question (Cookie Production Line 1) was the hardest question in PSS test. Timing might have an effect in this result, because there were some troubles observed during the applications for some students regarding not having enough time to solve the last question in the problem-solving skills test. Nevertheless, all four of the item difficulty values shows a level of not too hard or not too easy. When the item discrimination indices were considered, all four items showed a discrimination value above .40. Therefore, according to Ebel's (1965) criteria, each item in the problem-solving skills test functioned quite satisfactorily (as cited in Crocker & Algina, 2008).

As reliability consideration, internal consistency of problem-solving skills test was assessed with Cronbach's alpha coefficient and it was found as .528 for four questions. When item-total statistics were assessed, it was seen that the Cronbach's alpha would increase to .662 without the last item, which showed the timing problem one more time. As considering the measurement model and factor

loadings later, researcher decided to eliminate this last question from the problem-solving skills test.

Data Analysis

The purpose of this study is to investigate the effect of programme type (IB vs non-IB) on students' perceptions of 21st century skills and their problem-solving skills in relation with students' English comprehension level, age, Mathematics and English achievement and gender. Before starting to analysis, English comprehension variable was recoded reversely in order to make values of the variable increasing when students' English comprehension becomes better. This makes the variable coded as; 1: Not so good, 2: Good, 3: Very good, 4: Mother tongue.

For data analysis, the researcher used MS Office – Excel 2019 (for Data Organization), SPSS 23.0 (for data purification, descriptive statistics and one-way ANOVA), AMOS 23.0 for Structural Equation Modeling (SEM)) in order to investigate the research questions. Out of 304 collected questionnaires, one case was deleted because of missing gender information, one case was deleted because age was filled as "smaller than 12", three cases were deleted because of zero standard deviation (no variation) for the PoTCS scale items, and three cases were deleted because of not answering Problem-solving skills test. Therefore, analysis proceeded with 296 cases.

To answer the first research question, firstly average points for each factor and for the whole perception of 21st century skills (PoTCS) scale were calculated as new variables and formed five new columns on SPSS. Then descriptive statistics of these four factors' (GC, ICT, LtoL and ComCol) and the whole PoTCS scale were calculated regarding the programme type (IB vs non-IB). To answer the second sub-research question, descriptive statistics of the problem-solving skills (PSS) test for each question and for the whole test were calculated for IB and non-IB school students (N=296). There were three questions (the last question [TSQ1] was eliminated from the test) in PSS test and each question was awarded with maximum of three points and in total nine points.

For both research questions, if there was a consistent difference between two groups regarding the programme type, statistical significance of the group

differences in terms of PoTCS scale and PSS test were assessed. Before any analysis, variables were tested in terms of normality with Kolmogorov-Smirnov test, stem-and-leaf plots, Normal Q-Q plots and Box-and-Whisker diagrams. For PoTCS scale, almost each factor showed a nonnormal distribution when considered separately for IB and non-IB groups (only GC was normal for non-IB group and the whole scale average was normal for IB group). Therefore, nonparametric analysis technique, Mann Whitney-U was performed to investigate the group differences on PoTCS scale and its factors in .05 significance level. Similarly, questions of PSS test and its total score distribution was nonnormal as well and so statistical significance of group differences were questioned with Mann Whitney-U test.

To address the third, fourth, fifth and sixth sub-research questions, Structural Equation Modeling (SEM) was preferred by the researcher. For the third and fourth sub-research questions model established by the researcher was tested with the related paths, and for the fifth and sixth sub-research questions multigroup analysis were conducted in AMOS based on the established model.

For the seventh sub-research question, one-way ANOVA was used to compare the group means of multiple groups (duration of attendance to the IB programme). For this research question, only the IB school students' data was included into the analysis. There were 142 students in the current study group who followed the IB curriculum. Two cases were eliminated because of no variation in the answers of PoTCS scale. Therefore, analyses were done for 140 IB students. Prior to analysis, IB duration variable (duration of attendance to the IB programme) was recoded as; 1 year or less were coded as 1 (short duration), 2 and 3 years are coded as 2 (moderate duration), and 4 and more years were coded as 3 (long duration) in order to make the interpretations easier.

Assumptions for one-way ANOVA was checked before running the analysis. The dependent variable students' PoTCS scale scores are the average of each factor of PoTCS, which provided the condition of the dependent variable to be at least an interval scale. To check the normality assumption for each factor at each level of the independent variable, Kolmogorov-Smirnov test was run and Comcol was the only factor provided the normality for each category, whereas GC, ICT and LtoL factors did not provide the normality for the moderate and long duration

groups ($p<.05$). Homogeneity of variance was controlled with Levene's test and homogeneity of variance was provided for each factor (Büyüköztürk, 2012). Therefore, to analyze the group differences of PoTCS according to the IB duration, one-way ANOVA was used for ComCol and Kruskal-Wallis test was used for the three factors (GC, ICT and LtoL).

Structural Equation Modeling (SEM). SEM is a confirmatory statistical technique which needs to have models based on prior knowledge and hypothesis about the variables and possible relations in between those variables (Tabachnick & Fidell, 2001). In other words, to form a model which includes multiple relations between miscellaneous variables in SEM, this model needs to be supported by theory. SEM gives some advantages to the user such as using multiple independent and dependent, observed and latent, continuous or categorical variables in one model and estimating various relations at the same time. There are some other statistical methods, such as multiple regression, logistic regression or canonical correlation which take multiple and various variables into account as well except that SEM can analyze more complex multivariable relationships as considering both direct and indirect effects (Walker, n.d.). As Tabachnick and Fidell (2001) states "When exploratory factor analysis is combined with multiple regression analyses, you have SEM" (p.653). As Raykov and Marcoulides (2006) mentioned "…SEM provides researchers with a comprehensive method for the quantification and testing of substantive theories" (p.1). Additionally, SEM provides an advantage about the measurement error. SEM is able to estimate and remove the measurement error as examining all possible relations of factors (Raykov & Marcoulides, 2006; Şimşek, 2007; Tabachnick & Fidell, 2001).

During the current study, to analyze the research questions, researcher chose SEM which can investigate complex relations as taking the measurement errors into account and remove them during the model testing analysis. Steps of SEM is explained in following sections, but before starting to analysis of SEM, the assumptions were assessed.

Assessing the assumptions of SEM. Assumptions prior to assess SEM assumptions and data requirements are basically same with CFA. Missing data, absence of univariate and multivariate outliers, sample size, providing univariate

and multivariate normality and linearity to use ML based analysis, absence of multicollinearity and singularity, and residuals needed to be checked.

Firstly, all variables included into the model were checked in terms of missing values. Only English and Mathematics score variables had missing values, and ratio of missing value to the number of people in the study group was 5%. Because the variables are continuous, imputation with series mean was used to complete the data set for English and Mathematics scores (Çokluk, et al., 2012). This method gives the advantage to have a complete data set and running the SEM analysis with different programmes without an error. Mean of the variable does not change when imputing the missing values with the series mean, but the variation might get little lower as well.

Univariate outliers were identified for each observed variable considering the standard *z* scores. All standardized scores except six cases have provided the two-tailed 3.29 condition and these six cases were eliminated from the data set and data set remained with 290 cases. Mahalanobis distances were calculated to evaluate the presence of multivariate outliers. Three of remaining 290 cases produced p values smaller than .001 which made them considered as outliers. These cases were eliminated from the data set and data set for SEM remained with 287 participants.

To determine the adequacy of the sample size, there are different suggestions in the literature as mentioned for the EFA section. Generally, the number of cases needed for SEM is related with some characteristics of the model estimated and distributional characteristics of the data. For instance, more complex models which have more parameters would need bigger samples to get more stable results. Additionally, if distributions of continuous variables in a model are normal and have linear relations with one another, for such models smaller sample sizes would be enough (Tabachnick & Fidell, 2001). Klein (2011) stated that about 200 cases can be considered as a recommended sample size for a medium size model with regard to the previous published articles about SEM. Therefore, 287 cases were accepted sufficient to proceed the structural equation modeling.

Univariate skewness and kurtosis values (±2) were in the normal distribution thresholds for the continuous variables (average points of PoTCS scale factors and Math and English scores) included into the model (Çokluk et al., 2012; Mertler & Vannatta, 2005). Programme type and gender are dichotomous variables; item scoring of PSS test, age and English comprehension are entered as ordered variables. Univariate normality check is a general practice to provide the multivariate normality, and SEM actually only assumes multivariate normality for some of its estimation methods (Tabachnick & Fidell, 2001). For multivariate normality Mardia's (1970) multivariate test of kurtosis was calculated based on the constructed model. Mardia's (1970) multivariate test of kurtosis for SEM model was calculated as 5.257 in AMOS where all the univariate normality values (skewness and kurtosis) were in between ±2. As considering the suggestions about using kurtosis values in the context of SEM rather than skewness (DeCarlo, 1997; cited in Byrne, 2010), and accepting kurtosis value larger than 7 as an indication of a departure from normality (West et al.; cited in Byrne, 2010), multivariate normality was considered as not an issue. Because of different views and thresholds about kurtosis critical ratio calculated by AMOS, Mardia's multivariate test of skewness and kurtosis were run as using the Web application (https://webpower.psychstat.org/models/kurtosis/) developed by Cain, Zhang and Yuan (2017) based on R as well. Multivariate kurtosis was calculated as insignificant (0.345, $p>.05$), which indicated multivariate kurtosis is not a problem. Hence, Maximum likelihood (ML) was considered as proper estimators for SEM analysis.

Existence of multicollinearity issues was assessed with VIF (and TV) values and correlations between variables. For each variable, VIF values were less than 2 and TV values were larger than .6. Additionally, all correlation coefficients (r) between variables were smaller than .90. Hence, it was concluded that multicollinearity was not a problem for any of the variables even for stricter threshold of 2.5 for VIF. Standardized residuals were assessed together with analysis, and it was spotted that there was not a sizable residual and standardized residuals were normally distributed. Therefore, after all the pre-requisite analyses had been considered, constructed model was tested for the collected data using AMOS with ML estimator (N=287).

***Steps of structural equation modeling*.** Steps to design a SEM was summarized by Kline (2011, p.92) as a flowchart given in *Figure 8.*

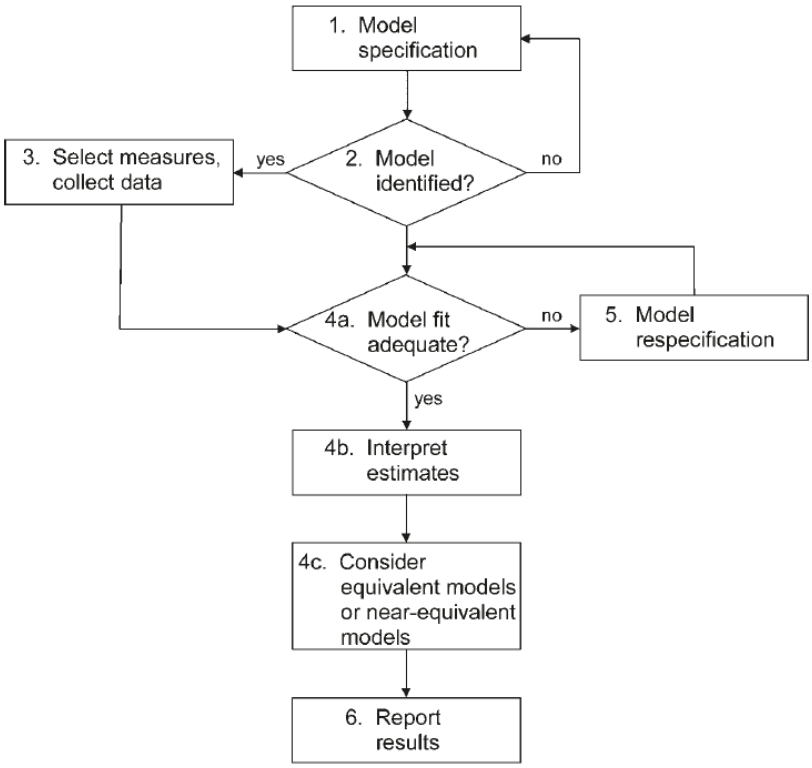


Figure 8. Flow chart of the basic steps of SEM

In order to construct, test and interpret the model during the current study, flow chart given in *Figure 8* had been followed. First step of SEM is model specification and it refers forming an initial model as putting hypotheses of the research together. As mentioned earlier, SEM is a confirmatory analysis. It is mostly applied in order to test a model arises from a theory. Following stages of SEM assume that model specification is fundamentally correct, which makes model specification the most important stage of SEM. This stage is mostly formed as a diagram which proposes the relations between the latent and measured variables based on the literature or the expertise of the researcher (Klein, 2011). In line with relative research findings, problem-solving skills (PSS) and perception of 21st century skills, were included as latent variables and factors of PoTCS; *Being a Responsible Global Citizen (GC)*, *Using Information and Technology Effectively (ICT)*, *Learning to Learn (LtoL)* and *Communication and Collaboration (ComCol)*, questions of PSS test, programme type (IB vs non-IB), age, and English comprehension were included as observed variables, Mathematics and English

scores were included as moderators, and gender was included as the multigroup variable into the model.

With the model established given in *Figure 9,* the direct effect of English comprehension on Perception of 21st century skills (PoTCS) and Problem-solving skills (PSS) will be assessed as well as the mediation effects of English comprehension in between programme type and these two latent variables. It is expected that the programme type effects the English comprehension and English comprehension effects PoTCS. In other words, it is expected that students who follow the IB curriculum to have higher English comprehension because of the international curriculum and 100% English instruction. Additionally, it is expected to have higher PoTCS with a better English comprehension. This expectation based on the more students have a better English comprehension, the more they can reach resources and so widen their perspectives, adapt new situations easier, communicate or collaborate with people from different countries, and so on. In short, they may have a better understanding and implication on 21st century skills. Because English is known and reported as the most spoken language on the world with 1.348 billion people according to the Szmigiera's, (2021) research and the most commonly used language on the internet in 2020 with 25.9% according to the Johnson's (2021) research, it is expected the better the English comprehension, the better the 21st century skills perception. Therefore, English comprehension is expected to be a causal result of the programme type and a causal antecedent of PoTCS. On the other hand, there is no expectation about the Mathematics or English school achievements to be the causal result of PoTCS, whereas they might strengthen the effect of PoTCS on PSS, which will be tested with moderation effect.

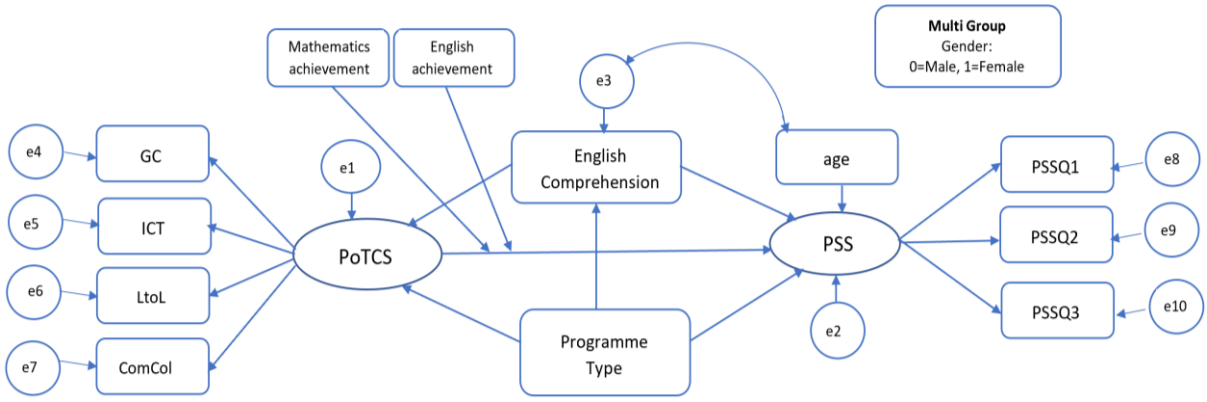


Figure 9. Constructed model to be tested with SEM

In the model, programme type and age are expected to be uncorrelated, based on each programme having similar age groups. Correlation between programme type and age was analyzed and very low statistically insignificant correlation was obtained (Spearman's rho=-.094, p=.114>.05). Additionally, there is a residual connected to English comprehension variable, because students have stated their English comprehension level subjectively and some error may be included into the answers. Lastly, English comprehension and age are expected to be correlated because it is expected students to have better language comprehension when they become older.

Second step of SEM is model identification. There are three steps to determine the identifiability of a model. First one is about the numbers of data points and the number of parameters to be estimated. In the current model, with 10 observed variables there are (10. (10+1)/2) = 55 data points. The hypothesized model indicated that 24 parameters needed to be estimated (13 regression coefficients, 1 covariance and 10 variances). Since, there are 31 (55-24) more data points than parameters to be estimated, the suggested model can be considered as over identified (Tabachnick & Fidell, 2001).

Second one is to examine the measurement model. For this part, "It is necessary both to establish the scale of each factor and to assess the identifiability of this portion of the model" (Tabachnick & Fidell, 2001, p.691). Hence, the measurement model was assessed and for each latent variable, related one observed variable's path was fixed to 1. According to Şimşek (2007),

measurement models can be assessed all together in a confirmatory factor analysis instead of running separate analysis for each measurement model. Therefore, confirmatory factor analysis was run with AMOS for two latent variables to assess how much these constructs were represented by the items described under these variables.

After running the analysis, last question of the problem-solving skills (PSS) test (TSQ1) was eliminated from the test, because of very low factor loading ($\lambda = .21$) and decreasing the reliability of the test (Cronbach's alpha .528 with TSQ1 and .662 without TSQ1). This was an expected situation, because during the data collection, it was observed that many students had trouble about timing, and could not answer the last question on time. Therefore, this last question accepted as not representative about the skill that it supposed to measure. Confirmatory factor analysis was repeated with the remaining three questions of the PSS together with four factors of PoTCS scale and the diagram with standardized path coefficients, which shows the measurement models, is given in *Figure 10*. Standardized factor loadings of PoTCS scale ranged between 0.59 and 0.71 and standardized factor loadings of *Problem-solving skills* ranged between 0.55 and 0.71.

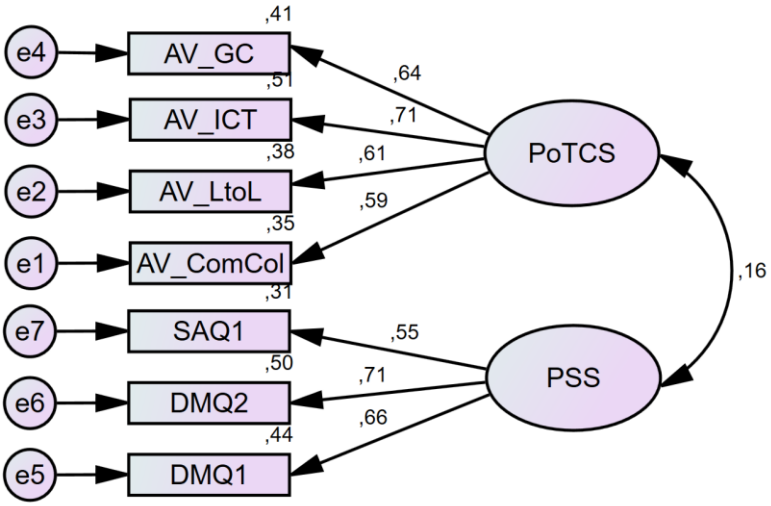


Figure 10. Standardized path coefficients of the measurement model

As seen in *Figure 10*, each factor has at least three indicators and each indicator/item load only on one latent variable. The errors for each indicator are uncorrelated, and the covariance between the latent variables is not zero. Therefore, the model may be identifiable.

