UNIVERSITY OF TWENTE

THE INFLUENCE OF FOLLOWING AN EXTRACURRICULAR HONORS PROGRAM ON THE DEVELOPMENT OF TWENTY FIRST CENTURY SKILLS

Author: Maschja Baas

Master thesis (final project) EST (201200036)
Enschede, June 2020

First supervisor: Dr. Marieke van Geel
Second supervisor: Dr.ir. Jean-Paul Fox

University of Twente
Department ELAN

UNIVERSITEIT TWENTE.

Table of contents

Abstract (English version)

Knowledge and skills that students need for success in the global workplace of the future are conceptualized as 'Twenty First Century Skills' (TFCS). Even though it is essential that students develop TFCS, learning these skills is currently insufficiently present in regular secondary education. When students do not learn to master these skills, this can cause problems in their role in society. In contrast to regular education, the focus on TFCS is visibly present in honors programs like the honors program of the Pre-University program in Twente. This is a university offered program specially aimed for high school students. Since the Pre-U honors program focuses on TFCS, it is interesting to study the effect of this program on the development of TFCS. Hence, in this study we evaluated what the influence was of following the Pre-U extracurricular honors program on the development of TFCS among 11th grade honors students (experimental group), compared to students who only follow regular 11th grade education (control group). A longitudinal study was conducted amongst 11th grade students of 25 high schools, following the "VWO track". Data was collected three times over the course of eight months through a self-assessment questionnaire that was developed for this study. The results showed as hypothesized that students in the Pre-U honors program (experimental group) scored significantly higher than regular 11th grade students (control group) on TFCS at the beginning and at the end of the honors intervention. Contrary to our expectations, students in the control group did develop their TFCS over time, whilst students in the honors program did not to develop their skills over time. Implications of these results and suggestions for future research are discussed in the thesis.

Abstract (Dutch version)

Kennis en vaardigheden die studenten nodig hebben voor succes in de (wereldwijde) arbeidsmarkt van de toekomst, worden gedefinieerd als '21e-eeuwse vaardigheden'. Hoewel het essentieel is dat leerlingen 21e-eeuwse vaardigheden ontwikkelen, leren leerlingen deze vaardigheden momenteel onvoldoende in het reguliere secundair onderwijs. Wanneer leerlingen deze vaardigheden niet leren beheersen, kan dit problemen veroorzaken in hun rol in de samenleving. In tegenstelling tot het reguliere onderwijs is de focus op 21e-eeuwse vaardigheden zichtbaar aanwezig in honoursprogramma's zoals het honoursprogramma van het Pre-University in Twente. Dit is een door de universiteit aangeboden programma dat speciaal is bedoeld voor middelbare scholieren. Omdat het Pre-U honoursprogramma zich richt op 21e-eeuwse vaardigheden, is het interessant om het effect van dit programma op de ontwikkeling van 21e-eeuwse vaardigheden te bestuderen. Daarom hebben we in deze studie geëvalueerd wat de invloed was van het volgen van het Pre-U extra curriculaire honoursprogramma op de ontwikkeling van 21e-eeuwse vaardigheden bij honoursleerlingen (experimentele groep), vergeleken met leerlingen die alleen het reguliere onderwijs volgen (controlegroep). Er is een longitudinaal onderzoek uitgevoerd onder leerlingen in 5 VWO van 25 middelbare scholen. De data is in de loop van acht maanden driemaal verzameld via een zelfevaluatievragenlijst die voor deze studie is ontwikkeld. De resultaten lieten zoals verwacht zien dat leerlingen in het Pre-U honoursprogramma (experimentele groep) aan het begin en aan het einde van de honours-interventie significant hoger scoorden dan reguliere leerlingen (controlegroep) op 21e-eeuwse vaardigheden. In tegenstelling tot onze verwachtingen ontwikkelden leerlingen in de controlegroep hun 21e-eeuwse vaardigheden in de loop van de tijd wel, terwijl leerlingen in het honoursprogramma hun vaardigheden niet in de loop van de tijd ontwikkelden. Implicaties van deze resultaten en suggesties voor toekomstig onderzoek worden besproken in het verslag.

1. Introduction

The modern society changes and evolves rapidly with extensive (technological) developments. Information is always available and increases at a fast pace. As a result, for students to thrive in this world they must not only possess basic skills in science and mathematics but must make proficient use of skills as creativity, critical thinking and problem solving, collaboration and communication (World Economic Forum, 2015). A change of focus in the content and shape of education might therefore be necessary (Thijs, Fisser, & Van der Hoeven, 2014). In an ideal state of affairs, students thus learn the knowledge and skills that fit these developments in their school curriculum (Kaufman, 2013). Knowledge and skills that students need for success in the (global workplace of the) future are conceptualized with the overarching term 'Twenty First Century Skills' (TFCS, Germaine, Richards, Koeller, & Schubert-Irastorza, 2016; Larson & Miller, 2011; World Economic Forum, 2015).

One can wonder how new these skills to function in the 21st century society really are. The answer is simple: the concept of TFCS is not new. Throughout history, several trends in renewing education have occurred (Kaufman, 2013). Skills that now fall under the umbrella term TFCS have been major influencers of human progress throughout history (Rotherham & William, 2009). According to Silva (2009, p. 631) 'TFCS aren't new, just newly important". This means that there is a need in society for individuals to develop these skills (Kaufman, 2013). It thus has become clear that TFCS are essential to learn in education. Even though in research there is a slight variation of which skills considered part of TFCS (Gutman & Schoon, 2013) and which skills are most important (Thijs et al., 2014; Ledoux et al., 2013), consensus can be found in skills such as creativity and innovation, critical thinking and problem-solving skills, communication and collaboration skills (Thijs et al., 2014). These skills are also called the four core learning TFCS, as selected by the "Partnership of TFCS" (P21) who uses the corresponding name for the international commonly used model "P21" (P21, 2019).