Construct reliability (CR) and Average explained variation (AVE) were calculated in order to exhibit how much the observed variables included into the model representing the latent variables. In order to calculate the CR and AVE, formulas (4) and (5) were used respectively (Fornell & Larcker, 1981).

$$CR = \frac{(\sum_{i=1}^{n} \lambda_i)^2}{(\sum_{i=1}^{n} \lambda_i)^2 + (\sum_{i=1}^{n} \delta_i)} \quad (4)$$

$$AVE = \frac{\sum_{i=1}^{n} \lambda_i{}^2}{\sum_{i=1}^{n} \lambda_i{}^2 + (\sum_{i=1}^{n} \delta_i)} \quad (5)$$

For the construct reliability formula, (λ) represents the standardized factor loadings of each observed variable on the related latent variable, and (i) represents the item number (total n items). In the formula of AVE, (δ) represents the error variance.

There are two latent variables in the model, PoTCS and PSS. PoTCS has four observed variables with standardized factor loadings between 0.59 and 0.71. PSS has three observed variables with standardized factor loadings between 0.55 and 0.71. Therefore, CR values for PoTCS and PSS were calculated as 0.74 and 0.68 respectively and AVE values for PoTCS and PSS were calculated as 0.41 and 0.42 respectively. CR values showed acceptable construct reliability despite the multidimensional structure of PoTCS scale which was formed with single-response situational judgment test items (SRSJTs). On the other hand, AVE values were less than .50. According to Fornell and Larcker (1981), these AVE values (<.50) show less than 50% of the explained variance was captured by the construct and the rest was due to measurement error. On the other hand, again according to the discussion of Fornell and Larcker (1981) calculation of AVE is also not free of limitations and two important limitations for the current study could be considered as 1) both CR and AVE are summary statistics and they “cannot capture the full complexity of multivariate relationships” (Fornell & Larcker, 1981, p.48), 2) and as Bess (2001) mentioned about situational judgment tests, structure of the items (situational judgment tests) measuring the latent constructs may cause the variance problems. Additionally, as Sorrel et al. (2016) stated

multidimensional nature of SJTs might cause problems for factor analysis techniques and assumptions under alpha coefficients. In short, as considering the literature about SJTs' nature and CR being sufficient and AVE being not so low, it is decided to proceed as accepting the values of CR and AVE sufficient enough for the composite reliability and convergent validity.

Third step of model identification is assessing the structural model with only latent variables as keeping the measured/observed variables out. "If none of the latent DVs predict each other (the beta matrix is all zeros) the structural part of the model may be identified." (Tabachnick & Fidell, 2001, p.692). There are two dependent variables in the current model, but there is no feedback loop in between these two variables. Therefore, we can conclude that the model is identified, because the number of data points exceeded the number of parameters, and both the measurement and structural parts of the model were identified. After the model identification, the model was tested for the data set with AMOS. Model fit indices were assessed based on the thresholds given in *Table 17* and *18*.

After the model assessed and necessary modifications were made on the model, final model was assessed in terms of the direct effects to answer *the sub-research questions 3a, 3b, 3c and 4a, 4b, 4c*. Then, final model was assessed for the mediation effect of English comprehension in between programme type and PoTCS (*sub-research question 3d)*. Additionally, mediation effect of English comprehension was analyzed in between programme type and PSS (*sub-research question 3e)*. With the aim of checking the moderation effects, SEM analysis were run two more times with interaction effects. Firstly, model was tested with the *Mathscore* variable and its standardized interactions with the independent variable (PoTCS) of the relation (*sub-research question 4d*), and then with *Englishscore* variable and its standardized interactions with the independent variable (PoTCS) of the relation (*sub-research question 4e*). Eventually, multigroup analyses were run to assess possible differences in the final model by gender. Multigroup analyses were evaluated both for the whole model and separate paths.

Chapter 4
Findings

In this section of the dissertation findings are presented in line with the sub-research questions of the study.

Findings of Perceptions of 21st Century Skills Regarding the Programme Type

The first research question has two sub-questions as 'a) *How are the 21st century skill perceptions of students who follow an IB programme? b) How are the 21st century skill perceptions of students who do not follow an IB programme?*', which were considered together and the findings of data analysis related to these questions are presented in *Table 22*.

Table 22

Students' Levels of PoTCS Regarding the Programme Type

Variables		N	$\bar{X}$*	s.d.	Skewness		Kurtosis	
					Statistic	S.E.	Statistic	S.E.
Perception of 21st century skills (PoTCS)	IB	137	**3.81**	.49	-.180	.207	-.628	.411
	Non-IB	159	**3.53**	.40	.389	.192	1.105	.383
Being a Responsible Global Citizen	IB	137	**4.03**	.58	-.737	.207	.171	.411
	Non-IB	159	**3.65**	.58	-.002	.192	-.235	.383
Using Information and Technology Effectively	IB	137	**4.04**	.61	-.177	.207	-.786	.411
	Non-IB	159	**3.94**	.54	-.318	.192	.114	.383
Learning to Learn	IB	137	**3.82**	.76	-.564	.207	-.055	.411
	Non-IB	159	**3.28**	.72	-.082	.192	-.335	.383
Communication and Collaboration	IB	137	**3.52**	.67	-.228	.207	-.290	.411
	Non-IB	159	**3.47**	.54	.031	.192	-.017	.383

**1.00-1.79=Very Low; 1.80-2.59=Low; 2.60-3.39=Medium; 3.40-4.19=High; 4.20-5.00=Very High*.

According to the findings in *Table 22*, students' perceptions of their 21st century skills (PoTCS) are at a high level for both students who follow the IB programme and who do not follow the IB programme. For factors of PoTCS, IB students show high level of perceptions on all factors. Students who do not follow an IB programme show medium level perception on Learning to Learn factor ($\bar{X} = 3.28$, s.d.=.72) and high levels of perception towards being a responsible global

citizen, using information and technology effectively, communication and collaboration. For all 21st century skills, IB students have a higher perception about their skills compared to non-IB students. In order to assess, if these differences are statistically significant or not, Mann Whitney-U test was conducted and findings are presented in *Table 23*.

Table 23

Mann Whitney-U Test Results of PoTCS Regarding the Programme Type

Variables		n	Mean Rank	Sum of Ranks	U	p
Perception of 21st century skills	IB	137	177.89	24371	7126.50	***
	Non-IB	159	123.18	19585		
Being a Responsible Global Citizen	IB	137	177.89	24371	6865.00	***
	Non-IB	159	123.18	19585		
Using Information and Technology Effectively	IB	137	156.41	21428	9808.00	.137
	Non-IB	159	141.69	22528		
Learning to Learn	IB	137	180.88	24780	6456.00	***
	Non-IB	159	120.60	19176		
Communication and Collaboration	IB	137	153.69	21056	10180.00	.331
	Non-IB	159	144.03	22900		

** p<.05 **<.01 ***<.001*

Table 23 presents Mann Whitney-U results for students' perceptions of their 21st century skills (PoTCS) for IB and non-IB programme. According to findings, there is a significant difference between students' PoTCS scale average point regarding the programme type, U=7126.50, p<.05, r=.30, in favor of students who follow the IB programme. The differences between students' PoTCS were especially towards being a responsible global citizen (U=6865.00, p<.05, r=.32) and learning to learn (U=6456.00, p<.05, r=.35) with medium effects (DATAtab Team, 2022). Therefore, according to findings, it can be interpreted that students from both IB and non-IB programme schools perceive themselves similarly in terms *of using ICT effectively* and *communication and collaboration skills*.

Findings of Problem-Solving Skills Regarding the Programme Type

The second research question has two parts as '*a) How are the problem-solving skills of students who follow an IB programme? b) How are the problem-solving skills of students who do not follow an IB programme?*', which were considered together and the findings of data analysis related to these questions are presented in *Table 24*.

Table 24

Students' Levels of Problem-Solving Skills Regarding the Programme Type

Variables		n	$\bar{X}$	%	s.d.	Skewness		Kurtosis	
						Statistic	S.E.	Statistic	S.E.
Problem-solving skills (PSS) (max. 9 points)	IB	137	4.98	**55.3**	2.53	-.240	.207	-1.002	.411
	Non-IB	159	5.60	**62.2**	2.82	-.520	.192	-1.059	.383
Flying to Turkey 1 (max. 3 points)	IB	137	2.00	**66.7**	1.07	-.328	.207	-1.525	.411
	Non-IB	159	2.12	**70.7**	1.06	-.529	.192	-1.408	.383
Flying to Turkey 2 (max. 3 points)	IB	137	1.34	**44.7**	1.02	.058	.207	-1.169	.411
	Non-IB	159	1.67	**55.7**	1.06	-.268	.192	-1.135	.383
Pet Sitter 1 (max. 3 points)	IB	137	1.64	**54.7**	1.34	-.228	.207	-1.757	.411
	Non-IB	159	1.82	**60.7**	1.36	-.461	.192	-1.648	.383

According to the findings in *Table 24*, contrary to the findings of PoTCS scale, students' problem-solving skills (PSS) test results not only in terms of the total test score but also in terms of each question score appear in favor of the students who follow non-IB programme. Total test score was evaluated out of 9 points, and both IB and non-IB students achieved above 55% (5.5 out of 10), which is the minimum percentage required from a student to pass any subject in the Netherlands (Nuffic, 2009). In another words, both groups showed a sufficient level of problem-solving skills in the Netherlands educational system. On the other hand, IB students' PSS test average is just a pass ($\bar{X}$=4.98 [55.3%], s.d.=2.53), whereas non-IB students' PSS test average can be considered as a satisfactory level. This difference shows itself for each question, but the highest difference between IB and non-IB students appear for the second decision-making problem

(Flying to Turkey 2), which was about calculating the travel time for different trip options and deciding for the shortest travel time. Considering different stages of the trip (partly by car and partly by plane), finding the total time without missing any stage and converting the time into a common unit, comparing the times for different options and choosing the shortest and longest travel time were the skills expected for this question. IB students remained under the 55% threshold for this question ($\bar{X}$=1.34 [44.7%], s.d.=1.02), whereas non-IB students could achieve the 55% threshold ($\bar{X}$=1.67 [55.7%], s.d.=1.06). In order to assess, if these differences in PSS test which are in favor of non-IB students statistically significant or not, Mann Whitney-U test was conducted and findings are presented in *Table 25*.

Table 25

Mann Whitney-U Test Results of PSS Test Regarding the Programme Type

Variables		n	Mean Rank	Sum of Ranks	U	p
Problem-solving skills	IB	137	135.85	18611	9158.00	.017*
	Non-IB	159	159.40	25345		
Flying to Turkey 1	IB	137	143.69	19685	10232.00	.314
	Non-IB	159	152.65	24271		
Flying to Turkey 2	IB	137	134.82	18471	9017.50	.008**
	Non-IB	159	160.29	25486		
Pet Sitter 1	IB	137	141.84	19432	9978.50	.178
	Non-IB	159	154.24	24525		

** p<.05 **<.01 ***<.001*

Table 25 presents Mann Whitney-U results for students' PSS test for IB and non-IB programme. According to findings, there is a significant difference between students' PSS test total points regarding the programme type, U=9158.00, p<.05, r=.14 in favor of students who follow a programme different than the IB programme with a small effect size (DATAtab Team, 2022). Although IB students had a lower percentage of points for each PSS question, there is no statistically significant difference between the groups for the first decision making question (Flying to Turkey 1), U=10232.00, p>.05, r=.06, and for the system and analysis question (Pet Sitter 1), U=9978.50, p>.05, r=.08. The difference between students' problem-solving skills test results is especially based on the second part of the decision-making problem (Flying to Turkey 2), U=9017.50, p<.05 r=.15, despite the small effect size (DATAtab Team, 2022). Therefore, according to findings, it can be

interpreted that students from both IB and non-IB programme schools achieved the 55% threshold in almost each question of the PSS test, whereas for each question students who follow a programme different than IB programme reached to higher percentages. This difference between two groups is significant for the whole test and for the more complicated decision-making question.

Examination of the Structural Equation Model Related to PoTCS and PSS

Following four sub-research questions (3-6) are based on the model established by the researcher. First of all, the initial model had to be tested and modifications needed to be made where needed. Model test was performed by Structural Equation Modeling (SEM) in AMOS 23.0 Software related to these questions. AMOS required the variables programme and age to be uncorrelated, because there was not any path or correlation was stated on the model. So, the correlation between programme and age was assessed with Spearman's rho and the correlation was very low and statistically insignificant r=-.094 (p=.114>.05). This was coherent with the expectation as mentioned in methodology section as well. Additionally, there was a covariance stated between age and the error of English comprehension in the initial model. The reason of this covariance is the possible relation between the age and the error in reporting students' English comprehension, because students' self-awareness about their English comprehension was expected to be more accurate when they became older (Demetriou & Kazi, 2006), and so more error in reporting the English comprehension was expected for the younger students attended to the current study. *Figure 11* shows the standardized regression weight estimates and covariance between the related variables on the model.

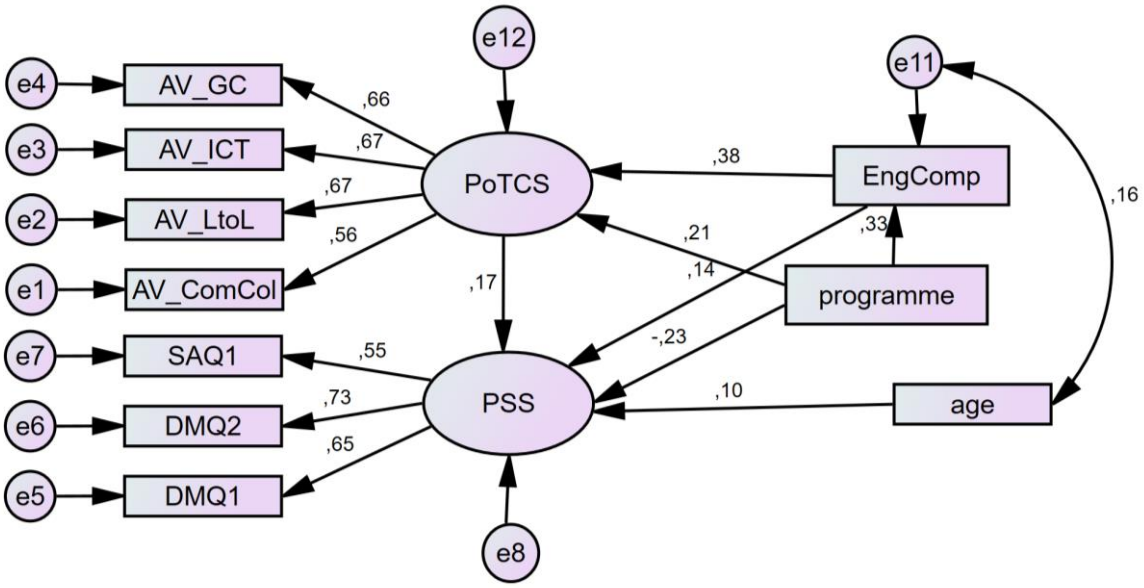


Figure 11. SEM path diagram of the initial model with standardized weights

When the model was examined, based on of Gaskin's (2021) suggestion statistical findings were assessed in an order of global to local tests, meaning first the model fit was checked (global), and then p values (local) for the paths, because if the model fit cannot be provided, p-values might have an alternative explanation (Gaskin, 2016). Fit indices of the model given in *Figure 11* can be seen in *Table 26*.

Table 26

Fit Indices of the Initial SEM in Comparison with the Thresholds

Fit indices	Good fit criteria	Fair fit Criteria	Calculated fit indices	Result
x^2/df	$0 \leq \mathrm{x}^2/\mathrm{df} \leq 2$	$2 < x^2/df \leq 3$	2.326	Fair fit
NFI	$.95 < \mathrm{NFI} \leq 1.00$	$.90 \leq \mathrm{NFI} \leq 0.95$	.870	Poor fit
CFI	$.95 < \mathrm{CFI} \leq 1.00$	$.90 \leq \mathrm{CFI} \leq 0.95$	.919	Fair fit
TLI	$.95 < \mathrm{TLI} \leq 1.00$	$.90 \leq \mathrm{TLI} \leq 0.95$	.879	Poor fit
GFI	$.95 < \mathrm{GFI} \leq 1.00$	$.90 \leq \mathrm{GFI} \leq .95$	.954	Good fit
AGFI	$.90 < \mathrm{AGFI} \leq 1.00$	$.85 \leq \mathrm{AGFI} \leq .90$	.915	Good fit
PGFI	$.95 < \mathrm{PGFI} \leq 1.00$	$.50 \leq \mathrm{PGFI} \leq .95$	.520	Fair fit
RMSEA	$.00 \leq \mathrm{RMSEA} \leq .05$	$.05 < \mathrm{RMSEA} \leq .08$	.068	Fair fit
RMR	$.00 \leq \mathrm{RMR} \leq .05$	$.05 < \mathrm{RMR} \leq .08$	.036	Good fit

Note. $x^2 = 69.778, df = 30, p = .000$.

As stated in *Table 26*, for the model tested, GFI, AGFI and RMR show good fit, x^2/df, CFI, PGFI and RMSEA show fair fit, but NFI and TLI show poor fit. Modification fit indices were examined from the AMOS output, the largest univariate modification index was given for the regression path predicting Learning to Learn factor of PoTCS from the type of programme, x^2=10.747, with an approximate standardized parameter value of .249. Because Learning to Learn factor is a part of the 21st century skills construct, and part of the research is based on questioning the effect of programme type on perception of 21st century skills (PoTCS), adding a path from the programme type to Learning to Learn (LtoL) variable was considered proper, and a model was run with this path and estimated, x^2(29, N = 287) = 54.644, *p=.003*, x^2/df=1.884, CFI=.948, PGFI=.508 and RMSEA=.056, *NFI=.898* and TLI=.919. The estimated (first) model and the modified model are nested within one another (Tabachnick & Fidell, 2001), meaning the modified model includes the estimated model in itself, so a chi-squared difference test was performed and it was concluded that adding the path predicting learning to learn factor from the type of programme improved the model significantly, x^2_{diff} (1, N = 287) = 15.134, p<.01.

After adding a regression line from programme type to Learning to Learn factor, two fit indices, NFI and x^2 significance were still indicating poor fit, and there was one more large univariate modification index given for the regression path predicting being a responsible global citizen factor of PoTCS from the type of programme, x^2=15.182, with an approximate standardized parameter value of .233. Because being a responsible global citizen factor is a part of the 21st century skills construct, and part of the research is based on questioning the effect of programme type on PoTCS, adding a path from the programme type to being a responsible global citizen (GC) variable was considered proper, a model was run with a path from the programme type to being a responsible global citizen variable was considered proper, and a model was run with this path and estimated, x^2(28, N = 287) = 34.724, p=.178, x^2/df=1.240, CFI=.986, *PGFI=.497* and RMSEA=.029, NFI=.935 and TLI=.978. The estimated (first) model, second model and the last modified model are nested within one another (Tabachnick & Fidell, 2001), so a chi-squared difference test was performed and it was concluded together with the improvement of fit indices that adding the path predicting being a responsible

global citizen factor from the type of programme improved the model significantly, x^2_{diff} (1, N = 287) = 19.920, p<.01. Modified model can be seen in *Figure 12*.

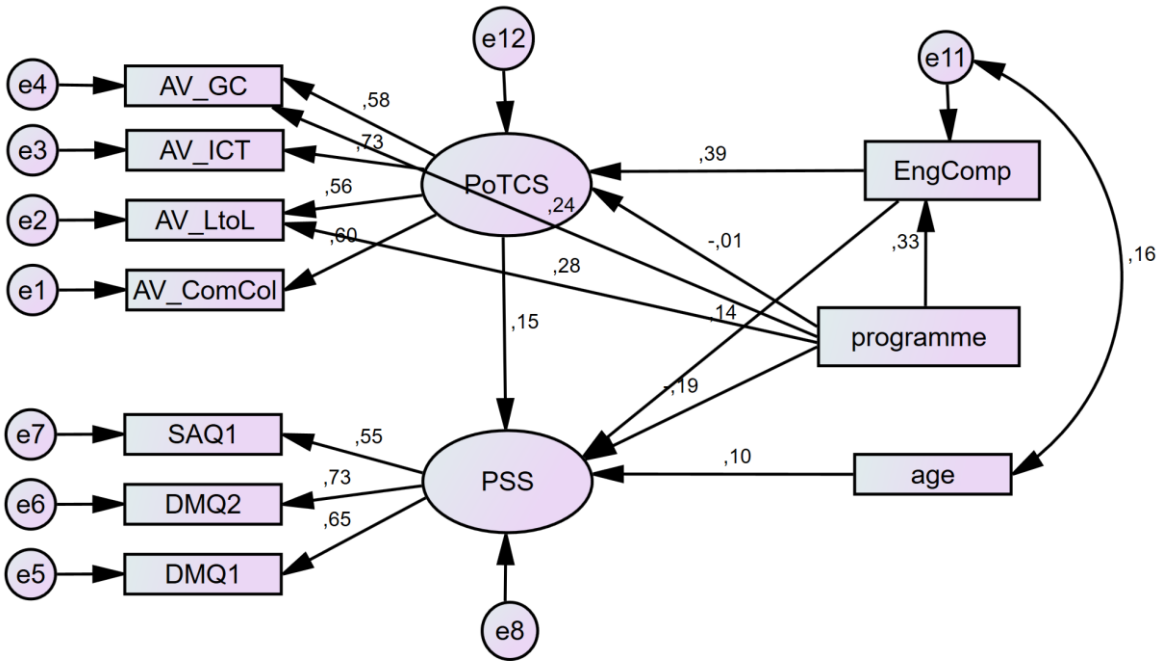


Figure 12. SEM path diagram of the model after two modifications

Two model modifications were performed in an attempt to develop a better fitting, and possibly more parsimonious model. As a conclusion, all the fit indices indicated a good fitting model, except parsimony goodness of fit index (PGFI) (.497). As Tabachnick and Fidell (2001) stated “there is a heavy penalty for estimating a lot of parameters with this index.” (p.701). Because there were two extra parameters to estimate added into the model, naturally PGFI decreased from .520 (initial model) to .497 (after adding two paths). Hence, it can be considered to eliminate some paths which are not significant on the diagram and do not disturb the theoretical background. When the model assessed in terms of the significance of regression weights (see *Figure 12*), the effect of programme on PoTCS (-.013, p=.860>.05) and the effect of age on problem-solving skills (PSS) (0.96, p=0.171>.05) were not statistically significant. The relation between age and PSS might have changed for different groups (male vs female) of multigroup analysis, so it was decided to keep the path between age and PSS. On the other hand, the effect of programme type on students’ PoTCS became insignificant, after two direct paths were added from programme type to two factors of POTCS. Additionally, programme type has significant effects on *learning to learn (LtoL)* and

being a responsible global citizen (GC), which are dimensions of PoTCS scale. Therefore, it was concluded that the effect of programme type on PoTCS latent variable was stem from *learning to learn* and *being a responsible global citizen* dimensions of PoTCS for the current model, and since the direct effects from programme type to these two dimensions were added on the model, deleting the path from programme type to PoTCS considered as reasonable. Fit indices of the final model after this last modification can be seen in *Table 27*.

Table 27

Fit Indices of the Final SEM in Comparison with the Thresholds

Fit indices	Good fit criteria	Fair fit Criteria	Calculated fit indices	Result
x^2/df	$0 \leq \mathrm{x}^2/\mathrm{df} \leq 2$	$2 < \ x^2/df \leq 3$	1.198	Good fit
NFI	$.95 < \mathrm{NFI} \leq 1.00$	$.90 \leq \mathrm{NFI} \leq 0.95$	.935	Fair fit
CFI	$.95 < \mathrm{CFI} \leq 1.00$	$.90 \leq \mathrm{CFI} \leq 0.95$	.988	Good fit
TLI	$.95 < \mathrm{TLI} \leq 1.00$	$.90 \leq \mathrm{TLI} \leq 0.95$	.982	Good fit
GFI	$.95 < \mathrm{GFI} \leq 1.00$	$.90 \leq \mathrm{GFI} \leq .95$	.976	Good fit
AGFI	$.90 < \mathrm{AGFI} \leq 1.00$	$.85 \leq \mathrm{AGFI} \leq .90$	.954	Good fit
PGFI	$.95 < \mathrm{PGFI} \leq 1.00$	$.50 \leq \mathrm{PGFI} \leq .95$	.514	Fair fit
RMSEA	$.00 \leq \mathrm{RMSEA} \leq .05$	$.05 < \mathrm{RMSEA} \leq .08$	.026	Good fit
RMR	$.00 \leq \mathrm{RMR} \leq .05$	$.05 < \mathrm{RMR} \leq .08$	.034	Good fit

Note. $x^2 = 34.755, df = 29, p = .213.$

As seen in *Table 27*, all fit indices indicate a fair or good fitting model, meaning the final model fit the data well, $x^2(26, \mathrm{N} = 347) = 34.755, \mathrm{p} = .213, CFI = .988$. Modifications were performed in an attempt to develop a better fitting and parsimonious model. On the basis of chi-square comparison tests two paths were added and on the basis of parsimony goodness fit index one path was deleted. *Table 28* presents the scaled x^2 degrees of freedom (df), CFI, and x^2 difference test.

Table 28

Comparison of the Models with Step-by-Step x^2 Differences

Model	Scaled x^2	df	CFI	x^2 Difference Test
Model 1: Initial Model	69.778	30	.919	
Model 2: Path added Programme→LtoL	54.644	29	.948	Model 1 - Model2=15.134*
Model 3: Path added Programme→GC	34.724	28	.986	Model 2 – Model3=19.920*
Model 4: Path deleted Programme→PoTCS	34.755	29	.988	Model 4 – Model3=-0.031

* *p<.01*

Because model modifications were performed, correlation coefficient was calculated between the standardized regression weights of the initially established model and the final model. Because all the skewness values were in between ±1 and kurtosis values were in between ±2, normality was assured and Pearson's Correlation coefficient was calculated, r (13) = .974, p<.01. Because there were only 13 common parameters Spearman's rho was calculated as well, r (13) = .951, p<.01. These high correlations indicates that parameter estimates of the initial model and the final model are highly related to each other (Tabachnick & Fidell, 2001). Path diagram of the final model with the standardized estimates can be seen in *Figure 13*.

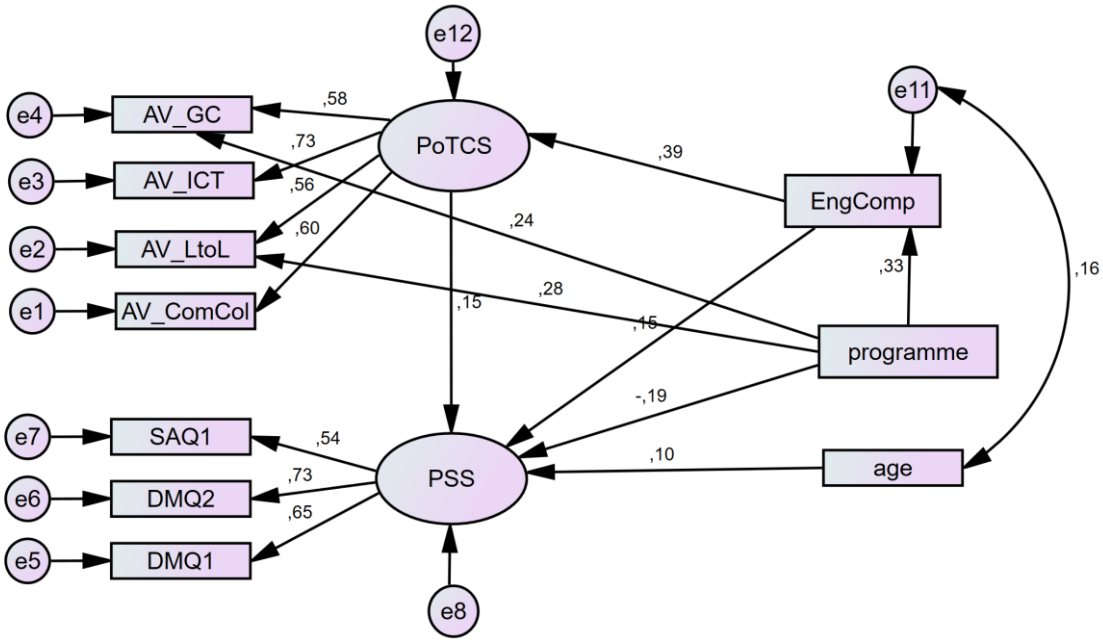


Figure 13. Final SEM model with standardized weights

Figure 13 demonstrates the directions and sizes of the relations between observed and latent variables. After setting a good fitting model with SEM, standardized regression weights, critical ratios (t-values) (C.R.=estimates ÷ standard errors [S.E.]) and p-values showing the significance of the paths are presented in *Table 29*.

Table 29

Direct Effects of Final SEM with Standardized Weights, t-values and Significance

Relation		Standardized Estimates	t-values	p-value
Programme	→PoTCS			
Programme	→LtoL	.281	5.675	***
Programme	→GC	.239	4.773	***
Programme	→PSS	-.190	-2.533	.011*
Programme	→English Comprehension	.332	6.032	***
English Comprehension	→PoTCS	.386	5.229	***
Age	→PSS	.096	1.367	.171
English Comprehension	→PSS	.145	1.755	.079
PoTCS	→PSS	.155	1.700	.089

** p<.05 **<.01 ***<.001 (*Programme→PoTCS direct effect is deleted during model modification)

Findings about the Examination of the Direct Effects of the Model

The third research question has five parts questioning the relation between programme type which students attend and other variables. Three of these questions are about the direct effects of programme type on a) *perceptions of 21st century skills*, b) *problem-solving skills*, and c) *English comprehension.* Additionally, fourth research question has five parts questioning the relation between students' problem-solving skills and other variables. Three of these questions are about the direct effects of students' *a) age*, *b) English comprehension, and c) perceptions of 21st century skills on* students' problem-solving skills. Sub-research questions which question the direct effects were assessed under the current title and the mediation and moderation effects are presented under separate titles.

As presented in *Table 29*, the direct relation of programme type (IB vs non-IB) to perception of 21st century skills (PoTCS) latent variable was omitted based on the insignificant relation after adding two extra paths in the model (-.013, p=.860) and the modification suggestions. So, final model indicated that programme type has significant effects on *learning to learn* [LtoL] (standardized coefficient= .281, p<.001) and *being a responsible global citizen* [GC] (standardized coefficient= .239, p<.001). Finding about the effect of programme type on LtoL means IB curriculum has a positive effect on students' perceptions about their skill of learning to learn. In other words, IB students have a more positive view about their ability to control and direct their own learning. Additionally, finding about the effect of programme type on GC means IB curriculum has a positive effect on students' perceptions about their skill of being a responsible global citizen. In other words, IB students have a more positive view about their awareness on their responsibilities for global matters. *[Research question 3a]* These findings came out from the established model are consistent with the Mann-Whitney U analysis findings of the first research question.

According to the relation between the programme type and problem-solving skills (PSS), there is a significant difference between students' problem-solving skills in terms of programme in the favor of non-IB students (standardized coefficient= -.190, p<.05). In other words, for the current study group, IB programme did not affect the problem-solving skills of the students positively. This finding is consistent with the Mann Whitney U test result of the second research question. The selection process of students to the related schools can be the reason of students from non-IB schools to achieve higher in PSS test independent from the programme type. Because non-IB students, who continue to a tto (tweetalig onderwijs) school for the current study, are selected students who study in homogenous groups (havo and vwo level [for more information see Dutch government page), wheras IB schools accept their students from many different backgrounds and ability levels without a selection process. Yet, this finding indicates that IB programme did not increase students' problem-solving skills for the current study. *[Research question 3b]*

On the other hand, programme type is a significant predictor of students' English comprehension in the favor of IB students (standardized coefficient=.332,

p<.001). This was an expected finding, because IB schools generally preferred by students who need to learn multiple languages or who have an international background. Of course, knowing multiple languages or coming from an international background do not necessarily guarantee IB students to have a good level of English comprehension compared to other bilingual schools. On the other hand, IBO (2022a) mentioned language as a central key to their curriculum and all IB schools in the Netherlands offer English in their language courses. Therefore, it is reasonable to see a significant English comprehension level difference in the favor of students who follow an IB programme. *[Research question 3c]*

The effect from age to problem-solving skills was not significant (standardized coefficient=.096, p=.171). It was expected that students' problem-solving skills would get better, when students get older, whereas contrary to the expectation students' problem-solving skills were not significantly higher for the older students. *[Research question 4a]* Additionally, the direct effect of English comprehension on students' problem-solving skills was not significant at p<.05 level, (standardized *[Research question 4b]*. Finding about the relation between PoTCS and PSS was similar. The direct effect of PoTCS on PSS was not significant at p<.05 level, (standardized coefficient=.155, p=.089). *[Research question 4c]*

Examination of English Comprehension as a Mediator Between Programme Type, PoTCS and PSS

The last two sub-question of the third research-question are about the mediation effect of English comprehension. Research questions are d) '*Is there a significant relation between the programme type students attend and their perceptions of 21st century skills with the mediation effect of English comprehension?' and e) 'Is there a significant relation between the programme type students attend and their problem-solving skills with the mediation effect of English comprehension?'*. *Table 30* shows the standardized regression weights, standardized errors (S.E.), C.R. (t-value) and significance levels (p-value) of these indirect effects together with the direct effects which form these indirect effects.

Table 30

Mediation Effects of English Comprehension on Final SEM

Relation	Standardized Estimates	S.E.	t-value	p-value
Programme type→English Comprehension → PoTCS	.092	.023		.001**
Programme type→ English Comprehension	.332	.083	6.032	***
English Comprehension → PoTCS	.386	.035	5.229	***
Programme type→ English Comprehension → PSS	.067	.044		.078
Programme type→ English Comprehension	.332	.083	6.032	***
English Comprehension → PSS	.145	.076	1.755	.079

** p<.05 **<.01 ***<.001*

As considering the direct effects, findings showed that IB students' English comprehensions were significantly better than non-IB students (standardized coefficient=.332, p<.001) and having a better English comprehension predicts higher perceptions of 21st century skills (PoTCS) (standardized coefficient=.386, p<.001). In addition to these findings, the relation between programme type and PoTCS was mediated by students' English comprehension levels significantly (standardized coefficient for indirect effect=.092, p<.01). In another words, IB students who have better English comprehension had better perception about their 21st century skills. *[Research question 3d]*

On the other hand, the relation between programme type and problem-solving skills was not mediated by students' English comprehension (standardized coefficient for indirect effect=.067, p=.078). In another words, when IB students' English comprehension become better, this did not increase their problem-solving skills significantly or having low English comprehension did not decrease non-IB students' PSS test scores *[Research question 3e]*

Examination of the Moderation Effects of Mathematics and English Achievement on PoTCS and PSS Relation

The last two parts of the fourth research-question are about the moderation effects of Mathematics (Math) and English achievements. Research questions are d) '*Is there a significant relation between students' problem-solving skills and their perceptions of 21st century skills with the moderation effect of Mathematics achievement?'* and e) *'Is there a significant relation between students' problem-solving skills and their perceptions of 21st century skills with the moderation effect of English achievement?'*. In order to check the moderation effects, SEM analysis were run two more times with interaction effects. Firstly, model was tested with the *Mathscore* variable and its standardized interactions with the independent variable of the relation (PoTCS), and then with *Englishscore* variable and its standardized interactions with the independent variable of the relation (PoTCS). Path diagrams of these analyses with Mathscore and Englishscore can be seen in *Figure 14* and *Figure 15* respectively.

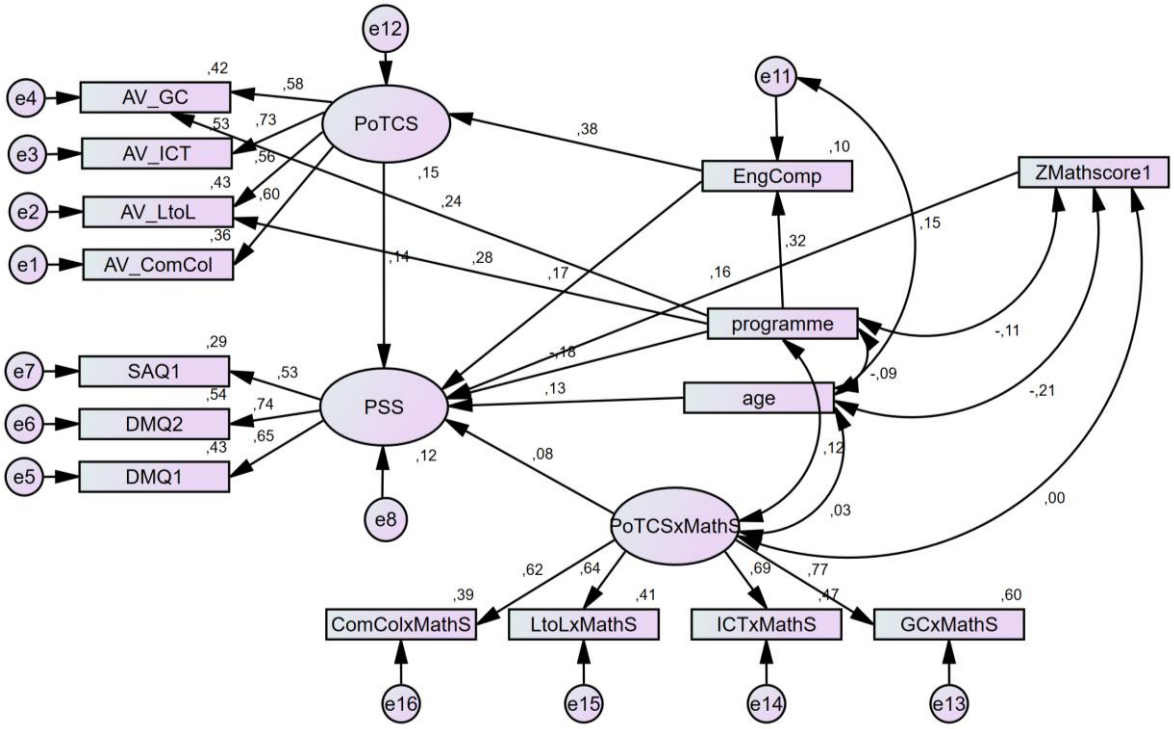


Figure 14. Final Model with the moderation effect of Math achievement

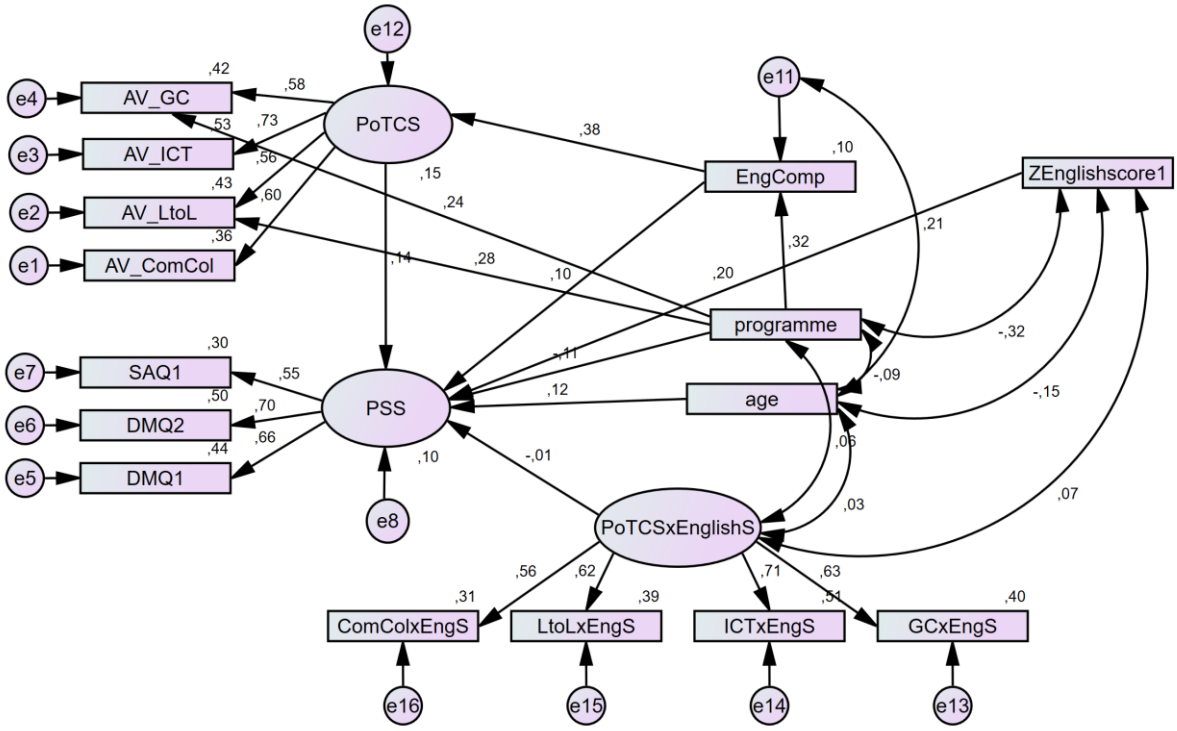


Figure 15. Final Model with the moderation effect of English achievement

Before checking the standardized regression coefficients and their significance, for both versions of the model, fit indices were assessed. For the model as Math achievement as the moderator, fit indices were x^2(77, N = 287) = 111.636, p=.002, x^2/df=1.450, CFI=.958, *PGFI=.610*, RMSEA=.040. For the model as English achievement as the moderator, fit indices were x^2(77, N = 287) = 99.763, p=.042, x^2/df=1.296, CFI=.970, *PGFI=.613*, RMSEA=.032. It was concluded that the models had still good fit after adding the moderation effects, and although the variance (r^2) of PSS (dependent variable of the relation) was not brilliant (r^2 =.12 in the model with MathScore moderation, r^2 =.10 in the model with EnglishScore moderation), it was accepted as substantial enough, and moderation effects were checked.