The current problem is that students do not learn TFCS enough in their education and thus have not mastered these skills when they go into the job market (Saavedra & Opfer, 2012; World Economic Forum, 2015; WRR, 2014). In 2014, in global companies more than a third (36%) of these companies had problems in filling their positions due to a shortage of available employees with key TFCS (Manpower Group, 2014). Explanations why students do not learn the skills enough are for example that the skills are not explicitly taught in the curriculum (Rotherham & Willingham, 2009; Saavedra & Opfer, 2012; Willingham, 2008), which can be

due to limited time because of the general focus on (the raise of) test scores (Germaine et al., 2016). Furthermore, research claimed that it is harder to teach students TFCS as a teacher (Saavedra & Opfer, 2012; Willingham, 2008), probably since it cannot be learned through textbooks and lectures (Saavedra & Opfer, 2012). Next to the lack of focus and challenge in teaching, it is also hard(er) to asses TFCS (Greiff & Kyllonen, 2016; Saavedra & Opfer, 2012; Silva, 2009).

It is quite common for employees to be able to critically and creatively evaluate knowledge and work as a collaborator in a team (Salmons, 2019; World Economic Forum, 2015). Many researchers have claimed the importance of developing critical thinking skills (Samuel, 2019) and communication skills (Covey, 1989) in the last decades. Thus, when students will not develop these essential TFCS, they have a hard time keeping up in the world in all areas of life. They are more likely to encounter problems in higher education (Kaufman, 2013), participating in the labor market (Wagner, 2008; Saavedra & Opfer, 2012), using the rights and responsibilities that contribute to a healthy society (Saavedra & Opfer, 2012) and keeping up with the globalization in the world (Saavedra & Opfer, 2012).

Since TFCS are not explicitly taught in the curriculum of regular education (Rotherham & Willingham, 2009; Saavedra & Opfer, 2012; Willingham, 2008) it is important to research if students do learn these skills elsewhere, like in alternative programs, to see whether the learning of TFCS could be improved in regular education. In recent years in education and in the job market, there has been a growing focus on talented students, their extracurricular (honors) programs and the skills that are learned there (Renzulli, 2012; Segers & Hoogeveen, 2012). An example of such an honors program is the Pre-University (Pre-U) honors program in which the university offers courses specifically aimed at 11th and 12th grade high school students. In these programs, the main focus lies on the development of skills such as critical and creative thinking, collaboration and communication (Pre-U, 2019a). This focus thus lies on the development of TFCS, where students learn knowledge and skills in addition to what the regular secondary education curriculum offers. Due to this specific focus, following a program like the Pre-U program could thus possibly be a solution to the problem, not mastering TFCS, and potentially stimulate the growth of TFCS to a large extent. Therefore, the goal of this study is to explore the effect of following the extracurricular Pre-U honors program on the development of TFCS and to compare that effect to the development of TFCS of their peers who only follow regular secondary education.

2. Theoretical framework

In the following paragraphs, the most common TFCS, that are selected for this research, will be described in more detail. After TFCS are discussed it is important to take a closer look towards the context in which they are researched: the extracurricular honors programs. Therefore, general information will be provided regarding extracurricular honors programs following a description of how the TFCS are explicitly taught in the Pre-U program, which will be the intervention for this research.

2.1 Twenty First Century Skills

This research focuses on the four core learning TFCS, also called the four c's, as selected by the international commonly used model "P21" (P21, 2019): creativity and innovation (C1), critical thinking and problem-solving (C2), collaboration (C3) and communication (C4). Note that it is useful to make a distinction between the four skills in theory, however in practice the skills are related and connected to each other (Germaine et al., 2016). In the subsequent paragraphs, the four core learning TFCS will be explained in more detail, together with some examples about their relation towards each other.

2.1.1 Creativity and innovation (C1)

Creativity and innovation relate to each other since creativity asks to connect existing elements together, so that something new, innovative, is created (Tigchelaar, 2015). Creativity is the "ability to imagine and devise innovative new ways of addressing problems, answering questions or expressing meaning through the application, synthesis or repurposing of knowledge" (World Economic Forum, 2015, chapter 1). It regards generating new ideas and be open and responsive to them or redefine already existing ideas (NEA, 2012 as mentioned in Germaine et al., 2016). Creative people are interested to try out new and original ideas, brainstorm or mind map about them and working to improve them, which often involves working together with others (Germaine et al., 2016) Therefore, creativity and innovation is closely related to collaboration. Being creative provides one with more opportunities for critical thinking and problem solving as well (Padget, 2012), linking it to another TFCS. Creative people view failure as an opportunity to learn from their mistakes (NEA, 2012 as mentioned in Germaine et al., 2016). In our ever changing world creative thinking skills are crucial to learn (and therefore crucial to teach) in the future (Gardner, 2007).

2.1.2 Critical thinking and problem-solving (C2)

Critical thinking means "to reason effectively, recognizing connections between systems, concepts, and disciplines to solve problems and make decisions" (Germaine et al., 2016, p21). It requires "clarity, accuracy, and precision of expression; relevance of arguments or questions; logic of thought; and thinking with sufficient depth and breadth to consider complexities and perspectives of an issue" (Germaine et al., 2016, p21). Critical thinking and problem-solving skills are intertwined since people who apply critical thinking by identifying, analysing and evaluating information and ideas can solve social, scientific and practical problems effectively (Snyder & Snyder, 2008; World Economic Forum, 2015, chapter 1). Critical thinkers needs to be open-minded to alternative views and be able to reflect on one's own learning process (Thijs et al., 2014).

2.1.3 Collaboration (C3)

Collaboration means working effectively in different and diverse groups, to be willing to make compromises, value each members' contribution and share responsibility for the work (P21, 2019). Furthermore, it is important to have respect for each other's (cultural) differences (Thijs et al., 2014) and to be able to communicate effectively, stating a relationship with the skill 'communication' (Thijs et al., 2014; Germaine et al., 2016). Lastly, people who collaborate work more efficient, solve problems and innovate faster, having access to the best skills and strengths and develop their own skills (Trilling & Fadel, 2009).