Table 31 shows the standardized regression weights, standardized errors (S.E.), C.R. (t-value) and significance levels (p-value) of moderation effects of Math and English achievement on PSS.

Table 31

Moderation Effects of Math and English Achievements on PSS on Final SEM

Relation	Standardized Estimates	S.E.	t-values	p-value
PoTCS*Math Achievement → PSS	.082	.070	1.062	.288
Math Achievement → PSS	.158	.050	2.218	.027*
PoTCS*English Achievement → PSS	-.012	.099	-.144	.886
English Achievement → PSS	.198	.054	2.598	.009**

** p<.05 **<.01 ***<.001*

According to the parameters estimated for the final model, effect of PoTCS on PSS was positive, but insignificant for p<.05 (standardized coefficient =.155, p=.089). As adding Math achievement and English achievement as moderators, it was aimed to check if Math achievement or English achievement would strengthen the positive effect of PoTCS on PSS. According to the findings displayed in *Table 31*, direct effect of Math achievement on PSS (standardized coefficient =.158, p=.027<.05) and direct effect of English achievement on PSS (standardized coefficient =.198, p=.009<.01) are significant. But, contrary to expectation, neither Math achievement (standardized coefficient=.082, p=.288) nor English achievement (standardized coefficient =-.012, p=.886) strengthen the positive effects of PoTCS on PSS significantly. *[Research questions 4d & 4e]*

Differences in the Effects of Variables Affecting PoTCS and PSS by Gender

The fifth research question is '*Do the effects of variables affecting students' perceptions of 21st century skills differ according to gender?*' and the sixth research question is '*Do the effects of variables affecting students' problem-solving skills differ according to gender?*'. These two research questions were assessed together. Among the effects, the followings were examined for differences by gender: 1) whole model, 2) programme type to PoTCS (LtoL and GC), 3) programme type to PSS, 4) English comprehension to PoTCS, 5) age to PSS, 6) English comprehension to PSS, 7) PoTCS to PSS. Chi-square difference tests results by gender related to each effect are given in *Table 32*.

Table 32

Chi-square Difference Tests Results by Gender about the Effects in Final SEM

Relation		x^2	df	p-value
Final SEM (whole model)		7.325	13	.885
Programme	→PoTCS			
Programme	→LtoL	3.331	1	.068
Programme	→GC	.017	1	.898
Programme	→PSS	.049	1	.824
English Comprehension	→PoTCS	.842	1	.359
Age	→PSS	1.109	1	.292
English Comprehension	→PSS	.782	1	.377
PoTCS	→PSS	.279	1	.597

** p<.05 **<.01 ***<.001*

First of all, final model was examined as a whole and model did not show a significant chi-square difference between male and female groups, $x^2\left(13, N_{male} = 137\ N_{female} = 150\right) = 7.325, \mathrm{p} = .885$, which meant there was not a significant difference between the effects in the total model by gender. Chi-square difference test was repeated for each direct effect in the model to see if there was any difference between the effects in terms of gender. With each test two models (one for male and one for female) were freely estimated except constraining one path as equal across groups. Chi-square difference tests were insignificant for each direct effect in the model (p<.05). This result exposed that the effects in the final SEM related to PoTCS or PSS did not change by gender.

Effect of Duration of Enrollment in the IB Programme on PoTCS

The last research question is '*Are there significant differences between IB students' perception of 21st century skills according to the number of years that students attend to the IB programme?'*. Kruskal-Wallis test was conducted to examine the differences on students' perceptions of *being a responsible global citizen* (GC), *using ICT effectively* (ICT), and Learning to Learn (LtoL) skills according to the time of enrollment in the IB programme. Additionally, one-way ANOVA was conducted to examine the differences in students' perceptions of

communication and collaboration (ComCol) skills according to the time of enrollment in the IB programme. The descriptive statistics of PoTCS factors based on IB duration are given in *Table 33*.

Table 33

Descriptive Statistics of PoTCS Scores According to IB Programme Duration

Factor	Duration *	N	$\bar{X}$	s.d.
Being a responsible global citizen (GC)	1	45	4.07	.516
	2	51	4.01	.582
	3	44	3.96	.641
Using ICT effectively (ICT)	1	45	3.93	.600
	2	51	3.99	.616
	3	44	4.23	.567
Learning to Learn (LtoL)	1	45	3.64	.836
	2	51	3.90	.716
	3	44	3.86	.741
Communication & Collaboration (ComCol)	1	45	3.39	.627
	2	51	3.56	.681
	3	44	3.60	.689

* ***1:*** *1 year or less,* ***2:*** *2 to 3 years,* ***3:*** *4 years more*

According to the average point of each duration period (1: 1 year or less, 2: 2 to 3 years, 3: 4 years or more) students' perceptions of using ICT effectively, learning to learn and communication and collaboration skills generally increase when duration of the enrollment in the IB system increases. On the other hand, being a responsible global citizen has a slight decrease when the duration of enrollment increases. Kruskal-Wallis Test was conducted to examine the differences on perceptions of being a responsible global citizen, learning to learn and using information and technology effectively skills according to the time of enrollment in the IB programme. There was no significant difference found for being a responsible global citizen (x^2= .381, p = .826, df = 2) and learning to learn (x^2= 2.933, p = .231, df = 2). On the other hand, the differences on perceptions of

using information and technology effectively skill had a significant difference according to the time of enrollment in the IB programme (x^2= .048, p= .048<.05, df = 2). Groups were compared two by two (1-2, 2-3, 1-3) with Mann-Whitney U test in order to spot the differences between groups. According to the results, only students who enrolled into the IB programme for long time (four years or more: 3) had a higher perception about their effective use of ICT compare to the group who enrolled into the IB programme for short time (one year or less: 1).

Additionally, one-way ANOVA was performed to examine the differences in the perceptions of communication and collaboration skills according to the time of enrollment in the IB programme and there were no significant differences found (F (2,137) = 1.244, p>.05). Kruskal-Wallis and one-way ANOVA results for total 140 IB students are stated in *Table 34*.

Table 34

One-way ANOVA and Kruskal-Wallis Test Results of PoTCS Scores According to IB Programme Duration

	Duration	n	Rank Averages	df	x^2	p-value	Significant Differences
GC	1	45	73.26	2	.381	.826	-
	2	51	70.24				
	3	44	67.99				
ICT	1	45	62.77	2	6.066	.048*	3-1
	2	51	66.86				
	3	44	82.63				
LtoL	1	45	62.17	2	2.933	.231	-
	2	51	75.48				
	3	44	73.25				

	Duration	Sum of Squares	df	Mean Square	F	p-value	Significant Differences
ComCol	1	1.105	2	.552	1.244	.291	-
	2	60.836	137	.444			
	3	61.941	139				

** p<.05 **<.01 ***<.001*

Chapter 5
Conclusion, Discussion and Suggestions

In this chapter, findings of the research are summarized and discussed in accordance with the related literature in order of the research questions. Suggestions are presented for the practitioners and policy makers as well as for the researchers.

Conclusion and Discussion

The focus of the current research is to identify the effect of programme type on students' perceptions of their 21st century skills (PoTCS) and problem-solving skills (PSS), as well as to examine the relationships between these variables and English comprehension, age, Mathematics and English achievement and gender. Additionally, it is aimed to conceive if the duration of the IB enrollment makes a difference in students' PoTCS. It is aimed to reveal if IB middle years programme and its joint assessment processes help students to develop a more positive image about their 21st century skills and if they improve students' problem-solving skills better than other schools which follow a different programme. In order to measure the perceptions of 21st century skills, a scale with situational judgment items were created based on the aim of assessing the related characteristics more validly and reliably across different cultures and different reaction styles of respondents (Herde, et al., 2019; Yalçın, 2018). Conclusions and related discussions are presented with following points.

- Students who follow the IB programme perceived themselves better in their 21st century skills.

All students who attended to the current study showed a certain level of perceptions of 21st century skills regardless of programme type, but students who follow the IB programme had a higher level of perception for each dimension of the PoTCS scale compared to non-IB students. This finding is parallel with Dulun (2018)'s research findings, which stated that students who came from IBMYP had higher research skills, critical thinking skills and communication skills compared to students who came from two other programmes. Additionally, Öztermiyeci's (2019) findings supported this difference that IBDP students scored higher for all domains of perceptions of 21st century skills.

- The differences in perception of 21st century skills were significant in being a responsible global citizen and learning to learn dimensions of PoTCS scale in favor of IB students.

The differences in PoTCS scale scores appeared in favor of IB students on two dimensions with medium effect size for the current study. It is valuable to see the significant differences occurred for the skill sets which are comparingly newer skills (global citizenship and learning to learn) which were not studied as much as the other two skill sets (using information and technology effectively and communication and collaboration) as seen in a systematic literature review study (Van Laar, Van Deursen, Van Dijk & Haan, 2020).

Consistent with the findings of Mann Whitney-U test, constructed model showed significant effect of programme type on being a responsible global citizen and learning to learn factors of 21st century skills in favor of IB students. These findings showed that IB curriculum achieved to give their students (who attended to the current study) a more positive perspective about themselves on global citizenship and learning to learn, which match with IB's mission statement (IBO, 2017a). This outcome was not surprising as supporting the core IB values. Additionally, according to World schools' (2022) article, experts from IB schools listed international-mindedness and self-directed learning as two of the top advantages of IB schools.

On the other hand, the results can also stem from the structure of the questions. Situational Judgment test items are formed with short stories and these contextual items could be understood and answered differently by different group of people (Herde et al., 2019). Because of the international background of IB students, they could have a certain way of approaching to the given short scenarios.

Educators in the Netherlands in all levels of schools have also given attention into the concept of "global citizenship", whereas a common description and so a common targeted goal has not been set so far (Duarte, 2021). Hogeling (2012) reported that a structural attention for the concept of global citizenship is lagging behind in Dutch schools, which explains the difference between students' perceptions about global citizenship in terms of IB and non-IB students.

Additionally, secondary school teachers in Dutch education had the issue of lack of time to put attention on global citizenship in their lessons, and primary teachers emphasize the complexity of the global citizenship themes (Hogeling, 2012).

According to Mihelich's (2003) description, youth is concerned about environmental issues and have a high sense of responsibility. Similar projects offered by IB curriculum (SA or CAS) could be integrated for other school curricula in order to increase students' self-image about their global awareness as well. According to Perry, Ledger & Dickson's (2018) comprehensive study, conducted in Australia, participants who are MYP coordinators, teachers and principles', MYP is highly effective for students' non-scholastic in addition to their academic development. According to the same study, participants support the view that MYP encourages independent learning and valued local and global citizenship (Perry, Ledger & Dickson, 2018).

Learning to learn, or meta-learning, is listed as one of the skills which is not sufficient for the current and future work force on the World (Horvathova, 2019). This brings us as educators the need to find ways to improve students' learning to learn skills. According to the findings of the current study, despite the demanding and challenging nature of IB programme, it seems IB programme and integrated assessments around it achieving to give more autonomy to their students in terms of their learning and educate more aware and responsible learners. This finding could be interpreted if there was a data about the students' PoTCS levels before they started to follow the IB programme.

- Students who follow the non-IB programme scored better in problem-solving skills test.

According to the findings, problem-solving skills of students who follow the IB MYP curriculum and assessment systems could not achieve high in the problem-solving skills test for the current study. Students who follow a curriculum other than IB could achieve higher than IB students not only for the total of the PSS test but also for each question. The difference was significant only for one of the questions with a small effect size. This finding could stem from the schools' student acceptance policies and non-IB students being a homogenous and higher achiever group. Yet, if IB schools strive to improve their students in terms of 21st

century skills including problem-solving skills regardless of students' backgrounds, according to the current study findings this aim has not been fully achieved yet. This finding is parallel with Swartz and McGuinness' (2014) report which suggest curriculum team of IB MYP to extend the current taxonomy that IB MYP curriculum uses to involve problem-solving and decision-making.

Similar with the findings of Mann Whitney-U test, constructed model showed significant effect of programme type on problem-solving skills of students in favor of non-IB students. This finding can be based on problem-solving skills test used in the current study being closer to an academic test and tto (non-IB programme) students who attend to the current study being selected and having high academic profile. On the other hand, the test was improved in the frame of PISA problem-solving skills and IB programme might still need to improve on developing complex problem-solving skills.

Problem-solving skills, especially complex, unfamiliar and ill-structured ones need to be introduced and improved in all age groups of today's students. As Coombs (2013) mentioned, "Technology is a part of their [new generations'] identity and they are tech savvy but lack of problem-solving skills and have not demonstrated the ability to look at a situation, put in context, analyze it and make a decision". Similar with the previous finding, knowing students' levels of problem-solving skills before they started to follow IB or non-IB programme could give more insight for further interpretations.

- Students who follow the IB programme had better English comprehension.

Programme had a significant effect on English comprehension in favor of IB students. In another words, students who follow the IB programme chose their English comprehension levels as 'very good' or 'mother tongue' 'more instead of 'not so good' or 'good' compared to non-IB students. This was an expected effect because of the background of IB students and acceptance process of IB schools. IB schools are generally promoted for students who have an international background or possibility to live abroad in the future which bring the necessity of a good English comprehension level. This brings the discussion if IB programme effect students' English comprehension positively or students who have better English comprehension choses IB schools. On the other hand, it should be

mentioned that IB schools claimed to be dedicated to provide a good language education (IBO, 2022a), so this significant effect can stem from the programmes IB follows as well. Furthermore, International schools in the Netherlands use English as common communication tool and medium for teaching and learning (Dutch International Schools Annual Report 2020).

- IB students' who have higher English comprehension had higher perceptions of 21st century skills.

This finding is related with the mediation effect of English comprehension between programme type and students PoTCS scale scores. According to findings English comprehension appeared as a significant mediator in between programme type and PoTCS. In another words, IB programme effect students' English comprehension positively and students who had better English comprehension had a higher perception of 21st century skills. It is surprising that there was no study could be found investigating the effect of knowing English on improving 21st century skills. English is the most commonly spoken language on the world (Szmigiera, 2021) and most used language in the internet (Johnson, 2021), so it seems reasonable that better English comprehension will naturally mean more people to communicate and collaborate, more sources to be reached, more different perspectives to gain etc. Hence, better English comprehension would mean for a person to improve on expected skills of the new century better.

- IB or non-IB students who have better English comprehension did not have better scores in problem-solving skills test.

English comprehension did not have a significant effect on students' problem-solving skills, which is contradicting with study findings of Al-Bado (2021), which reveals that problem-solving skills can be improved with a reading comprehension strategy. Another contradicting study finding showed that English language skills predicted performance in mathematical problem solving (Beal, Adams & Cohen, 2010). Additionally, the mediation effect of English comprehension between program type and problem-solving skills was also appeared as insignificant. Understanding the questions in an English problem-solving skills test is indeed a requirement, first step, in order to solve a problem,

whereas it is not the only requirement. According to the findings, better English comprehension did not affect problem-solving skills of non-IB student positively.

- Age did not affect students' problem-solving skills significantly.

Age did not have a significant effect on problem solving skills, which was against the expectation based on, "Problem-solving skills peak around the age of 30 and decline thereafter" (Horvathova, 2019, p.30). On the other hand, according to Van Laar et al.'s (2020) review demographic determinants like age appeared on research results as nonsignificant more than significant, which is consistent with the current study finding.

- Students' perceptions of 21st century skills did not have a significant effect on their problem-solving skills.

For the current study PoTCS scale aimed to measure students' perceptions about their 21st century skills, whereas problem-solving skills test aimed to measure students' problem-solving skills which is considered as one of the 21st century skills. It is assumed that when students' perceptions about themselves on a matter is better, for instance high math self-efficacy predicts high math test performance (Schulz, 2005), they could perform better on that matter. For instance, according to Lee and Stankov's (2018) study as using large-scale international assessment data, students' perceptions about their abilities appear as very important predictors for their future achievements, whereas there could not be such an effect found with the current study findings.

- Students' Mathematics or English achievement (report grades) did not strengthen the effect of their perceptions of 21st century skills on their problem-solving skills.

Problem-solving is one of the 21st century skills and it was assessed by PSS test for the current study. With the model constructed in SEM, variables which affect or predict this skill was tried to be revealed. One of the possible variables which can affect PSS positively was students' perceptions about their 21st century skills, whereas the effect was not statistically significant. Therefore, as the next step, variables which can strengthen the effect of PoTCS on PSS were evaluated. Achievement in Mathematics was added as the moderator into the model as expecting it would strengthen the effect of PoTCS on PSS, because

students' perceptions about their 21st century skills were high and their achievement in a subject related to problem-solving skills were high as well, it was expected to get high level of problem-solving skills, but it did not happen. On the other hand, achievement in English was added as the moderator into the model as well, based on similar expectations with Mathematics achievement, but English achievement did not strengthen the effect either.

- The effects of variables affecting students' perceptions of 21st century skills did not differ according to gender.

The effects of variables affecting PoTCS did not differ by gender. This finding was contradicting with Van Laar et al.'s (2020) findings of a systematic literature review study. According to Van Laar et al.'s (2020) review for 21st century skills demographic factors such as gender appeared as significant more than nonsignificant in the literature. In another word, literature showed more studies which found significant effect of gender on 21st century skills than insignificant ones.

- The effects of variables affecting students' problem-solving skills did not differ according to gender.

The effects of variables affecting PSS did not differ by gender as well. This finding was parallel with Van Laar et al.'s (2020) findings of a systematic literature review study. According to Van Laar et al.'s (2020) review for problem-solving skills demographic factors such as gender appeared as nonsignificant more than significant in the literature.

- The longer the IB students' had attended to an IB school the better they perceived themselves in effective information and technology use.

According to the findings, only using information and technology effectively dimension of PoTCS scale had a significant difference in terms of the duration of enrollment into IB programme in favor of longer enrollment. In another words, the longer students had followed the IB MYP, the better they perceive themselves in terms of their effective use of information and technology. Nowadays, using personal laptops are normal at higher education, which started to become widespread in early 2000s. For secondary schools, bringing personal laptops to school has not become widespread yet in most of the countries with different

reasons, such as security issues, internet access problems or improper use of devices (Blair & Briggs, 2016).