2.1.4 Communication (C4)

Communication refers to the ability to express thoughts and ideas in a written, verbal and non-verbal way in varieties of contexts. Communication can be used for different reasons, such as to inform or persuade someone or to listen effectively to others (P21, 2019). According to Germaine et al. (2016, p22) "effective communication requires that the message be transmitted, heard, and understood within the context of intellectual virtues".

All in all, developing these four TFCS are crucial to be able to work in the workplace (of the future) and it became the norm that employees possess these skills (World Economic Forum, 2015). The section below will further explain the existence and role of honors programs and how they can stimulate TFCS growth.

2.2 Honors education

Honors programs are developed to stimulate high potential students to develop themselves beyond the scope of their primary field of study and the offer of programs in higher education keeps growing. According to Wolfensberger et al. (2012) the programs are laboratories for educational innovation and this has a strong influence on the educational policy not only in higher education, where most honors programs are offered, but in primary and secondary education as well.

Honors programs in universities can typically be divided into three different types of programs: mono-disciplinary programs, interdisciplinary programs and multidisciplinary programs (Wolfensberger, van Eijl, & Pilot, 2012). A mono-disciplinary honors program defines itself by being organized per university major or department, and thus has a specific subject. Interdisciplinary honors programs bring together different departments and their subjects and are offered for students of all disciplines. A multidisciplinary honors program usually is designed as an entire bachelor program for talented students at a university, which teaches in Liberal Arts and Sciences (Wolfensberger, 2015). Over the years, not only the offer in honors programs grew but also the number of bachelor students (3,3% of all students in 2012) and master students (2,6% of all master students in 2012) who participated in these programs (Sirius expertcommissie, 2014).

The growing offer in honors programs seems like a good trend, since the need for honors programs in secondary education exists as well. Research showed that a quarter of all secondary school students were bored by their curriculum and found it to be too easy, and this number even raised up until 56% among the top 20% students (Rijksoverheid, 2014). In order to stimulate these high-potential but now bored students, an extensive provisions plan was created in 2014. What is important to mention is that these programs focus on the 20% best performing students, at all levels. This thus means that it does not regard only the student who knows a lot but also a student who is creative, shows craftmanship or has other interesting competences (Rijksoverheid, 2014).

Because of the demand for educational programs for high potential secondary school students, a lot of new programs were developed. First of all, many secondary schools offer extra or different courses (Breetveld, Meijer and Koopman, 2012), falling under the category of a mono-disciplinary program. Second of all, different national networks and programs exist to support education for high talented students (Sirius Programma, 2014 as discussed in

Wolfensberger, 2015). Local school initiatives for high-potential students can lead to a nationwide offered program (Stichting het Zelfstandig Gymnasium, 2020). Other examples of nationwide programs are the national Olympiads which are offered in different subjects (Wolfensberger, 2015). This thus seems to fall under the category of either a mono-disciplinary program (Olympiads) or an interdisciplinary program (Gymnasium), only it is not offered on one school or university but on multiple. Last of all, a majority of universities, in collaboration with national networks offer enrichment programs for talented secondary students as well (Sirius Expertcommissie, 2014). The universities offer secondary students enrichment programs on specific topics like science and mathematics (which can be compared to a mono-disciplinary program) (Wolfensberger, 2015; Pre-U, 2020a).

Even though several universities offer enrichment programs, Junior College Utrecht and Pre-U seem to be the only two offering a program specifically designed as an honors program for high school students, which is organized for students in 11th and 12th grade (Wolfensberger, 2015; Pre-U, 2019a). The programs are respectively called "U-Talent" (Junior College Utrecht) and "Pre-U honors program" (Pre-U). Both programs share similarities in organization and collaboration with their 27 partner schools (Wolfensberger, 2015; Pre-U, 2020b), organizing activities at the university campus (Wolfensberger, 2015; Pre-U, 2019a), creating interaction amongst students, have the goal to stimulate talent and offer programs like masterclasses and guest lectures (Michels & Eijkelhof, 2018; Pre-U, 2020a). The programs differ mostly on content in the provided modules. The U-Talent program focuses mostly on orientation and development of natural science and mathematics, whereas the Pre-U honors program focuses on preparing students for a relevant role in the society of the future where students develop TFCS, which means they learn to think creatively and critically, collaborate with other students and develop academic skills in all fields: beta, gamma and alfa (Pre-U, 2019a). These programs seem to overlap the most with a real interdisciplinary honors program as mentioned above. The next section will describe how the TFCS are taught and are present in the interdisciplinary Pre-U honors program. The "C" corresponding with the text will be in brackets, indicating creativity and innovation (C1), critical thinking and problem solving skills (C2), collaboration (C3) and communication (C4).

2.3 Description of intervention: Pre-U honors education

The Pre-U honors program, focusing on students in 11th and 12th grade consists of three honors modules, two masterclasses of choice and an extensive research project which together

takes one-and-a-half-year to complete (Pre-U, 2019b). Modules and masterclasses are lesson series of six classes. A masterclass is a scientific education program in which the student delves into a subject (C2) for six weeks (Pre-U, 2019d). The student will attend lectures, conduct practicals, collaborate on projects and sometimes also prepare at home (C1, C3, C4) (Pre-U, 2019d). The masterclasses are also open to non-honors students. They are therefore not an explicit part of the intervention and will not be discussed in relation to the intervention.

Since this research focuses on the Pre-U honors program it is relevant to know how TFCS are incorporated in the program. Therefore, the goals, approach and assessment of the honors program are analysed. These three topics are discussed in the subsequent paragraphs with emphasis on the four TFCS. The current study will focus on two third of the first year of the program. Table 1 below provides an overview of the Pre-U honors program, the current study period is marked blue.

Table 1

Overview of Pre-U honors education

	September October	November December	January February	March April	May June
5 VWO (11th grade)	Module 1	Masterclass 1	Module 2	Masterclass 2	Module 3
6 VWO (12th grade)	Conducting your own scientific research			Graduation	

2.3.1 Goals of the Pre-U honors program

The overall goal of the honors program is to develop a scientific knowledge base and to learn to creatively and critically make use of this knowledge. The honors program aims to convey why a critical attitude towards scientific knowledge and methods results in trust in the scientific method (C2) (Pre-U, 2019c). It is therefore essential for students to understand the origins of science, how it works and its continuous development. In addition, the program works

on creating a student attitude characterized by critical thinking (C2), optimism, creativity (C3), ambition and social vision (Pre-U, 2019c; Pre-U, 2019d).

The program is taught by four secondary education teachers. These teachers work one day a week for the Pre-U honors program, and spent the rest of their time teaching in regular secondary education. The four teachers develop, teach and evaluate the program intensively and function as a personal mentor and coach to the students. Since the program is therefore largely teacher dependent and not everything is documented, the explanations in the subsequent paragraphs about how TFCS are present in different parts of the programs are complemented by interviews and personal communication with these teachers of cohort 2019-2021.

2.3.2 Approach of the Pre-U honors program

2.3.2.1 Group composition. To achieve the goals, students work together with talented students from other schools on various challenging modules and students normally follow two masterclasses according to their own interests (C3) (Pre-U, 2019c). Studies show that students in high school honors programs especially value the cooperation with enthusiastic and motivated peers from other schools (Michels & Eijkelhof, 2018), which adds great value to their development (Wolfensberger & Pilot, 2014). Together with teachers the students create an enthusiastic community in which they pay in-depth attention to knowledge, skills and attitudes (Wolfensberger & Pilot, 2014) (C3). This is also the case in the Pre-U honors program where 38 students and 4 teachers together form one class. Students are also guided in this personal and group formation process. All meetings take place in lecture halls of the university where there is space for students and teachers to discuss and work together (C3, C4) (Pre-U, 2019c). The meetings are interactive. Students prepare material in advance and discuss it with each other during class (C3, C4) (Pre-U, 2019c).

2.3.2.2 Process of sessions: working and communicating together. Students in the Pre-U program collaborate during all modules, and in general as well. Next to the personal process that the students experience their group process is evenly important and ideally they strengthen each other (C3). All sessions start and end with a check-in and check-out assignment to connect to each other and the session. When the group functions well, students are able to grow to their fullest potential (ten Cate, Niemeijer & Bijman, personal communication, 2020).

Each session contains various working methods in which consultation between students is central (C4). During the sessions discussions occur and are stimulated with fellow students and teachers. During the first two modules the teachers work with group randomizers, to make

sure that for each collaboration assignment, a different team is formed (C3). Due to this method the students can find out which type of other student is a successful collaboration partner. During the collaborative assignments, the teachers keep an eye on the process (not punitive, but corrective) (ten Cate et al., personal communication, 2020).

The students are also divided amongst the teachers who function as their personal mentor and are the first contact person for the student. Sometimes the collaboration between students fails. If so, the students (either in the modules or the research project) are invited for a group conversation with the teacher, with the aim of getting the collaboration back on track. During the conversations, collaboration is continuously referred to as a skill, in which all parties are responsible for making further collaboration possible (C3, C4) (ten Cate et al., personal communication, 2020). The same occurs on individual level, if teachers feel that someone is not feeling well, they start a conversation with that student (C3, C4), normally the mentor of the student would do this. All this is very important for group formation and contributes to safety in the group, which must be as optimal as possible if the program wants students to develop and grow fully (ten Cate et al., personal communication, 2020).

2.3.2.3 Content of sessions. The content of the different modules are discussed in the following sections. It seems when looking at this content as if the content is mainly focused on critical and creative thinking. It is important to mention that content and process are interwoven and that during the sessions students therefore continuously cooperate and communicate and that they are therefore always working on the four skills (ten Cate et al., personal communication, 2020).

Focus module 1: philosophy of science. The aim of this first module is to cultivate curiosity about the origin of knowledge and the way in which this knowledge is used (C1, C2) (Pre-U, 2019d). In this way the teachers and students together build a foundation to understand how and why research is done (C2, C3) (Pre-U, 2019d). Students need this concept in order to be able to conduct their own research in the modules below (C1) (Pre-U, 2019d). In this module, students learn that truth is much more volatile than they might think, that truth is bound by time and context and that truth is even socially determined in a way (ten Cate et al, personal communication, 2020).

Focus module 2: Research methods. The aim of the module is to convert the philosophical foundation of the first module into the beginning of a concrete toolbox that students can apply in their own research (C1, C2) (Pre-U, 2019d). The accumulated understanding of truth and

truth-finding is used to get acquainted with various scientific methods. During this module, the student will gain crucial experiences with setting up and conducting academically sound scientific research and the alpha, gamma and beta domain. These experiences will help the student to draw up his or her own methodology and to be able to evaluate the reliability, validity and accuracy of a / their own research (C1, C2) (Pre-U, 2019d).

Focus module 3 and year 2 (not part of this intervention). During the third module, students combine what they have learned and start their own research with guidance from the university, which will be further carried out in the second year. The end product is a scientific "paper" (C1, C2, C3, C4) (Pre-U, 2019c). The student learns to take initiative and to take a critical look at their own research, but also at the research of others, for example by means of so-called peer reviews (C2, C3, C4) (Pre-U, 2019d).

2.3.3 Assessment in the Pre-U honors program

Students create a portfolio throughout the program (Pre-U, 2019c). A lot of attention is paid towards these portfolio assignments, which can be seen as a red thread throughout the program whereby the students' personal development is monitored. Through the portfolio assignments the students are challenged, bringing all TFCS elements together. Hereby, (and during the sessions) students are continuously trained to express their thinking into words and communicate with it and if they get stuck they get personal coaching by the teachers (C2, C3, C4) (ten Cate et al., personal communication, 2020).

The goal of the portfolio is not that students jump through a hoop, but rather involve the teachers in the thinking process. The student is encouraged to submit intermediate forms of the portfolio assignments to the teacher (their mentor), so that the teacher can provide feedback and together they can discuss the content via email or telephone (C1, C2, C3, C4).