IB schools in the Netherlands use an internet platform as the common way of communication between students and teachers, for assigning work, giving feedback, sharing documents, announcing any lesson related news etc. Additionally, investigation-based curricula direct students and teachers to use search engines and create investigation reports as well, which forms the need of personal device use. Therefore, bringing laptop seems to become more normal than bringing a textbook to the class for an IB student in the Netherlands. This could be one of the reasons, the longer a student has been enrolled into the IB system, the better they perceive themselves in terms of effective information and technology use. Additionally, IB school students have to pay a certain fee for attending to these schools in the Netherlands, whereas students who follow the other curricula do not have an enrolment fee. So, it is expected that students who attend to an IB school come from a higher income family and so technological devices can be more accessible by those students.

Suggestions

The suggestions for practitioners and researchers resulting from the current study findings are presented separately in this section.

Suggestions for practitioners. Suggestions for school teachers, managers, coordinators and policy makers about the effect of programme type and assessment practices on 21st century skills are proposed below. Mostly, it is easy to plan and set objectives in education systems or schools, but way harder to actually implement these objectives. This matter is a limitation for any suggestion would be presented.

- To advance students' 21st century skills in schools, clear operational definitions for 21st century skills for the use of secondary school teachers and managers may be created. IB students have a more positive perspective about their 21st century skills, because these skills are clearly stated in IB core values as the IB learner profile. This clarity might help for all schools to develop these skills.

- Providing professional development for teachers and coordinators to give a clear picture about the definitions of 21st century skills and to show good practices about how to integrate and efficiently implement 21st century skills practices in various curricula and assessment practices may help to improve a better understanding for professionals and implementation in classroom settings.
- Revising current curricula in terms of 21st century skills in country level and adapting the features from IB schools' curricula such as community projects may help improving 21st century skills.
- Revising current curricula, where necessary to include research-based aspects of successful applications of 21st century skills education and assessment practices may help improving 21st century skills. Especially, it may help to education parties to integrate *formative assessment*s more into all teaching learning processes as "assessment for learning" and use only *summative assessments* as "assessment of learning" (Harlen & Johnson, 2014).).

Suggestions for researchers. Suggestions for researchers about investigating how to integrate, develop and assess 21st century skills, problem-solving skills, and programme type in the current education systems are presented below.

- Conducting further research on different types of curricula other than IB to investigate whether there are possibilities to collaborate in between different curricula and assessment practices may help to utilize in implementing 21st century skills in secondary schools.
- Using item response theory to assess item and test statistics to develop a 21st century skills scale instead of classical test theory might bring more detailed results.
- Instead of single response situational judgement tests, multiple response situational judgement tests may be created to assess the multi-faceted features of 21st century skills better.

- Considering the effect of IB MYP criterion-based assessment with other assessments based on supporting 21st century skills in an English-speaking country could bring a more representative sampling and more generalizability to the findings related to the matter.
- Combining different forms of instruments to assess 21st century skills may give a deeper understanding of the matter and give direction about which kind of measurement instruments can be used. Examples for different forms of assessment can be listed as: structured interviews to assess communication skills and global citizenship; critical thinking skills tests; assessing how students analyze a text to summarize the necessary information; assessment of web searching skills to find relative information about a topic.
- Assessing problem-solving skills prior to start following an IB programme and after a certain time, and doing the same for other programmes and comparing the improvements in problem-solving skills could give a better picture about the contribution of the programme on problem-solving skills.

References

Akçay, A. (2019). *An analysis of ELT coursebooks in terms of 21st century skills: communication, collaboration, critical thinking, creativity* [Master's thesis]. Akdeniz University Institute of Educational Sciences, Antalya.

Al-Bado, A.M.A. (2021). The effect of using reasing comprehension strategy to acquire problem-solving skills in mathematics among tenth-grade students in Amman. *Journal of Social Sciences*, 21.

Allen, J., & van der Velden, R. (2013). Skills for the 21st century implications for Dutch education. In Larry Robert Smit (ed.) *Higher Education Recent Trends, Emerging Issues and Future Outlook (pp.1-40)*. Nova Publishers, New York.

Alkış, M. (2020). *Üniversite öğrencilerinin 21. Yüzyıl becerilerinin bazı değişkenler açısından incelenmesi* [Master's thesis]. Necmettin Erbakan Üniversitesi Eğitim Bilimleri Enstitüsü, Konya.

Altun, B. (2021). *21.yüzyıl becerilerinin Z kuşağı üniversite öğrencilerinin kariyer planlamasına ve teknoloji yönelimlerine etkisi* [Master's thesis]. İstanbul Ticaret Üniversitesi Sosyal Bilimler Enstitüsü.

Anagün, Ş.S., Atalay, N., Kılıç, Z. & Yaşar, S. (2016). Öğretmen adaylarına yönelik 21. Yüzyıl becerileri ölçeğinin geliştirilmesi: geçerlik ve güvenirlik çalışması. *PAU Eğitim Fakültesi Dergisi*, 40, 160-175.

Ananiadou, K. & Claro, M. (2009). 21st century skills and competences for new millennium learners in OECD countries. OECD Education Working Papers, No. 41, *OECD Publishing*. Retrieved from http://dx.doi.org/10.1787/218525261154

Assessment & teaching of 21st century skills (ATC21S). (2009-2012). Retrieved from http://www.atc21s.org/about.html

Assessment & teaching of 21st century skills (ATC21S). (2010). Retrieved from https://www.cisco.com/c/dam/en_us/about/citizenship/socio-economic/docs/ATC21S_Exec_Summary.pdf

Atalay, N. (2015). *Fen bilimleri dersinde öğrencilerin öğrenme ve yenilenme becerilerinin gelişiminde yavaş geçişli animasyon (slowmotion) uygulaması* [Doctoral dissertation]. Anadolu Üniversitesi Eğitim Bilimleri Enstitüsü, Eskişehir.

Atakişi, T. (2019). *Öğretmenlerin ve öğretmen adaylarının 21. Yüzyıl becerilerini esas alan bilgi okuryazarlık durumlarının çeşitli değişkenler açısından incelenmesi* [Master's thesis]. Kafkas Üniversitesi Sosyal Bilimleri Enstitüsü, Kars.

Aydın, A. (2019). *İngilizce öğretmen adaylarının görüşleri çerçevesinde öğretmen eğitiminde 21. yüzyıl becerilerinin incelenmesi* [Master's thesis]. Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.

Aydın, S. (2021). *Yabancı dil hazırlık sınıfı öğrencilerinin 21. yüzyıl becerileri, başarı odaklı motivasyonları ve ingilizce öğrenmeye yönelik motivasyonları arasındaki ilişkiler* [Master's thesis]. Bartın Üniversitesi Lisansüstü Eğitim Enstitüsü, Bartın.

Babakuş, E. & Mangold, W. G. (1992). Adapting the SERVQUAL scale to hospital services: an empirical investigation. *HSR: Health Services Research*, 26 (6), 767-786. https://www.researchgate.net/post/Should_we_use_a_5_or_7_point_Likert_scale_Whats_better_and_why

Baker, E. L. (2007). The end(s) of testing. *Educational Researcher*, 36(6), 309-317. Retrieved from http://er.aera.net

Bal İncebacak, B. & Ersoy, E. (2016). Problem solving skills if secondary school students. *China-USA Business Review*, 15, 6, 275-285.

Bayülgen, N. (2012). *Comparison of cognitive and affective skills of students which take international baccalaureate program Turkish A1 course and students which take national program Turkish literature course*. [Doctoral dissertation]. Uludağ University, Bursa.

Beal, C.R., Adams, N.M., Cohen, P.R. (2010). Reading Proficiency and Mathematics Problem Solving by High School English Language Learners. *Urban Education*. 45(1), 58-74. doi:10.1177/0042085909352143

Benek, İ. (2019). *Sosyobilimsel stem etkinliklerinin öğrencilerin tutumlarına ve 21. Yüzyıl becerilerine etkisinin incelenmesi* [Doctoral dissertation]. İstanbul Üniversitesi Cerrahpaşa Lisansüstü Eğitim Enstitüsü, İstanbul.

Bennet, R. E., Jenkins, F., Persky, H. & Weiss, A. (2003). Assessing complex problem solving performances. *Assessment in Education*, 10 (3), 347-359.

Bentler, P. M. (2006). *EQS 6 structural equations program manual*. Encino, CA: Multivariate Software.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588–606. https://doi.org/10.1037/0033-2909.88.3.588

Bess, T. L. (2001). *Exploring The Dimensionality of Situational Judgment: Task and Contextual Knowledge*. (Master's thesis). The Faculty of Virginia Polytechnic Institute and State University, Virginia.

Binkley, M., Erstad, O., Hermna, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining Twenty-First Century Skills. In Griffin, P., Care, E., & McGaw, B. (eds.) *Assessment and Teaching of 21st Century Skills (pp.17-66)*. Springer, Dordrecht.

Bircan, M.A. (2019). *Stem etkinliklerinin ilokul dördüncü sınıf öğrencilerinin stem'e yönelik tutumlarına, 21. Yüzyıl becerilerine ve matematik başarılarına etkisi* [Doctoral dissertation]. Ondokuz Mayıs Üniversitesi Eğitim Bilimleri Enstitüsü, Samsun.

Briggs, G., & Blair, E. (2014). Everyday personal laptop usage in secondary schools in Trinidad and Tobago. *Education and Information Technologies*, 21(3), 545–558. https://doi.org/10.1007/S10639-014-9338-4

Bunnell, T. Fertig, M. & James, C. (2016). What is international about International Schools? An institutional legitimacy perspective. *Oxford Review of Education.* DOI:10.1080/03054985.2016.1195735

Büyüköztürk, Ş. (2012). *Sosyal Bilimler için Veri Analizi El Kitabı* (17. Baskı). Ankara: Pegem Akademi Yayıncılık.

Byrne, B.M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed). New York: Taylor & Francis.

Cain, M.K., Zhang, Z. & Yuan, KH. (2017). Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence and estimation. *Behavioral Research*, 49**,** 1716–1735. https://doi.org/10.3758/s13428-016-0814-1

Care, E., Griffin, P., Woods, K. & Mountain, R. (2009-2012). Module 2: Defining and assessing 21st century skills. Assessment Research Centre, Melbourne Graduate School of Education.

Care, E., Kim, H. & Scoular, C. (2017). 21st century skills in 20th century classrooms. *Ducadores* (December).

Cohen, R. J. & Swerdlik, M. E. (2010). *Psychological testing and assessment: An introduction to tests and measurement* (7th ed.). Mountain View, Calif: Mayfield Pub. Co.

Confirmatory Factor Analysis (CFA) in R with Lavaan (n.d.). UCLA: Statistical Consulting Group. Retrieved December 27, 2021, from https://stats.oarc.ucla.edu/r/seminars/rcfa/#s4e.

Conley, D, McGaughy, C, Davis-Molin, W, Farkas, R and Fukuda, E. (2014). *International Baccalaureate Diploma Programme: Examining college readiness.* Bethesda, MD: USA. International Baccalaureate Organization

Coombs, J. (2013). Generation Z: Why HR Must Be Prepared for Its Arrival. Retrieved from https://www.shrm.org/resourcesandtools/hr-topics/talent-acquisition/pages/prepare-for-generation-z.aspx

Cortina, J.M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, (78), 1, 98-104.

Crocker, L. & Algina, J. (2008). *Introduction to classical & modern test theory* (2nd Edition). Ohio: Cengage Learning.

Çevik, M. & Şentürk, C. (2019). Multidimensional 21st century skills scale: validity and reliability study. *Cypriot Journal of Educational Sciences*, 14(1), 11-28.

Çınar, F.S. (2019). *Ortaokul öğretmenlerinin 21. Yüzyıl becerilerine ilişkin algılarının ve görüşlerinin incelenmesi (Çorum ili örneği)* (Master's thesis). Kafkas Üniversitesi Sosyal Bilimler Enstitüsü, Kars.

Çokluk, Ö., Şekercioğlu, G. & Büyüköztürk, Ş. (2012). Sosyal bilimler için çok değişkenli istatistik SPSS ve Lisrel Uygulamaları (2nd ed.). Pegem Akademi: Ankara.

DATAtab Team (2022). DATAtab: Online Statistics Calculator. DATAtab e.U. Graz, Austria. URL https://datatab.net

Dede, C. (2009). Comparing frameworks for "21st century skills". Harvard Graduate School of Education. Retrieved from http://sttechnology.pbworks.com/f/Dede_%282010%29_Comparing%20Frameworks%20for%2021st%20Century%20Skills.pdf

Demetriou, A. & Kazi, S. (2006). Self-awareness in g (with processing efficiency and reasoning). *Intelligence*, 34(3), 297-317.

Dinç Bilgin, S. (2021). *2D ve 3D destekli modellemeye dayalı öğretimin öğrencilerin akademik başarılarına ve 21. Yüzyıl becerilerine etkisi* (Master's thesis). Kütahya Dumlupınar Üniversitesi Eğitim Bilimleri Enstitüsü, Kütahya.

DiStefano, C., Zhu, M. & Mindrila, D. (2009). Understanding and using factor scores: considerations for the applied researcher. *Practical Assessment, Research & Evaluation*, 14(20). DOI: https://doi.org/10.7275/da8t-4g52

Doğan, E. (2020). *Ortaöğretim programlarına 21. Yüzyıl becerilerinin yansıması ve öğretmenlerin bu becerilere ilişkin görüşleri (Ankara ili örneği)*. Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.

Doorman, M., Drijvers, P. Dekker, T., van den Heuvel-Panhuizen, M., de Lange, J. & Wijers, M. (2007). Problem solving as a challenge for mathematics education in the Netherlands, *ZDM Mathematics Education*, 39, 405-418.

Duarte, J. (2021). "Global citizenship means different things to different people": Visions and implementation of global citizenship education in Dutch secondary education. *Prospects*. https://doi.org/10.1007/s11125-021-09595-1

Dulun, Ö. (2018). *Student perceptions of successful preparation for IBDP: implications for developing 21st century skills* [Doctoral dissertation]. İhsan Doğramacı Bilkent University, Ankara.

Dutch International Schools Annual Report 2020. Retrieved from https://riss.wolfert.nl/

Edens, K. M. (2000) Preparing problem solvers for the 21st century through problem-based learning, *College Teaching*, 48(2), 55-60, DOI: 10.1080/87567550009595813

Education GPS, *OECD*, 04/12/2021, 12:29:11 http://gpseducation.oecd.org

Education Research Center (2013). Developing and assessing students' collaboration in the IB programme. A&M University, Texas.

Erbek, İ. (2021). *Öğretmen adaylarının ve aday öğretmenlerin öğretmenlik mesleği genel yeterlikleri ile 21.yüzyıl becerileri öz yeterlikleri* (Master's thesis). Siirt Üniversitesi Sosyal Bilimler Enstitüsü, Siirt.

Erdoğan, Ö. (2019). *Robotik lego uygulamaların fen bilgisi öğretmen adaylarının 21. yüzyıl becerileri üzerindeki etkilerinin incelenmesi* (Master's thesis). Amasya Üniversitesi Fen Bilimleri Enstitüsü, Amasya.

Erkılıç, M. (2020). *21. Yüzyıl becerilerinin fizik başarısına etkisinin araştırılması. Necmettin Erbakan Üniversitesi* (Master's thesis). Necmettin Erbakan Üniversitesi Eğitim Bilimleri Enstitüsü, Konya.

Facione, P. A. (1990). *Critical thinking: a statement of expert consensus for purposes of educational assessment and instruction*. American Philosophical Association, ERIC document ED 315-423.

Facione, P. A. (2000). The disposition toward critical thinking: Its character, measurement, and relationship to critical skill. *Informal Logic*, 20(1), pp.61-84.

Fornell, C. & Larcker, D.F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39-50.

Frankel, J. R., Wallen, N. E. & Hyun, H. H. (2012). *How to design and evaluate research in education* (8th ed.). New York, NY: McGraw-Hill.

Fuller Jr, E. L., & Hemmerle, W. J. (1966) *Robustness of the maximum-likelihood estimation procedure in factor analysis.* Psychometrika, 31 (2), 255-266.

Gaskin, J. (2013), *SEM Series Part 3: Exploratory Factor Analysis*, Gaskination's Statistics. http://youtube.com/Gaskination

Gaskin, J. (2016), *SEM Series (2016) 7. Setting Up Causal Model*, Gaskination's Statistics. http://youtube.com/Gaskination

Gaskin, J. (2021), *Causal Models*, Gaskination's StatWiki. http://statwiki.gaskination.com

Gay, L.R., Mills, G. E. & Airasian, P. (2006). *Educational research: Competencies for analysis and applications (8th ed.).* Boston: Pearson Education, Inc.

Geban, Ö. (2001). *21. Yüzyıla girerken Türk eğitim sisteminin ihtiyaç duyduğu çağdaş öğretmen profili*. Ankara: Milli Eğitim Basımevi.

George, D. & Mallery, M. (2010). *SPSS for windows step by step: A simple guide and reference 17.0 update* (10the Edition). Boston: Pearson.

Griffin, P., McGaw, B., Care, E. (Eds.) (2012). *Assessment and teaching of 21st century skills*. New York, NY: Springer.

Günüç, S., Odabaşı, H. F. & Kuzu, A. (2013). The defining characteristics of student of the 21st century by student teachers: a twitter activity. *Journal of Theory and Practice in Education*, 9(4), 436-455.

Haladyna, T. M. (1997). *Writing test items to evaluate higher order thinking*. Needham Heights, MA: Allyn & Bacon.

Hamarat, E. (2019). *21. Yüzyıl Becerileri Odağında Türkiye'nin Eğitim Politikaları*. Istanbul: SETA Yayınları.

Harlen, W. & Johnson, S. (2014). A review of current thinking and practices in assessment in relation to the primary years programme. Assessment Europe. Retrieved from https://www.ibo.org/globalassets/publications/ib-research/assessmentinthepypfinalreport.pdf

Herde, C. N., Lievens, F., Solberg, E. G., Harbaugh, J. L., Strong, M. H. & Burkholder, G.J. (2019). Situational judgement tests as measures of 21st century skills: evidence across Europe and Latin America. *Journal of Work and Organizational Psychology,* 35(2), 65-74.

Hilbert, M. (2020). Digital technology and social change: The digital transformation of society from a historical perspective. *Dialogues in Clinical Neuroscience*, *22*(2), 189–194. https://doi.org/10.31887/DCNS.2020.22.2/mhilbert

Hogeling, L. (2012). *Global citizenship in primary and secondary education in the Netherlands: the opinions, attitudes and experiences of primary and secondary education teachers in relation to global citizenship* (Research series 6). Amsterdam: NCDO.

Hopfenbeck, T. N. (2018). Assessment and learning in an uncertain world. *Assessment in Education: Principles, Policy & Practice,* 25(4), 351-354.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. Psychometrika, 30(20), pp.179-185.

Horvathova, M. (2019). Study on employability skills in the IB diploma programme and career-related programme curricula. *Center for Curriculum Redesign*.

Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. https://doi.org/10.1080/10705519909540118

IB Africa, Europe & Middle East regional conference (2013, October) [PowerPoint presentation]. Retrieved from https://www.ibo.org/contentassets/019c8bebf2a84f2c8493ae56de3c8c34/100b-sevaggertrudegomescreativelyembracinglifelonglearning.pdf

IBO, International Baccalaureate Organization, (2014). The IB Continuum of international education. Retrieved from https://www.ibo.org/globalassets/digital-tookit/brochures/ib-continuum-brochure-en.pdf

IBO, International Baccalaureate Organisation. (2017). What is an IB education? Retrieved from http://www.ibo.org/globalassets/what-is-an-ib-education-2017-en.pdf

IBO, International Baccalaureate Organisation. (2018a). What is the DP? Retrieved from https://www.ibo.org/programmes/diploma-programme/what-is-the-dp/

IBO, International Baccalaureate Organisation. (2018b). What is the MYP? Retrieved from http://www.ibo.org/programmes/middle-years-programme/what-is-the-myp/

IBO, International Baccalaureate Organisation. (2019a). DP assessment and exams Retrieved from https://www.ibo.org/programmes/diploma-programme/assessment-and-exams/

IBO, International Baccalaureate Organisation. (2019b). Find an IB school. Retrieved from http://www.ibo.org/programmes/find-an-ib-school/

IBO, International Baccalaureate Organisation. (2020). Find an IB World School: Netherlands. Retrieved from https://www.ibo.org/programmes/find-an-ib-school/?SearchFields.Country=NL&page=2

IBO, International Baccalaureate Organisation. (2022a). Language and Literature. Retrieved from https://www.ibo.org/programmes/middle-years-programme/curriculum/language-and-literature/

IBO, International Baccalaureate Organisation. (2022b). MYP Projects. Retrieved from https://www.ibo.org/programmes/middle-years-programme/curriculum/myp-projects/

Jongh, PJ de, Jongh, E de, Pienaar, M., Gordon-Grant, H., Oberholzer, M. & Santana, L. (2015). The impact of pre-selected variance inflation factor thresholds on the stability and predictive power of logistic regression models in credit scoring. *Orion*, 31(1), pp. 17-37.