All the things that the students have previously learned must be applied during the rest of the program. In the second portfolio assignment for instance, the student must prove that he or she can still work with the content of the first module, thereby the students thus use repetition and deepen the understanding of the learned skills and knowledge (C2) (ten Cate et al., personal communication, 2020).

2.3.3.1 Passing of first year. At the end of the first year, there is a go / no go moment where students are informed whether they are allowed to participate in the second year of the program. To be allowed to continue, the portfolio assignments of the students must have been

sufficiently completed and the research proposal for the honors research has been assessed with a pass. Students should not expect clear assignments with checklists for an assessment (Pre-U, 2019c). The main focus of assessment in the honors program is providing feedback and whether or not students apply this (ten Cate et al., personal communication, 2020). Grades are not part of the program. Ideally, continuing or quitting the program should come as no surprise, and the student and teacher come to such a conclusion together (C2, C4) (Pre-U, 2019c).

2.3.3.2 Graduation assessment. Everything the students' learned comes together in the second year in which students conduct a scientific research and write a paper. The focus lies on applying what has been learned, the importance of reflection on research and process and on the discussion of the research. The final product of the second year (not part of this intervention) is the paper report of the students' research, which a student also adds to their portfolio. The quality of this report determines whether the student has passed the entire honors program (Pre-U, 2019c). Quality however is not defined by a focus on truth finding by the student (although it is of course nice when a student "discovers" something), but it mainly focuses on the process and the reflective ability (C2). A failed experiment, with measurement errors and incorrect assumptions, but afterwards a great discussion with a critical look at one's own "failure", with well-founded self-insights and therefore the prospect of a future in which such an investigation is well tackled by the same student is seen as very valuable. In this, (self) critical thinking and problem-solving skills are therefore "key" (C2). All in all, this should make it possible for students to break free from their previous assumptions regarding truth, science and research (ten Cate et al., personal communication, 2020).

2.4 Research questions and model

Even though it is essential that students develop TFCS, learning these skills is currently insufficiently present in regular education. Since the focus on TFCS is visibly present in the Pre-U honors program, it is interesting to study the effect of such a program on the development of TFCS. Therefore, the research question of this study is: "What is the influence of following the extracurricular Pre-U honors program on the development of Twenty First Century Skills among 11th grade students, compared to students who follow only regular education?"

Data will be collected three times: at the start of the intervention (T1), in the middle of the intervention (T2) and at the end of the intervention (T3), which will occur over the course of eight months (see Table 2). Based on the theory and research question, two hypotheses were formulated:

Hypothesis 1: "*Honors students score higher on TFCS than regular students at baseline*".

Hypothesis 2: "*The growth in TFCS is higher for honors students than for regular students*".

Table 2

Overview of Pre-U honors education, with intervention and moments of data-collection marked green

	September - October	**November - December**	**January - February**	**March - April**	**May - June**
5 VWO (11th grade)	**(T1) Module 1**	**Masterclass 1 (T2)**	**Module 2**	**Masterclass 2 (T3)**	**Module 3**
6 VWO (12th grade)	**Conducting scientific research**			**Graduation**	

3. Methods

3.1 Respondents

Recruitment for respondents took place among 25 of the 27 Pre-U partner schools, since they teach an 11th grade class. This research had an experimental and a control group. Only participants who filled the questionnaire in entirely during one or more of the measurement moments were included.

3.1.1 Selection of experimental group

The experimental group started with N=38 students who were selected and followed honors education next to their regular 11th grade education at one of the 25 schools that were invited to participate. In total, 37 students finished the intervention (dropout rate 2.63%). Students applied for the honors program by writing a motivational letter, introducing themselves in an alternative way during an open assignment and by asking a teacher for a letter of recommendation (see the entire procedure in "Pre-U, 2019c"). In total, 76 students applied for the program. Due to practical reasons, the program had room for 35 to 40 students. Based on the students' application, on the profile of honors students that the program is looking for and based on a selection day, a selection committee selected the group of 38 students (rejection rate 50%). The profile that the selection committee looked at regarded a student who: is talented and motivated to further develop himself on all levels and it does not necessarily concern the "best" students with the highest grades, but students who thus want to develop themselves and could spent a whole day a week at the program (Pre-U, 2019b).

3.1.2 Selection of control group

All 25 schools were invited to participate as control group. Twenty schools decided to participate with their students (rejection rate 20%). This created a rather large control group. Since all the partner schools were invited to participate this also means that the 38 students who applied for the honors program but who were not selected were invited to participate in the control group. It is unknown how many of these students actually participated in the research and how they were divided amongst the participating schools. Respondents in the control group followed regular courses that are provided in the Dutch secondary education system.

3.1.3 General descriptives of the respondents

It differed per measurement how many students participated. Of the experimental group, the rejection and exclusion rate was zero during all measurements. The dropout rate was 2.63% for the last measurement. All students in the experimental group filled out the questionnaire

entirely during the first two measurements. In the control group, the rejection rate of students who did not want or were allowed to participate ranged between 3.30% and 5.22% during the three measurements. The dropout rate varied more across measurements with 6.55% during the first measurement and 25.24% during the last measurement. An explanation for the somewhat higher dropout rate during the last measurement can be found in the procedure. Due to the Covid-19 outbreak the schools were closed and students had to fill the questionnaire in from home, which might cause students to not finish the questionnaire. The exclusion rate, for students who were removed since they were unmatchable based on identical demographics, was rather low and varied between 0.49% and 0.86%. All in all, since data could only be used if the questionnaire was filled in entirely, only students who finished the questionnaire and were assigned a unique ID are seen as participants in this study. An overview of the gender and age of the participants can be found in Table 3 below.

Table 3

Participant overview age and gender

Measurement	Group	Age mean	Age SD	Gender Male	Gender Female	Gender 'other'
1	Experimental, N=38	16.45	0.58	44.7%	55.3%	0.0%
	Control, N=1324	16.48	0.81	44.7%	54.3%	1.0%
2	Experimental, N=38	16.72	0.58	44.7%	55.3%	0.0%
	Control, N=1045	16.73	0.72	44.2%	54.7%	1.1%
3	Experimental, N=37	17.03	0.59	43.2%	56.8%	0.0%
	Control, N=535	17.03	0.65	38.9%	60.7%	0.4%

3.2 Instrumentation

No validated questionnaire was yet available to properly measure TFCS, therefore the development of TFCS cannot be measured properly, which is considered a problem for policy-makers and educators. It is therefore extremely important that measurement instruments for these skills are developed (World Economic Forum, 2015). For the current research, several alternatives options were evaluated in order to determine TFCS-development of the students in the current study.

First of all, a choice had to be made regarding the form: either objective assessment, teacher assessment or student self-assessment. Objective validated scales could have been used to measure different components, like the "Watson Glaser Critical Thinking Appraisal" (Watson, 1980). Unfortunately, these questionnaires were too long or unreachable. Teachers could have been asked to assess the level of TFCS of their students since they are the expert who knows the student. However, this would cost too much time for the scope of this research. Taken both these arguments into account, self-evaluation was discussed as the best option. Using student self-assessment questionnaires is a well-established method of measuring TFCS (Child Trends, 2014 as mentioned in Lamb, Jackson & Rumberger, 2015). Next to that, using student self-assessment was a practical method of data collection since students could fill it in themselves during classes and it saves teachers a lot of time. Therefore, using self-assessment questionnaires was chosen as the best option.

Based on the 'TFCS standard rubric', which uses the official definitions of the P21-model of the skills to let teachers evaluate the level of the acquired skill in the student (CTE – Washington, 2019), a questionnaire was developed with 48 statements (see Appendix A). The statements were rephrased from a teacher perspective to a student perspective. All 48 statements (in Dutch) were answered with a 6-point-likert scale ranging from totally disagree (1) to totally agree (6). Six statements were reversed in order to control for a random completion of the questionnaire (see Appendix A).

- Creativity and innovation (C1): 17 statements, such as 'I use various creative techniques (such as brainstorming) to come up with new ideas'.
- Critical thinking and problem solving (C2): 16 statements, such as 'I like to analyze and evaluate perspectives that are very different from my own'.
- Collaboration (C3): 7 statements, such as 'In group work I feel co-responsible for the end product'.

- Communication (C4): 8 statements, such as 'I can always communicate well verbally and non-verbally with others, in different environments (think of school, sports, part-time job)'.

In order to check the face validity of the instrument, five students in 11th grade (who were not part of the main study) were asked to fill out the questionnaire. Furthermore they were asked to comment on statements if statements were unclear. After this feedback round, the TFCS self-assessment questions were revised. Next to the self-assessment questions, some questions regarding demographics were included, which will provide data to label the students of their gender, age, school and whether they follow honors education. This data is also used to create a unique student ID in order to match the students in the different data sets, which was later recoded into a random case ID in order to guarantee anonymity.

3.3 Procedure

The ethical committee of the University of Twente granted permission for this research. Students who were below the age of 16 needed permission from their parents in order to participate. Parents who give permission must sign a consent form (see Appendix B). Students who were above the age of 16 gave informed consent when they wanted to participate.

During both measurements in September and December, a 20-minute time frame was made available for students to fill in the questionnaire. Students in the control group filled out the questionnaire during regular class time at their secondary school, the students in the experimental group filled out the questionnaire during the Pre-U honors sessions. During the last measurement in April the schools were closed due to Covid-19 and all students were asked to fill in the questionnaire from home. The procedure of filling in the questionnaire stayed the same. The students' teacher shared the link to the questionnaire among the students. All students opened the link via their computer, tablet or mobile phone and they all received the same questionnaire. A hard copy version of the questionnaire was available for students who had no access to such a device, but in practice this was not necessary.

The questionnaire started with the informed consent page. The consent form explicitly mentioned that all information would be analyzed and handled anonymously and that students could drop out of the research anytime they wanted. Then, students filled in questions regarding demographics and all statements regarding TFCS. After the entire research was finished, the students were debriefed and 5 gift vouchers of 20 euros were spread by a randomizer to five randomly selected participants, who had left behind their email address so they could be

contacted. The email addresses were only used for the purpose of rendering the gift vouchers, and were deleted afterwards.

3.4 Analysis

3.4.1 Analysis of raw data

For each measurement, a new dataset was created including the students case ID, the measurement moment, their school number, whether or not they participated in the intervention and the answers to the 48 statements. The new datasets were merged together by cases. Table 4 below provides an overview of the sequential count of matching cases by measurement moment, which shows that 400 students in total filled in all three questionnaires.

Table 4

Sequential count of matching cases by measurement

Sequential count of matching cases	Measurement 1	Measurement 2	Measurement 3	Total
0	354	141	38	533
1	1008	34	0	1042
2	0	908	134	1042
3	0	0	400	400
Total	1362	1083	572	3017

3.4.2 Fitting a structural equation model

The merged file was loaded into R to fit a structural equation model of the data. The model was fitted in R version 3.6.2 (R Core Team, 2019) with the use of the lavaan package (Rosseel, 2012) in the default mode for ordinal data (see code in Appendix C). In lavaan, the first item loading of each latent variable is restricted to 1. When the first item was a reversed question, the second item was chosen to be restricted to 1 since the scores possibly contain response bias.

Based on the SEM model, the latent factor scores were estimated (see code in Appendix C). The choice to estimate latent factors using lavaan was made because when only average construct scores are used the added value of each item to the construct is lost, and every item counts equally to the construct, which would be the easy route. Instead of sum scores, therefore latent factor scores were estimated which take into account that some items are more important to the construct than others. After the dataset including the estimated latent factor scores was finished, indicator variables were created. The data was analyzed in long format where indicator variables were used to explain score differences between measurement moments and different TFCS subscales.

3.4.3 Shapiro-Wilk and Mann-Whitney tests

Now that the TFCS model fitted well and latent factor scores per construct were measured, analysis followed to answer the first hypothesis. Normally an independent sample t-test would have been used, which works on the assumption of normally distributed data. However, the Shapiro-Wilk test was conducted and showed significant departure from normality for all four subscales. Therefore, in order to continue further comparative analysis taken into account the non-normality of the data, Mann-Whitney tests were conducted for the four subscales for both hypotheses.