Johnson, J. (2021). Most common languages on the internet 2020. Statistica. Retrieved from https://www.statista.com/statistics/262946/share-of-the-most-common-languages-on-the-internet/

Kamata, A., Turhan, A. ve Darandari, E. (2003). Estimating reliability for multidimensional composite scale scores. Presented at the annual meeting of American Educational Research Association, Chicago, April 2003.

Kang, M., Heo, H., Jo, I., Shin, J. & Seo, J. (2010) Developing an educational performance indicator for new millennium learners. *Journal of Research on Technology in Education,* 43 (2), 157-170.

Kang, J. & Keinonen, T. (2016). Examining factors affecting implementation of inquiry-based learning in Finland and South Korea. *Problems of Education in the 21st Century*, 74, 31-48.

Karademir, E. (2020). *21. yüzyıl becerilerinin geliştirilmesinde dijital öyküleme uygulamaları: özel yetenekli öğrenciler örnekleminde öğrenme ve yenilenme becerileri* (Master's thesis). Eskişehir Anadolu Üniversitesi Eğitim Bilimleri Enstitüsü, Eskişehir.

Karakaş, M. M. (2015). *Ortaokul sekizinci sınıf öğrencilerinin fen bilimlerine yönelik 21. Yüzyıl beceri düzeylerinin ölçülmesi* (Master's thesis). Eskişehir Osmangazi Üniversitesi Eğitim Bilimleri Enstitüsü, Eskişehir.

Kaya, S. (2017). *The relation of high school students' 21st century skills with student burnout and school engagement* [Doctoral dissertation]. Hacettepe University, Ankara.

Kaya, A. (2020). *Sınıf öğretmeni adaylarının 21. Yüzyıl becerileri kapsamında girişimcilik becerilerinin incelenmesi* (Master's thesis). Çanakkale Onsekiz Mart Üniversitesi Lisansüstü Eğitim Enstitüsü, Çanakkale.

Kilmen, S. (2014). Madde analizi, madde seçimi ve yorumlanması. R. N. Demirtaşlı, (Ed.). *Eğitimde Ölçme ve Değerlendirme* içinde (2.Baskı), (171-251). Ankara: Edge Akademi.

Kline, P. (1993). Personality: the psychometric view. New York, NY: Routledge.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York & London: The Guilford Press.

Koğar, H. & Yılmaz Koğar, E. (2016). *Comparison of Different Estimation Methods for Categorical and Ordinal Data in Confirmatory Factor Analysis .* Journal of Measurement and Evaluation in Education and Psychology , 6 (2) , 351-364 . DOI: 10.21031/epod.94857

Korkmaz, H. (2019). *Argümantasyon tabanlı bilim öğrenme yaklaşımının 7. sınıf öğrencilerin bilimsel epistemolojik inançlarına ve 21. Yüzyıl becerilerine etkisi* (Master's thesis). Muğla Sıtkı Koçman Üniversitesi Eğitim Bilimleri Enstitüsü, Muğla.

Kozma, R. (2009). Transforming education: assessing and teaching 21st century skills. In Scheuermann, F. & Björnsson (Eds.), *The Transition to Computer-Based Assessment: New Approaches to Skills Assessment and Implications for Large-scale Testing* (pp. 13-23). Luxemburg: European Commission JRC.

König, C., Steinmetz, H., Frese, M., Rauch, A. & Wang, Z.M. (2007). Scenario-based scales measuring cultural orientations of business owners. *Journal of Evolutionary Economics*, 17(2), 171-200.

Kuramoto, N. & Koizumi R. (2018) Current issues in large-scale educational assessment in Japan: focus on national assessment of academic ability and university entrance examinations, *Assessment in Education: Principles, Policy & Practice*, 25 (4), 415-433. DOI: 10.1080/0969594X.2016.1225667

Kutlu, Ö. (2006). Üst düzey zihinsel süreçleri belirleme yolları: Yeni durum belirleme yaklaşımları. *Çağdaş Eğitim*, 335.

Kutlu, Ö., Doğan, C. D. & Karakaya, İ. (2010). Öğrenci başarısının belirlenmesi : Performansa dayalı durum belirleme (3rd ed.). Ankara: Pegem Akademi.

Külegel, S. (2020). *Çevre eğitimine dayalı fen, teknoloji, mühendislik, matematik temelli etkinliklerin özel yetenekli öğrencilerin 21. Yüzyıl becerilerini geliştirmesine yönelik araştırma* (Master's thesis). Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.

Kyllonen, P. C. (2012). Measurement of 21st century skills within the common core state standards. Paper presented at: *Invitational Research Symposium on Technology Enhanced Assessments*. Educational Testing Service (ETS).

Lee, J. & Stankov, L. (2018). Non-cognitive predictors of academic achievement: evidence from TIMSS and PISA. *Learning and Individual Differences*, 65, 50-64.

Li, C.H. (2016) The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. *Psychol Methods*. 21(3):369-87. doi: 10.1037/met0000093. PMID: 27571021.

Lievens, F. & Sackett, P. R. (2012). The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance. *Journal of Applied Psychology*, 97(2), 460-468.

Mardia, K.V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57, 519-530. https://doi.org/10.1093/biomet/57.3.519

Martin-Raugh, M.P., Anguiano-Carrasaco, C. Jackson, T., Brenneman, M. W., Carney, L., Barnwell, P. & Kochert, J. (2018). Effect of situational judgment test format on reliability and validity. *International Journal of Testing*, DOI: 10.1080/15305058.2018.1428981

MEB (2011). Meb 21. Yüzyıl Öğrenci profili. Eğitimi Araştırma ve Geliştirme Dairesi Başkanlığı (EARGED). Retrieved from https://www.meb.gov.tr/earged/earged/21.%20yy_og_pro.pdf

Mecklin, C. J. & Mundfrom, D. J. (2005). A Monte Carlo comparison of the Type 1 and Type 2 error rates of tests of multivariate normality. Journal of Statistical Computation and Simulation, 75(2), pp. 93-107.

Mertler, C. A. & Vannatta, R. A. (2005). *Advanced and multivariate statistical methods: practical application and interpretation* (3rd ed.). Glendale, CA: Pyrczak Publishing.

Mihelich, M. (2013). Another generation rises: looking beyond the millenials. Workforce, Retrieved from https://workforce.com/uk/news/another-generation-rises-looking-beyond-the-millennials

Mountain, R., Care, E., Scoular, C. & Woods, K. (2009-2012). Module 5: Teaching and learning 21st century skills. Assessment Research Centre, Melbourne Graduate School of Education.

Murat, M. (2018). *Ters yüz sınıf modelinin beşinci sınıf öğrencilerinin 21. Yüzyıl becerileri ve bilimsel epistemolojik inançlarına etkisi* (Master's thesis). Muğla Sıtkı Koçman Üniversitesi Eğitim Bilimleri Enstitüsü, Muğla.

National Centre for Education Statistics. (n.d.). *PISA problem solving items and scoring guides.* Retrieved from https://nces.ed.gov/surveys/pisa/pdf/items2_solving.pdf

Nieveen, N., & Plomp, T. (2017). *Five guiding principles for curriculum change*. Enschede: SLO.

Nuffic (July, 2009). Grading system in the Netherlands, the United States and the United Kingdom. Retrieved from https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjcy83SjZT1AhVyi_0HHQkXCBEQFnoECAQQAQ&url=https%3A%2F%2Fstudents.uu.nl%2Fsites%2Fdefault%2Ffiles%2Fgeo-grading-systems-holland-vs-us-uk.pdf&usg=AOvVaw0jJgAr02CwZqvlBPE0rmeU

Nuffic (n.d.-a). Education in the Netherlands. The Dutch organization for Internationalization in education. https://www.nuffic.nl/en/subjects/study-holland/education-netherlands

Nuffic. (n.d.-b). Alle tto-scholen in Nederland. De Nederlandse organisatie voor internationalisering in onderwijs. https://www.nuffic.nl/onderwerpen/tweetalig-onderwijs/alle-tto-scholen-nederland

Nunnally, J.C. & Bernstein, I.H. (1994). *Psychometric Theory* (3rd ed.). New York, NY: McGraw-Hill, Inc.

Nusche, D., Braun, H., Halász, G. & Santiago, P. (2014). *OECD reviews of evaluation and assessment in education: Netherlands*. OECD.

Organization for Economic Co-operation and Development (OECD) (2004). Problem Solving for Tomorrow's World: First Measures of Cross-Curricular Competencies from PISA 2003. Retrieved from http://www.oecd.org/education/school/programmeforinternationalstudentassessmentpisa/34009000.pdf

Özgün, D. (2019). *İnsan ve çevre ünitesinin yaratıcı drama destekli işlenmesinin beşinci sınıf öğrencilerinin 21. Yüzyıl becerileri üzerine etkisi* (Master's thesis). Muğla Sıtkı Koçman Üniversitesi Eğitim Bilimleri Enstitüsü, Muğla.

Öztermiyeci, E. (2019). Türkiye'de Ulusal Program ve Uluslararası Bakalorya Programı öğrencilerinin 21. yüzyıl becerilerine ilişkin algıları / The perceptions of the students educated in national programme and international baccalaureate programme in Turkey about the 21st century skills (Master's thesis). Maltepe Üniversitesi, Sosyal Bilimler Enstitüsü, İstanbul.

Pallant, J. (2013). *SPSS survival manual*. McGraw-Hill Education (UK).

Partnership for 21st Century Learning (2019). Retrieved from http://www.lsce-mena.org/p21-partnership-for-21st-century-learning

Perry, L.B., Ledger, S. & Dickson, A. (2018). What are the benefits of the International Baccalaureate middle years programme for teaching and learning? Perspectives from stakeholders in Australia. Murdoch University, Perth, Australia.

Popham, W. J. (2003). *Test better, teach better: the instructional role of assessment*. Alexandria, VA: Association for Supervision and Curriculum Development

Principal components (PCA) and Exploratory factor analysis (EFA) with SPSS. UCLA: Statistical Consulting Group from https://stats.idre.ucla.edu/spss/seminars/efa-spss/ (accessed February 26, 2020).

Raykov,T. & Marcoulides,G.A. (2006). *A first course in structural equation modeling*. Mahwah, NJ: Erlbaum.

Ron, T. H. (2019). *Bringing situational judgement tests to the 21st century: scoring of situational judgement tests using item response theory*. [Doctoral dissertation]. Graduate College of Bowling Green State University.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1-36. http://www.jstatsoft.org/v48/i02/

Rosseel, Y. (2021) The lavaan tutorial https://lavaan.ugent.be/tutorial/cat.html

Rönkkö, M. & Cho, E. (2020). An updated guideline for assessing discriminant validity. *Organizational Research Methods*, 25(1), 6-47.

Saavedra, A. R. & Opfer, V. D. (2012, October). Learning 21st-century skills requires 21st-century teaching. *Kappan Magazine*, 94(2), 8-13.

Salters-Pedneault, K. P. (2018). *The use of self-report data in psychology.* Retrieved from https://www.verywellmind.com/definition-of-self-report-425267

Schermelleh-Engel, K., Moosbrugger, H. & Müller, H. (2003). Evaluating the fit of structural equation models: tests of significance and goodness-of-fit measures. *Methods of Psychological Research Online*, 8 (2), 23-74.

Schulz, W. (2005). Mathematics self-efficacy and student expectations. Result from PISA 2003. Paper prepared for Annual Meetings of the American Educational Research Association in Montreal.

Scoular, C. & Awwal, N. (2009-2012). Module 3: ATC21S Collaborative problem solving assessments. Assessment Research Centre, Melbourne Graduate School of Education.

Serdar, R. A. (2015). *Searching for innovation, creativity and 21st century skills: A case study of a suburban elementary school district*. [Doctoral dissertation]. National-Louis University. https://digitalcommons.nl.edu/diss/124

Silva, E. (May, 2009). Measuring skills for 21st-century learning. *Phi Delta Kappan*, 630-634.

Singh, A.P. & Dangmei, J. (2016). Understanding the generation Z: the future workforce. *South-Asian Journal of Multidisciplinary Studies*, 3(3).

Soland, J. Hamilton, L.S. & Stecher, B. M. (2013). *Measuring 21st century competencies: Guidance for educators.* A Global Cities Education Network Report (GCEN).

Sorrel, M.A., Olea, J., Abad, F.J., Torre, J., Aguado, D. & Lievens, F. (2016). Validity and reliability of situational judgement test scores: a new approach based on cognitive diagnosis models. *Organizational Research Models*, 1-27.

Suldo, S. M., Shaunessy, E. & Hardesty, R. (2008). Relationships among stress, coping, and mental health in high-achieving high school students. *Psychology in the Schools*, 45(4), 273-290.

Suhr, D. D. (2006). Statistics and Data Analysis Paper 200-31 Exploratory or Confirmatory Factor Analysis?.

Swartz, R. & McGuinness, C. (June, 2014). Developing and assessing thinking skills. IB Continuum. Retrieved from https://www.ibo.org/globalassets/publications/ib-research/continuum/student-thinking-skills-en.pdf

Szmigiera, M. (2021). The most spoken languages worldwide 2021. *Statistica*. Retrieved from https://www.statista.com/statistics/266808/the-most-spoken-languages-worldwide/

Şencan, H. (2005). *Sosyal ve davranışsal ölçümlerde güvenirlik ve geçerlilik*. Ankara: Seçkin Yayıncılık.

Şimşek, Ö.F. (2007). *Yapısal eşitlik modellemesine giriş (Temel ilkeler ve Lisrel uygulamaları).* Ankara: Ekinoks.

Tabachnick, B. G. & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston, MA: Allyn & Bacon, Inc.

Thorndike, E. L. (1931). *Human learning* (2nd edition). New York, NY: Century Company.

Toe, D., Lang, J., Paatsch, L., Yim, B., Jobling, W., Doig, B. & Aranda, G. (2015, November). Assessment of student development and learning in IB PYP schools. International Baccalaureate Organization School of Education, Deakin University. Retrieved from https://www.ibo.org/contentassets/cc0cf81a37814d94b46b9711a50a1a98/final-report-ib-pyp-assessment.pdf

Trilling, B. & Fadel, C. (2009). *21st century skills: learning for life in our times*. San Francisco CA: John Wiley & Sons, Inc. Retrieved from https://books.google.nl/books?hl=nl&lr=&id=VUrAvc8OB1YC&oi=fnd&pg=PR7&dq=21st+century+skills&ots=DweUOvcxeR&sig=-bd46_iMxDl5WVJwUiPE97_COkk#v=onepage&q=21st%20century%20skills&f=false

Walker. (n.d.) *Chapter 14 Factor Analysis, Path Analysis, and Structural Equation Modeling.* Jones and Bartlett Publishers, pp.324-351. Retrieved from https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwiSz7CInJH1AhWPNOwKHTE9DqcQFnoECAYQAQ&url=https%3A%2F%2Fsamples.jblearning.com%2F0763755486%2F55485_CH14_Walker.pdf&usg=AOvVaw1oNdPTh32X19EfMDdr_LAQ

Weekley, J. A. & Jones, C. (1999). Further studies of situational tests. *Personal Psychology*, 52, 679-700.

World Schools (2022). International Baccalaureate: pros vs cons. Retrieved from https://world-schools.com/international-baccalaureate-pros-vs-cons/

Van Laar, E., Van Deursen, A.J.A.M., Van Dijk, J.A.G.M. & De Haan, J. (2020). Determinants of 21st-century skills and 21st-century digital skills for workers: a systematics literature review. *SAGE Open*, 1-14.

Van Veen, K. Bloemert, J. & Wolthuis, F. (2020). Preparing the Dutch educational system for the 21st century. *Oxford Research Encyclopedias*. https://doi.org/10.1093/acrefore/9780190264093.013.1482

Vivek, P. H, Singh, S. N., Mishra, S. & Donavan, D. T. (2017). Parallel Analysis Engine to Aid in Determining Number of Factors to Retain using R [Computer software], available from https://analytics.gonzaga.edu/parallelengine/.

Voogt, J., Dede, C. & Erstad, O. (2009). TWG 6: 21st century learning. International Summit on ICT in Education.

Voogt, J., Erstad, O., Dede,C. & Mishra, P. (2013). Challenges to learning and schooling in the digital networked world of the 21st century. *Journal of Computer Assisted Learning*, 29, 403-4013.

Yalçın, S. (2018). 21. Yüzyıl becerileri ve bu becerilerin ölçülmesinde kullanılan araçlar ve yaklaşımlar. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 51(1), 183-201.

Yavaş, K. (2021). *21. yüzyıl becerilerinin geliştirilmesine yönelik bir model önerisinin ortaokul öğrencilerinin bazı becerileri üzerine etkisi* (Master's thesis). Fırat Üniversitesi Eğitim Bilimleri Enstitüsü, Elazığ.

Yılmaz, E. & Alkış, M. (2019). 21. Yüzyıl yeterlikleri ölçeği'nin geliştirilmesi: Geçerlik ve Güvenirlik Çalışması. The Journal of International Lingual, Social and Educational Sciences, 5(1), 125-154.

Zygmont, C. & Smith, M.R. (2014). Robust factor analysis in the presence of normality violations, missing data, and outliers: Emprical questions and possible solutions. *The quantitative Methods for Psychology*, 10(1), 40-55.

APPENDIX-A: Parent Information Letter

Dear parents,

Your child's school is willing to participate in a scientific study into the perception of 21st century skills. Last decades, humanity is in a constant race to catch the improvements in technology and effects of them in daily life. Involved competencies for daily routine as well as expected skills for many professions have been evolved which has brought the requirement of change in how we perceive education. Future generations are expected to adapt to new forms of socialization and have an active part in economic development of a knowledge society which we call as 21st century skills.

Unfortunately, we do not yet know how to develop those skills in depth. As soon as we gain more insight into circumstances that hinder or reinforce these, we can make better statements about what our students need for the future expects them. This research aims to develop a scale in order to assess and gain more insight about students' perception of their 21st century skills competences. In this sense, it can contribute to the improvement and development of existing and new support programs. Additionally, the study will investigate the relation between students' perception of their 21st century skills and their problem-solving skills which is one of the critical skills of today and future.

What exactly does the research entail?

In this study we want pupils between 12 and 16 years old to complete a questionnaire and a problem-solving skills test. The students can complete both task at school in the classroom. It will take a maximum of 45 minutes to complete.
Application of the questionnaire and problem-solving skills test will be provided by the researcher and/or the class teacher at school.

No risk

There are no inconveniences or risks for the students in participating in this project. The investment consists of the time the students spend completing the questionnaire.
The research is a scientific research. In this study we look at the 21st century skills perception of the students as well as problem-solving skills.

What do I get for participating in the study?

In addition to participating in an important study and contributing to an increase in knowledge about 21st century skills of children, the school receives a report of the most important findings.

What about privacy?

We treat your child's details with the utmost care and strict confidentiality. All data that we receive and collect will be anonymous. We do this by saving everything with a code or number. We call this "data that cannot be traced back to the person". As a result, when we look at the results of the investigation, we do not see what information your child has. All collected data is stored securely, and can only be viewed by the researcher, and by people who check the quality of the research. With the research data, we only make general statements about groups of people. We never make statements about individuals.