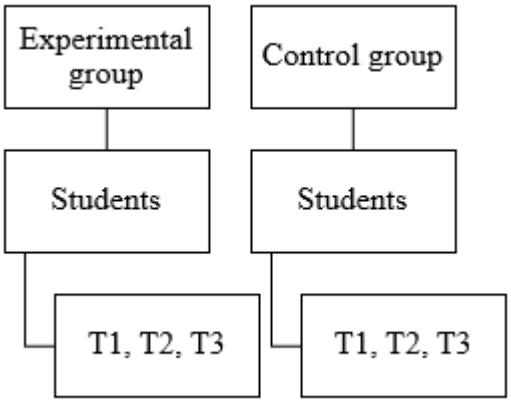
3.4.4 Using linear mixed effect modelling

Linear mixed effect models were analysed as following step, to be able to answer the second hypothesis. There is no perfect model to analyse this data. Ideally, one should have taken into account that the honors students are also nested in a school. However, this would have provided problems with honors students without participating schools (seven students; given the size of the experimental group and their added value these seven students had to be included) and schools without honors students (three schools). In addition, one of the components of the honors program is aimed at community building. The honors students therefore form their own group and cannot be placed back easily in a regular school. Both options were not perfect, to say that students do not belong to a school or to say that they do belong to a school. All in all after careful consideration, taking into account the original research question, while retaining the data and development of the honors group, the following two analyses were chosen: a longitudinal two and three level growth model. The growth models were fitted using the “linear mixed-effects model” (LME) package (Bates et al., 2015).

3.4.4.1 Longitudinal two-level growth model. A longitudinal two-level growth model was fitted in which the honors group is compared to the entire control group and whether there are significant differences in development. The predictor in this model is the honors intervention. In this two-level growth model (Figure 1 below) the school level was not taken into account. To evaluate growth in the latent factor scores (for each of the subscales, TFCS in general and for both the experimental and control group) it was tested whether the random intercept variance is greater than zero, which implies that the average score varied across the three measurement moments. This is represented by testing the ICC, which represents the same conclusion if tested to be positive. The ICC is tested by a Bayesian hypothesis test, which was conducted with the BFpackage based on Mulder et al. (2019). Nowadays, ICC plays an important role in modelling hierarchically structured educational data, providing insight in between-group and within-group variation (Mulder & Fox, 2019). Due to using the ICC in hypothesis testing, grouping effects across different group categories can be better assessed. The ICC scale is independent, making it easier to identify growth in latent factor scores (Mulder & Fox, 2019).

Figure 1

The longitudinal two level growth model as evaluated with LME - LME results were used for Bayesian hypothesis testing



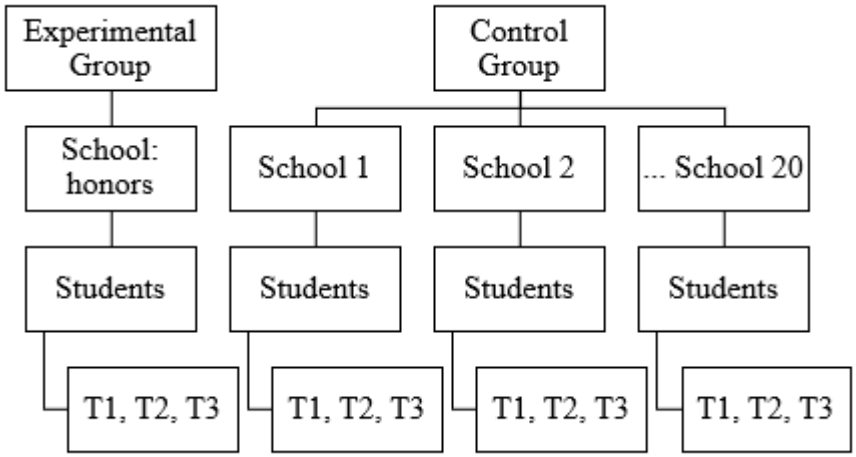
3.4.4.2 Longitudinal three level growth model.

In figure 1 above, students in the control group are presented as one group. It was important to also examine whether there were differences between the experimental and control group when we checked for the clustering of students in schools (see Figure 2 below). The Bayes factor testing of the random intercept variances while controlling for the clustering of

students in schools was not possible yet with BFPack. Therefore, an LME analysis was conducted.

Figure 2

The longitudinal three level growth model including the check for clustering in schools as tested in LME



4. Results

The following section provides the results of the several analysis that were conducted. First of all the descriptive statistics of the provided answers will be shown, which is the raw data. Then, the results of the structural equation model analysis will be provided. After that, the results of the LME analyses and Bayesian hypothesis testing are provided.

4.1 Descriptive statistics of raw data

The ordinal data was analysed with SPSS and provided results that show that not the whole range of answer options was used. Students in the experimental group seem to score much higher than students in the control group, and the range between minimum and maximum is low for both groups (see Table 5 below). Furthermore, looking on item level (Appendix D) the distribution of individual statement scores showed a 25% quartile around 3-4 and a 75% quartile around 4-5. This means that in general over/around 75% of the participants were in agreement with the statements on the four TFCS subscales. During the first and second measurement, several items were questioned reversed. The reversed items presented most diversity in scoring. Since this indicated that students might not read the questions well enough, the third measurement did not make use of reversed questions anymore.