Where can I go with complaints?

If you have complaints about the study, you can report it to the researcher, Buket Eren Janssen.

How can your child participate in the study?

Your child's school is willing to participate in this study. We make agreements with the school about a suitable time. However, participation is entirely voluntary. Your child will receive an explanation about the application at school. You can make this known by sending an email to Buket Eren Janssen: buket.eren2013@gmail.com. Please state the name, class and school of your child.

APPENDIX-B: Expert View Request Document and Sample Item for the Initial Form of PoTCS Scale

Dear language or assessment expert,

First of all, thank you very much for your help with my research.

Below, you will read some short stories, which aim to measure different aspects of a person's perception of his/her 21st century skills. You are expected to;

1) Make any editing needed on the stories as crossing the existing word or sentence and write the correct version with another color.
2) State whether the story is measuring the perception about the stated factor or not, and comment if needed.

The items are formed according the Binkley et al. (2012: p.36)'s framework. The framework is as follows;

Ways of Thinking:

11. Creativity and innovation
12. Critical thinking, problem solving, decision making
13. Learning to learn, Metacognition

Ways of Working:

14. Communication
15. Collaboration

Tools for Working:

16. Information literacy
17. ICT literacy

Living in the world:

18. Citizenship- local and global
19. Life and career
20. Personal and social responsibility-including cultural awareness and competence.

Answer options are ranked from 5: *Totally suits me* till *1: Totally not suits me.*

Thank you for your contribution,

B. Eren, MEd, MSc
International School Breda
Hacettepe University

Do the following items measure the perception of ***creativity and innovation****? Which one would measure "creativity and innovation"* ***the best (Rank 1)****? And which one would measure those* ***the least (Rank 3)****?*

Question/Item *(please make any editing needed on the text)*	***Yes***	***No***	***Rank (1 to 3)***	***Comment***
1. At visual arts or design lessons, our teacher expects us to produce original works. For example, designing an object like a doorstopper or creating a storybook with an original story or building a mind game are some of our tasks to complete**. I am very interested in such creative tasks and enjoy them a lot.**				

APPENDIX-C: Expert View Summary for the Initial Form of PoTCS Scale

Item Number	Expert 1	Expert 2	Expert 3	Expert 4
1	+2	+2	+2	1
2	+1	+3	+1	3
3	+3	+1	+3	2
4	+2	+	+	2
5	+1	+	+	3
6	-3	+	+	1
7		+2	+2	3
8		+1	+1	2
9		+3	-3	1
10	+1	+3		1
11	+2	+1	-	3
12	+3	+2		2
13	-3	+1	+1	3
14	-2	-3	+2	2
15	+1	+2	+3	1
16 (Reverse)	-3	-3	-3	3
17	+1	+2	+1	1
18	+2	+1	+2	2
19	-3	+1	+2	3
20	+1	+2	+1	2
21	+2	+3	+3	1
22 changed	-3	+3		3
23 changed	-2	+2		2
24	+1	+1		1
25 changed	+3	-3	-	3
26 changed	+2	-2	-	2
27	+1	+1	-	1
28	+1	+2	+2	3
29	-3	+3	+1	1
30	+2	+1	+3	2

APPENDIX-D: Initial Form of Perception of 21st Century Skills Scale

What is your age?.....

Dear students,
In the following part, you will read some short stories about different people. Please put yourself in that person's place and answer according to how much the **bold** sentence would suits to you if you were in that person's position.

Answer options are ranked from 5: *Totally suits me* till *1: Totally not suits me.*

Short Story	***5:Totally suits me***	***4:Somewhat suits me***	***3: Neutral***	***2:Somewhat not suits me***	***1:Totally not suits me***
1. In visual arts and design lessons, our teacher expects us to produce original works. For example, designing an object like a doorstopper, creating a storybook with an original story or building a mind game are some of our tasks to complete**. I am very interested in such creative tasks and enjoy them a lot.**					
2. At school, we sometimes need to brain storm to find a solution for a certain problem in order to develop some skills, such as being critical, creative or innovative. In such cases, **I can find the most different and original solutions.**					
3. When I look around and see all the technological items we use today, such as mobile phones or new tech cars, I start to think about what kinds of skills I need to create such an item. **I would be very interested to work in an innovative area and I would be very good at it.**					
4. I heard a piece of news on the radio yesterday that there is no solution for global warming anymore and the world is on an irreversible path. **This news made me think how that could be possible. I needed to learn how much that news was reliable and if there is still some solutions I needed to search for them.**					

Short Story	***5:Totally suits me***	***4:Somewhat suits me***	***3: Neutral***	***2:Somewhat not suits me***	***1:Totally not suits me***
5. This week in History class, we got research homework. I have looked through many different sources and obtained contradictory information. **In order to get reliable information, I decided to contact an expert on this topic and ask my questions.**					
6. Last week was all about protecting the environment and specifically plastic usage in our school. We discussed the problem of how we can reduce plastic consumption in our school. **I came up with various realistic solutions and shared those in student council.**					
7. We started a new topic in Math this week. I have never been introduced to that topic before**. I would like to get a more understanding so I read about this topic and its origins on the internet after school.**					
8. At language class last week, we had read a piece from a brilliant article. I was thinking about that article today to rephrase that piece to support my idea during history class**. I realized that I cannot remember the key point of the article, so I found the article from my last week notes and re-read it.**					
9. At science class, we had learned effect of forces last month. Today, our science teacher introduced us the new topic which is the solar system. **I instantly recognized the gravity as the force and its effect in the solar system.**					
10. We worked on a poster project today with two other friends. Our teacher expects us to explain our task visually as a poster and present it verbally to our classmates. **I can express my part in front of other classmates fluently and confidently.**					

Short Story	***5:Totally suits me***	***4:Somewhat suits me***	***3: Neutral***	***2:Somewhat not suits me***	***1:Totally not suits me***
11. Thanks to the internet, we can read countless sources from different countries and cultures. Understanding more than one language has a crucial role in understanding a wide range of sources. **I am happy to be able to understand multiple languages, so I can read different news or articles and understand different points of views.**					
12. This semester, I need to make a project about the use of technology in communication. I have decided to make interviews with people from diverse walks of life, **such as engineers, filmmakers, supermarket cashiers, and teachers. I can communicate with all these people comfortably.**					
13. “Four eyes are better than two” is a proverb which we can interpret as working together with someone makes it easier and more practical to complete a task. We often get assignments to work as a team at school, and **I think this makes our work more valuable because we can complement each other because of various perspectives we have.**					
14. During classwork sometimes we need to work as pairs or as small groups. Mostly, some group members put less effort and time on the task. **I prefer to make a chart with task division for group members and follow it to make the work more efficient and fair.**					
15. Last month, I had participated in a model competition with two other students from our school. I do not have any common view on anything with those students, but we are good at building the model. D**espite our differences, I could work collaboratively with these students.**					

Short Story	***5:Totally suits me***	***4:Somewhat suits me***	***3: Neutral***	***2:Somewhat not suits me***	***1:Totally not suits me***
16. In today's world, it is easy to access to an excessive amount of information. **I know how to find reliable sources, evaluate the information I got from there, and use it in different occasions.**					
17. As today's students, we are overwhelmed with the amount of information we can find via the internet. When we search for a school-related project, we come across loads of related information. N**ow I am more aware of approaching any information critically in terms of its correctness.**					
18. In the coming three weeks, we will work on an internal assessment in Mathematics class. We need to start writing an introduction about the background of our research question. **I found various sources to support my introduction and I am aware of how to cite those sources and form a bibliography.**					
19. Some subjects have many units based on visual understanding. Technology can be very helpful for a better understanding of such subjects. In our school, not only our teachers but also students need to use computers often. **I can use my computer confidently to find any related videos, maps or schemas, and improve my understanding of such units.**					
20. Each school year, we get a project or term paper to complete in a longer time period. **I can use the computer to do my project and try to learn different online tools to enrich it.**					
21. The internet allows us to communicate with a wider community and get better standards and different perspectives on our projects. **I ask my questions on online platforms and get answers from all over the world about a problem.**					

Short Story	*5:Totally suits me*	*4:Somewhat suits me*	*3: Neutral*	*2:Somewhat not suits me*	*1:Totally not suits me*
22. Democratic processes got influenced by each individual in a society who have enough maturity and information to contribute. **I am aware of my responsibility in terms of human rights, including diversity and equality of people.**					
23. Technology has not only brought us some practical devices but also changed some crucial definitions in our lives. For instance, being a 'citizen' does not only define the legal member of a country anymore but it also refers to be a member of the globe. **I can recognize the differences between people's identities and respect these differences.**					
24. We live in a production and consumption era now, whereas I am aware of the consequences of overconsumption for the globe. **I feel responsible as a local and global citizen to take measures against these consequences.**					
25. Improvements in technology affects the job descriptions as well. There is a chance that I will have a job which does not exist at the moment. **I am aware of the importance of adaptability and flexibility skills, and I work to improve these skills.**					
26. In an increasingly globalized world, many workplaces are multicultural, and we may need to cooperate with colleagues across the globe. **I know the importance of feedback and criticism from multicultural perspectives to improve myself for a global standard.**					
27. Education helps us to gather comprehensive knowledge about different concepts across different subject areas. **I care to be knowledgeable**, **because knowledge will help me to form my future learnings and make connections in real-life situations, so that I can become a lifelong learner.**					

Short Story	*5:Totally suits me*	*4:Somewhat suits me*	*3: Neutral*	*2:Somewhat not suits me*	*1:Totally not suits me*
28. We live in a consumption era. A wide variety of products have been produced and consumed by people. In this rapid change of items, **I do not like to renew but care my belongings.**					
29. There are 7.7 billion people and seven main cultural regions in the world. I love the culture and traditions of my family. On the other hand, each cultural practice, belief or custom is special. **I feel responsibility to support people from different cultures to follow their traditions, as long as they do not harm anyone.**					
30. Nowadays, biological products become much more popular. Of course, all of us prefer to eat healthier and less processed food in ideology. On the other hand, **I cannot make sense of the cost of biological products. Getting healthy food should not be only a right for rich people. I would like to take action to make this situation more fair.**					

APPENDIX-E: Demographic Information Questionnaire

Please mark the suitable option with ***X*** into the brackets ( ) if any given.

1. Gender: ( ) Female ( ) Male

2. Grade / Year Group: ……..

3. Age:

   ( ) younger than 12
   ( )12
   ( )13
   ( )14
   ( ) 15
   ( )16
   ( ) older than 16

4. What is your general point average GPA (overall grade average) on your last report?

   ……out of…… (For example: 58 out of maximum 100)

5. What is your English language grade on your last report?

   ……out of…… (For example: 58 out of maximum 100)

6. What is your Mathematics grade on your last report?

   ……out of…… (For example: 58 out of maximum 100)

7. How do you define your English level?

   ( ) Not so good ( ) Good ( ) Very good ( ) Mother tongue

8. Which programme you are enrolled in? ( ) IB ( ) Other

   If you are enrolled in IB, how long you have been in this system?

   ( ) Less than 1 year
   ( ) 1 years
   ( ) 2 years
   ( ) 3 years
   ( ) 4 years
   ( ) 5 years
   ( ) more than 5 years

APPENDIX-F: Instructions for Mentors

Dear year tutors/teachers,

This is the second phase application of a PhD dissertation project, which aims to measure students' perceptions about their 21st century skills competences and problem solving skills.

Please share the survey link with your students and ask them to answer each questions according to **what suits them personally the best**.

https://docs.google.com/forms/d/e/1FAIpQLScji6LYpgowCS7sw-AJxYPPbLUETExBoLSTxBv36tUf10W_eQ/viewform?usp=sf_link

When your group completes the survey, please e-mail me during the same class hour which class and how many students submit the survey. The survey will be anonymous and so I need to track if and which group complete the survey (i.e. MYP1A, 20 students).

You can mention the following points in front of the class.

1. Please read the part which starts with "Dear students" carefully. (According to the year group, you might prefer to read it out loud in front of the class)

   Dear students,

   As answering the following questions, you are helping for research which aims to create a questionnaire to observe students' perceptions about their 21st-century skills competences. Thank you for your help and contribution.
   All the information collected with this document will be anonymous.
   Please answer each question realistically but only as focusing what fits you the best. You need to state your age and read the following short stories about different people. Put yourself in that person's place and answer according to how much the quotation marked (" ") sentence would suits you if you were in that person's position.
   Answer options are ranked from 5: Totally suits me till 1: Totally not suits me.
   Thanks for your contribution,

2. Please do not skip any question.
3. For each statement click which option suits you the best and when the survey is completed click to submit.
4. There are 30 questions (except the question of age), and according to the trials expected time to complete the survey is maximum 15 minutes.

Thank you for your contribution to this scientific research and good luck with the application.

Kind regards,

Buket Eren, MEd, MSc

APPENDIX-G: Final Perception of 21st Century Skills Scale

In the following part, you will read some short stories about different people. Please put yourself in that person's place and answer according to how much the **bold** sentence would suits to you if you were in that person's position. Please mark the most suitable answer for each sentence as **X**.

Answer options are ranked from 5: *Totally suits me* till *1: Totally not suits me.*

Short Story	***5:Totally suits me***	***4:Somewhat suits me***	***3: Neutral***	***2:Somewhat not suits me***	***1:Totally not suits me***
1. At school, we sometimes need to brain-storm to find a solution for a certain problem in order to develop some skills, such as being critical, creative or innovative. In such cases, **I can find the most different and original solutions.**					
2. We started a new topic in Math this week. I have never been introduced to that topic before**. I would like to get a more understanding so I read about this topic and its origins on the internet after school.**					
3. Technology has not only brought us some practical devices but also changed some crucial definitions in our lives. For instance, being a 'citizen' does not only define the legal member of a country anymore but it also refers to be a member of the globe. There are different societies on the globe. **I can recognize that different societies have different identities and I respect these differences.**					
4. This week in History class, we got research homework. I have looked through many different sources and obtained contradictory information. **In order to get reliable information, I decided to contact an expert on this topic and ask my questions.**					
5. When I look around and see all the technological devices we use today, such as mobile phones or new tech cars, I start to think about what kinds of skills I need to create such a device. **I would be very interested to work in an innovative area.**					
6. At science class, we had learned effect of forces last month. Today, our science teacher introduced us the new topic which is the solar system. **I instantly recognized the gravity as the force and its effect in the solar system.**					

Short Story	***5:Totally suits me***	***4:Somewhat suits me***	***3: Neutral***	***2:Somewhat not suits me***	***1:Totally not suits me***
7. We worked on a poster project today with two other friends. Our teacher expects us to explain our task visually as a poster and present it verbally to our classmates. **I can express my part in front of other classmates fluently and confidently.**					
8. We live in a production and consumption era now, whereas I am aware of the consequences of overconsumption for the globe. **I feel responsible as a local and global citizen to take measures against these consequences.**					
9. As today's students, we are overwhelmed with the amount of information we can find via the internet. When we search for a school-related project, we come across loads of related information. N**ow I am more aware of approaching any information critically in terms of its correctness.**					
10. Last month, I had participated in a model competition with two other students from our school. I do not have any common view on anything with those students, but we are good at building the model. D**espite our differences, I could work collaboratively with these students.**					
11. This semester, I need to make a project about the use of technology in communication. I have decided to make interviews with people from diverse walks of life, such as engineers, filmmakers, supermarket cashiers, and teachers. **I can communicate with all these people comfortably.**					
12. "Four eyes are better than two" is a proverb which we can interpret as working together with someone makes it easier and more practical to complete a task. We often get assignments to work as a team at school. **I think this makes our work more valuable because we can complement each other with different perspectives we have.**					

Short Story	***5:Totally suits me***	***4:Somewhat suits me***	***3: Neutral***	***2:Somewhat not suits me***	***1:Totally not suits me***
13. In the coming three weeks, we will work on a research project in Mathematics class. We need to start writing an introduction about the background of our research question. **I found various sources to support my introduction and I am aware of how to cite those sources and form a bibliography.**					
14. Some subjects have many units based on visual understanding. Technology can be very helpful for a better understanding of such subjects. In our school, not only our teachers but also students need to use computers often. **I can use my computer confidently to find any related videos, maps or schemas, and improve my understanding of such units.**					
15. There are 7.7 billion people and seven main cultural regions in the world. I love the culture and traditions of my family. On the other hand, each cultural practice, belief or custom is special. **I feel responsibility to support people from different cultures to follow their traditions, as long as they do not harm anyone.**					
16. Democratic processes got influenced by each individual in a society who have enough maturity and information to contribute. **I am aware of my responsibility in terms of human rights, including diversity and equality of people.**					
17. I heard a piece of news on the radio yesterday that there is no solution for global warming anymore and the world is on an irreversible path. **This news made me think how that could be possible. I need to learn how much that news is reliable and if there is still some solutions, then I need to be aware of them.**					
18. In today's world, it is easy to access to an excessive amount of information. **I know how to find reliable sources, evaluate the information I got from there, and use it in different occasions.**					

Short Story	***5:Totally suits me***	***4:Somewhat suits me***	***3: Neutral***	***2:Somewhat not suits me***	***1:Totally not suits me***
19. In an increasingly globalized world, many workplaces are multicultural, and we may need to cooperate with colleagues across the globe. **I know the importance of feedback and criticism from multicultural perspectives to improve myself for a global standard.**					
20. Each school year, we get a project or term paper to complete in a longer time period. **I can use the computer to do my project and try to learn different online tools to enrich it.**					
21. Nowadays, biological products become much more popular. Of course, all of us prefer to eat healthier and less processed food in ideology. On the other hand, **I cannot make sense of the cost of biological products. Getting healthy food should not be only a right for rich people. I would like to take action to make this situation more fair.**					

APPENDIX-H: Expert View Request Document for Problem-Solving Skills Test

Dear language or assessment expert,

Thank you again for your contribution to this research.

Below, you will read three problem solving questions. You are expected to make any editing needed on the questions as crossing the existing word or sentence and write the correct version with another color. (You can use “Track changes” on word document.)

If you are an assessment expert, please also comment if the questions measure Decision making (Question 1), System analysis and design (Question 2), and Troubleshooting (Question 3).

Kind regards,

B. Eren, MEd, MSc
International School Breda
Hacettepe University

APPENDIX-I: Problem Solving Skills Test

This part of the survey aims to assess your problem-solving skills. Please read each question carefully and state your answer. You **can use calculator** and spare paper for your calculations.

FLYING TO TURKEY

Ella lives in the Netherlands with her family and she wants to visit her grandmother during the Autumn holiday. Ella's grandmother lives in a small city called Bilecik (Turkey). Ella would like to fly from one of the three airports closest to her house. The cost of travel from Ella's house to these airports (one way) are given in the following table.

	Cost of driving
Eindhoven airport EIN	€ 3
Amsterdam airport AMS	€ 20
Dusseldorf airport DUS	€ 23

There is no airport at Bilecik, so Ella needs to find a flight to İstanbul, which is the closest city with international airports. There are two airports in Istanbul. The cost of travel from these airports to Ella's grandmother's house (one way) are given as follows:

	Cost of driving
SAW airport	€ 25
IST airport	€ 35

Ella's Autumn Holiday is from the 12th of October until the 20th of October and she finds direct

Question 1. FLYING TO TURKEY

Taking into account the information Ella researched about flights, and travel costs to and from the different airports, which of the four flight options is **the cheapest** and which of the four flight options is **the most expensive** for Ella to go to her grandmother's? (Answer options and calculation area are on the next page). Please show your calculations where needed.

Flight Options	Airports and Timing*	Round Trip Flight Prices
Option 1	13:50 EIN — 3h 35 Direct → 18:25 SAW 10:30 SAW — 3h 40 Direct → 13:10 EIN	€ 325
Option 2	19:05 AMS — 3h 35 Direct → 23:40 SAW 10:40 IST — 3h 40 Direct → 13:20 AMS	€ 208
Option 3	14:35 AMS — 3h 35 Direct → 19:10 IST 10:40 IST — 3h 40 Direct → 13:20 AMS	€ 260
Option 4	08:00 DUS — 3h 15 Direct → 12:15 SAW 14:20 SAW — 3h 20 Direct → 16:40 DUS	€ 300

***Timing:** *Example:* 3h 35= 3 hours 35 minutes

Calculations:

Total Price for;

Option 1	
Option 2	
Option 3	
Option 4	

*Circle your answer below (**The cheapest**).*

Option 1	Option 2	Option 3	Option 4

*Circle your answer below (**The most expensive**).*

Option 1	Option 2	Option 3	Option 4

Question 2. FLYING TO TURKEY

For Ella's trip, time it takes to reach each airport differs because of traffic situations and timing. These times are given as an average in the following table. Considering flight time (on the previous page) and the time it takes to reach airports by car, which of the four flight options from the previous part takes **the shortest** and which of the four flight options takes **the longest *time*** for Ella to go to her grandmother's?

Airport names	Time between the destinations and the airports
EIN	10 min
AMS	1hour 10 min
DUS	1 hour 20 min
SAW	2 hour 20 min
IST	3 hour 10 min

Calculations:

Total Time for;

Option 1	
Option 2	
Option 3	
Option 4	

*Circle your answer below (**The shortest**).*

Option 1	Option 2	Option 3	Option 4

*Circle your answer below (**The longest**).*

Option 1	Option 2	Option 3	Option 4

PET SITTER

The Palace Pet Sitter is open during the winter holiday. They have reservations for 28 pets (18 dogs and 10 cats) for the winter holiday. There are six caretakers and five rooms. Three of the caretakers are specialized in dogs and the other three are specialized in cats. You can see the information about the caretakers, rooms and the Palace Pet Sitter **rules** on the following boards.

Board 1: Caretakers	
Dog	***Cat***
Nelly	Abby
Bruno	Martin
Bart	Ayla

Board 2: Rooms	
Room Name	Capacity
A	10
B	6
C	6
D	4
E	4

RULES

1. Cats and dogs must stay in separate rooms.
2. At least one caretaker should be responsible for a room according to their expertise.
3. One caretaker can be responsible for a maximum of 8 pets.

Question 1. PET SITTER

Complete the table to place the 28 pets and caretakers in the rooms, bearing all the rules in mind.

Room	**Number of dogs**	**Number of cats**	**Name(s) of caretakers**
A			
B			
C			
D			
E			

COOKIE PRODUCTION LINE

Diva works in a cookie factory as a production line supervisor. Production line is a rail system which carries the raw material to multiple processes and eventually to its packaging point. You can see all the sections of the cookie production line for Diva's on Diagram1. All the sections of this production line is furnished with an airtight cover for hygienic reasons, so if there is a malfunctioning section it is not possible to see from outside which section is malfunctioning.

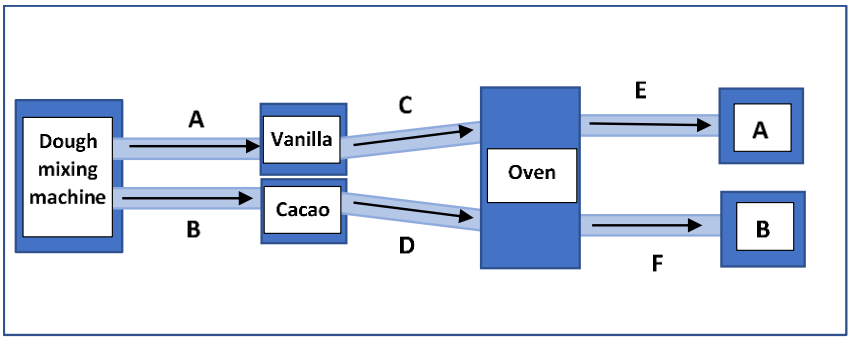


Diagram 1: Original Production Line (Version 1)

The sections of the production line includes:

- Dough mixing machine
- **A:** Oil spraying belt1
- **B:** Oil spraying belt2
- Vanilla and cacao adding machines
- **C:** Slicing belt1
- **D:** Slicing belt2
- Oven
- **E:** Cooling belt1
- **F:** Cooling belt2
- Packaging machines A and B

Diva checked the production line when it is designed as Diagram 1. She realized that cookies arrive to packaging point B, whereas **not** to the packaging point A. Diva has to find the malfunctioning section. She made some changes in the production line to figure out the problem and tried versions 2, 3 and 4 (check the following page). Belts do not intersect in any version of the production line. When she rearranged the production line as on given diagrams, she collected the following information:

	Cookie in A	Cookie in B
Diagram 1	No	Yes
Diagram 2	Yes	No
Diagram 3	Yes	No
Diagram 4	No	No

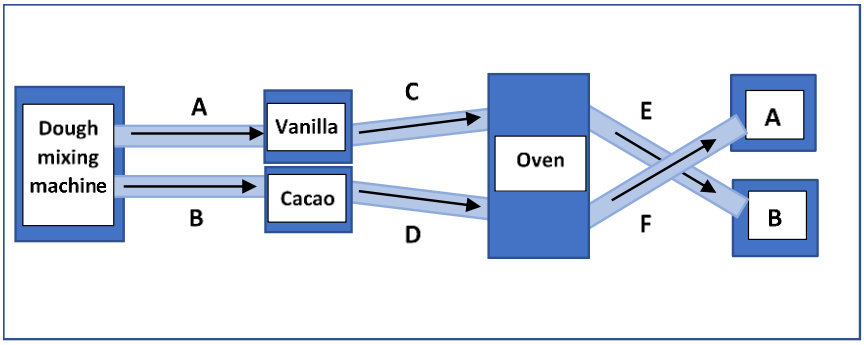


Diagram 2: Production Line Version 2

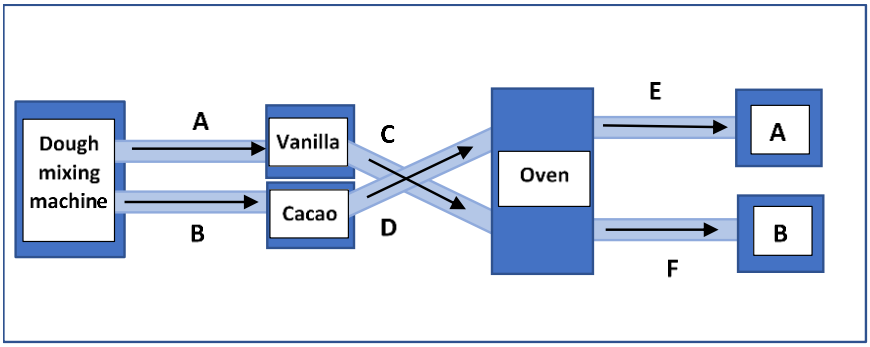


Diagram 3: Production Line Version 3

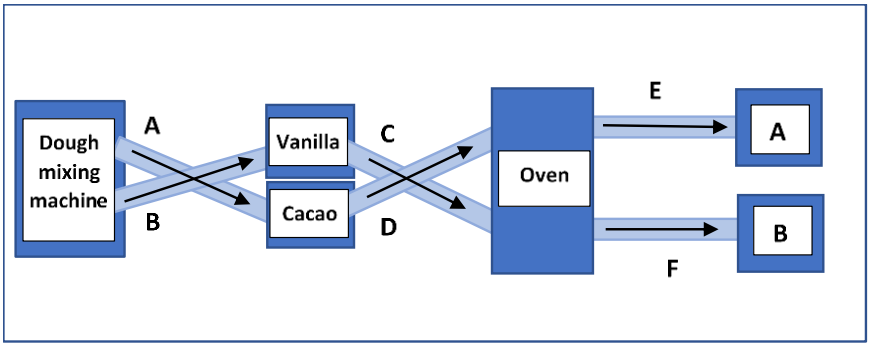


Diagram 4: Production Line Version 4

Question 1: COOKIE PRODUCTION LINE

Taking into account the information collected through different diagrams, which section(s) of the production line could be malfunctioning/broken? Please circle the possible correct option for each section.

Name of the section	*May be Malfunctioning?*	*Name of the section*	*May be Malfunctioning?*
A	Yes / No	D	Yes / No
B	Yes / No	E	Yes / No
C	Yes / No	F	Yes / No

APPENDIX-J: Problem Solving Skills Test Rubric

Calculations:

Marking:

- A3: Total prices are correct and the cheapest and the most expensive are correct.
- A2: Total prices are not given or wrong correct but the cheapest and the most expensive are correct.
- A1: One of the cheapest or the most expensive is correct.
- A0: Not answered or all wrong.

Total Price for;

Option 1	€ 381
Option 2	€ 308
Option 3	€ 370
Option 4	€ 396

*Circle your answer below (**The cheapest**).*

Option 1	Option 2	Option 3	Option 4

*Circle your answer below (**The most expensive**).*

Option 1	Option 2	Option 3	Option 4

Question 2. FLYING TO TURKEY

Calculations:

Marking:

- A3: Total times are correct and the shortest and the longest times are correct.
- A2: Total times are not given or wrong correct but the shortest and the longest times are correct.
- A1: One of the shortest or the longest times is correct.
- A0: Not answered or all wrong.

Total Time for;

Option 1	12 h 15 min (735 min)
Option 2	15 h 5 min (905 min)
Option 3	15 h 55 min (955 min)
Option 4	13 h 55 min (835 min)

*Circle your answer below (**The shortest**).*

Option 1	Option 2	Option 3	Option 4

*Circle your answer below (**The longest**).*

Option 1	Option 2	Option 3	Option 4

System analysis and design

Question 1. PET SITTER

Complete the table to place the 28 pets and caretakers in the rooms, bearing all the rules in mind.

Room	Number of dogs	Number of cats	Name(s) of caretakers
A	8		Nelly
B	6		Bruno
C	4		Bart
D		6	Abby, Martin
E		4	Ayla

Alternative

Room	Number of dogs	Number of cats	Name(s) of caretakers
A	6		Nelly
B	6		Bruno
C	6		Bart
D		5	Abby, Martin
E		5	Ayla

Marking:

- A3: All rules and information given were satisfied.
- A2: One rule or information given was violated.
- A1: Two of the rules or information given were violated.
- A0: Not answered or more than two rules/information given were violated.

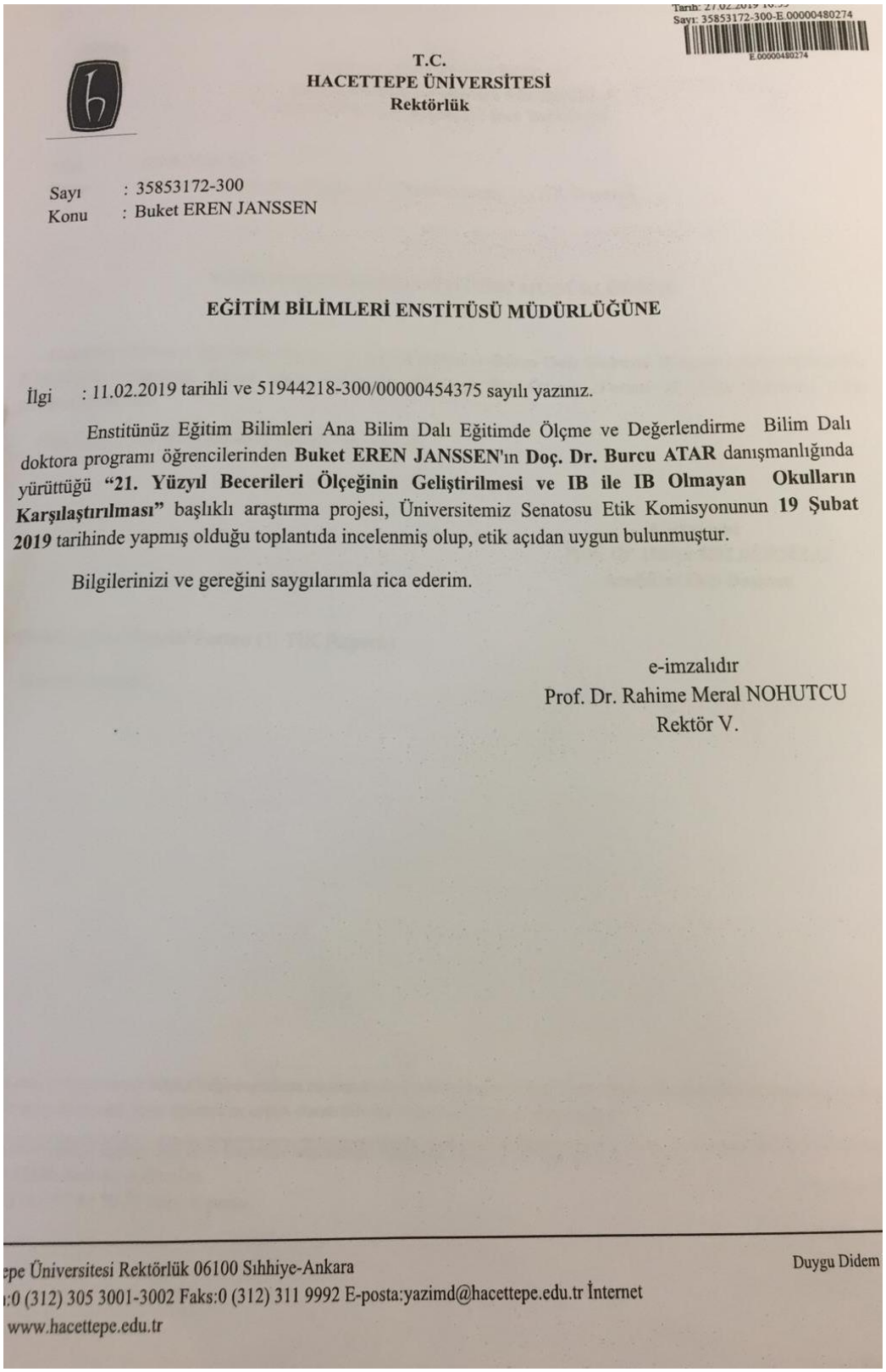
Trouble shooting

Question 1: COOKIE PRODUCTION LINE

Taking into account the information collected through different diagrams, which section(s) of the production line could be malfunctioning/broken? Please circle the possible correct option for each section. **Marking: Each correct answer is one mark.** Total: 6 Marks

Name of the section	*May be Malfunctioning?*	*Name of the section*	*May be Malfunctioning?*
A	Yes / No	D	Yes / No
B	Yes / No	E	Yes / No
C	Yes / No	F	Yes / No

APPENDIX-K Ethics Committee Approval



Tarih: 27.02.2019 10:39
Sayı: 35853172-300-E.00000480274

T.C.
HACETTEPE ÜNİVERSİTESİ
Rektörlük

Sayı : 35853172-300
Konu : Buket EREN JANSSEN

EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

İlgi : 11.02.2019 tarihli ve 51944218-300/00000454375 sayılı yazınız.

Enstitünüz Eğitim Bilimleri Ana Bilim Dalı Eğitimde Ölçme ve Değerlendirme Bilim Dalı doktora programı öğrencilerinden **Buket EREN JANSSEN**'ın **Doç. Dr. Burcu ATAR** danışmanlığında yürüttüğü **"21. Yüzyıl Becerileri Ölçeğinin Geliştirilmesi ve IB ile IB Olmayan Okulların Karşılaştırılması"** başlıklı araştırma projesi, Üniversitemiz Senatosu Etik Komisyonunun **19 Şubat 2019** tarihinde yapmış olduğu toplantıda incelenmiş olup, etik açıdan uygun bulunmuştur.

Bilgilerinizi ve gereğini saygılarımla rica ederim.

e-imzalıdır
Prof. Dr. Rahime Meral NOHUTCU
Rektör V.

epe Üniversitesi Rektörlük 06100 Sıhhiye-Ankara
:0 (312) 305 3001-3002 Faks:0 (312) 311 9992 E-posta:yazimd@hacettepe.edu.tr İnternet
www.hacettepe.edu.tr

Duygu Didem

APPENDIX-L Declaration of Ethical Conduct

I hereby declare that…

- I have prepared this thesis in accordance with the thesis writing guidelines of the Graduate School of Educational Sciences of Hacettepe University;
- all information and documents in the thesis/dissertation have been obtained in accordance with academic regulations;
- all audio visual and written information and results have been presented in compliance with scientific and ethical standards;
- in case of using other people's work, related studies have been cited in accordance with scientific and ethical standards;
- all cited studies have been fully and decently referenced and included in the list of References;
- I did not do any distortion and/or manipulation on the data set,
- and **NO** part of this work was presented as a part of any other thesis study at this or any other university.

(28) /(02)/(2022)

(Signature)
Buket Eren Janssen

APPENDIX-M: Thesis/Dissertation Originality Report

HACETTEPE UNIVERSITY

Graduate School of Educational Sciences

To The Department of Educational Sciences

Thesis Title: Investigating the variables affecting students' perceptions of 21st century skills and problem-solving skills

The whole thesis that includes the *title page, introduction, main chapters, conclusions and bibliography section* is checked by using **Turnitin** plagiarism detection software take into the consideration requested filtering options. According to the originality report obtained data are as below.

Time Submitted	Page Count	Character Count	Date of Thesis Defence	Similarity Index	Submission ID
28/02 /2022	134	228611	26/01 /2022	13%	1772586283

Filtering options applied:

1. Bibliography excluded
2. Quotes included
3. Match size up to 5 words excluded

I declare that I have carefully read Hacettepe University Graduate School of Educational Sciences Guidelines for Obtaining and Using Thesis Originality Reports; that according to the maximum similarity index values specified in the Guidelines, my thesis does not include any form of plagiarism; that in any future detection of possible infringement of the regulations I accept all legal responsibility; and that all the information I have provided is correct to the best of my knowledge.

I respectfully submit this for approval.

Name Lastname:	Buket Eren Janssen
Student No.:	N15141791
Department:	The Department of Educational Sciences
Program:	Program of Doctor of Philosophy in Educational Measurement and Evaluation
Status:	☐ Masters ☒ Ph.D. ☐ Integrated Ph.D.

Signature

ADVISOR APPROVAL

APPROVED
(Assoc. Prof. Dr. Burcu Atar)

APPENDIX-N: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kâğıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alınarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan **"Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına ilişkin Yönerge"** kapsamında tezim aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H.Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- Enstitü/Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihinden itibaren 2 yıl ertelenmiştir. (1)
- Enstitü/Fakülte yönetim kurulunun gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren … ay ertelenmiştir. (2)
- Tezimle ilgili gizlilik kararı verilmiştir. (3)

28 /02 /2022

(imza)

Buket EREN JANSSEN

"Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge"

(1) Madde 6. 1. Lisansüstü tezle ilgili patent başvurusu yapılması veya patent alma *sürecinin devam etmesi durumunda, tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü Üzerine enstitü veya fakülte yönetim kurulu iki yıl süre ile tezin erişime açılmasının ertelenmesine karar verebilir.*

(2) Madde 6.2. Yeni teknik, materyal ve metotların kullanıldığı, henüz makaleye dönüşmemiş veya patent gibi yöntemlerle korunmamış ve internetten paylaşılması durumunda 3. şahıslara veya kurumlara haksız kazanç; imkânı oluşturabilecek bilgi ve bulguları içeren tezler hakkında tez danışmanın önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulunun gerekçeli kararı ile altı ayı aşmamak üzere tezin erişime açılması engellenebilir.

(3) Madde 7. 1. Ulusal çıkarları veya güvenliği ilgilendiren, emniyet, istihbarat, savunma ve güvenlik, sağlık vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı, tezin yapıldığı kurum tarafından verilir. Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü tezlere ilişkin gizlilik kararı ise, ilgili kurum ve kuruluşun önerisi ile enstitü veya fakültenin uygun görüşü Üzerine üniversite yönetim kurulu tarafından verilir. Gizlilik kararı verilen tezler Yükseköğretim Kuruluna bildirilir.*
Madde 7.2. Gizlilik kararı verilen tezler gizlilik süresince enstitü veya fakülte tarafından gizlilik kuralları çerçevesinde muhafaza edilir, gizlilik kararının kaldırılması halinde Tez Otomasyon Sistemine yüklenir

** Tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu tarafından karar verilir.*