Table 5

Descriptive statistics at scale level including percentiles

Scale	Group	Measure-ment	Min	Max	Mean	SD
Creativity and innovation	Experimental	1	3.24	5.29	4.41	.43
		2	3.18	5.35	4.35	.47
		3	3.35	5.47	4.40	.50
	Control	1	1.29	5.65	4.01	.56
		2	1.59	5.88	4.09	.54

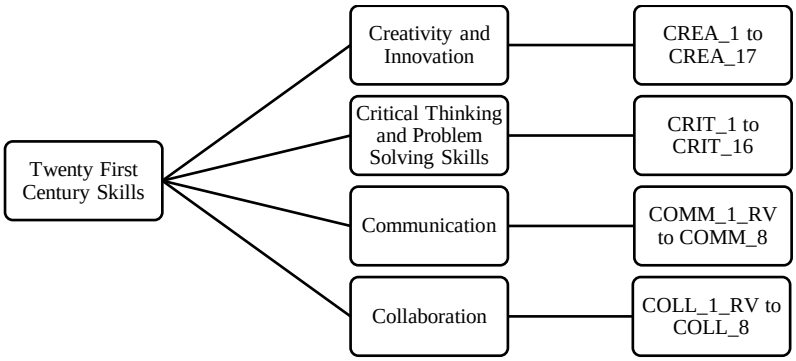
		3	1.00	5.82	4.22	.60
Critical thinking and problem solving skills	Experimental	1	3.75	5.25	4.56	.36
		2	3.50	5.31	4.45	.43
		3	3.81	5.38	4.62	.39
	Control	1	1.00	5.88	4.10	.53
		2	1.63	5.81	4.15	.53
		3	1.00	6.00	4.32	.59
Communication	Experimental	1	2.88	5.88	4.51	.70
		2	2.75	5.88	4.48	.68
		3	2.13	5.50	4.45	.79
	Control	1	1.00	5.88	4.25	.66
		2	1.00	6.00	4.24	.66
		3	1.00	6.00	4.41	.72
Collaboration	Experimental	1	3.29	5.86	4.80	.56
		2	3.57	5.86	4.65	.55
		3	3.57	5.86	4.75	.55
	Control	1	2.43	6.00	4.61	.60
		2	1.71	6.00	4.49	.61
		3	1.00	6.00	4.76	.61

4.2 Results of structural equation model

A four-subfactor-model was fitted where the four subscales represent latent variables with 7 to 17 items per subscale (see Figure 3 below).

Figure 3

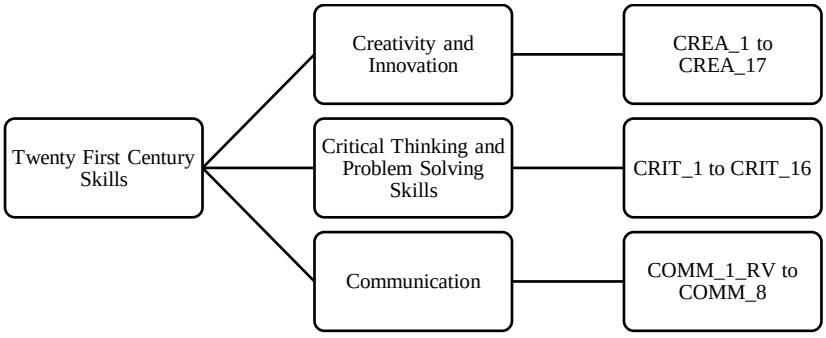
Twenty First Century Skills as a four-factor-model as tested in lavaan



Based on this analysis it became clear that, for this research and based on the collected data the collaboration subscale could not measure the construct ‘collaboration’ on individual student level well enough as part of Twenty First Century Skills. The model with the collaboration scale could not be fitted since the algorithm did not converge normally and the factor variance of this subscale was only 0.007. Therefore unreliable results were obtained. It was decided to run a factor model without the collaboration scale in further analyses (see Figure 4 below).

Figure 4

Twenty First Century Skills as a three-factor-model as tested in lavaan



Parameter estimates were obtained after 51 iterations using all 3019 observations. The model provided three widely-used fit indices: CFI, TLI and RMSEA. The CFI equalled .922, where a value larger than .9 indicates an acceptable level of fit (Smith & McMillan, 2001). The TLI equalled .917, a value at or above .9 is considered to indicate a good to excellent fit (Bentler & Bonnet, 1980). The RMSEA was estimated to be .090, with a 90% confidence interval from

.089 to .091., where the .090 RMSEA indicates mediocre fit (Byrne, 1998). The model thus provided good fit.

The latent factor scales all showed significant positive factor loadings, with standardized coefficients ranging from 0.94 to 1.32 (see Table 6). Factor loadings on item level can be found in appendix E, in order to estimate to what extent an individual item contributes towards one of the three different twenty first century subscales.

Table 6

Factor loadings 'Twenty First Century Skills' on scale level (with main latent variable)

Latent factor (scale)	Indicator	B (factor loading)	SE (standard error)	Z	Beta coefficient (standardized)	Sig P(>|z|)
Twenty First Century Skills	Creativity and innovation	1.00			.83	.00
	Critical thinking and problem solving skills	1.32	0.05	25.56	.93	.00
	Communication	0.94	0.04	24.90	.74	.00

4.3 Development of Twenty First Century Skills

4.3.1 Comparing the experimental and control group at the start of the intervention

Now that the TFCS model fitted well and latent factor scores per subscale were measured, the results of the evaluation of the first hypothesis can be presented. The hypothesis regarded that at the start of the intervention (T1) honors students have a higher TFCS score as compared to the control group. Normally an independent sample t-test would have been used, which works on the assumption of normally distributed data. However, the Shapiro-Wilk test was conducted and showed significant departure from normality for all four constructs. Creativity and innovation (C1) showed a significant departure from normality, W(1362) = .995, p=0.00. Critical thinking and problem solving skills (C2) showed a significant departure from

normality, W(1362) =.990, p= 0.00. Communication (C4) showed a significant departure from normality, W(1362) = .996, p=0.01. Lastly, TFCS as entire construct showed a significant departure from normality, W(1362) = .991, p=0.00.

In order to continue further comparative analysis taken into account the non-normality of the data, Mann-Whitney tests were conducted for the four constructs. The results, presented in Table 7 below, showed that the scores on all the four subscales were significantly greater for the experimental group than for the control group, indicating that these students scored higher on these subscales and TFCS as a whole at the start of the intervention. Therefore the first hypothesis is found to be true.

Table 7

Results of Mann-Whitney test per construct

	Creativity and innovation	Critical thinking and problem solving skills	Communication	TFCS
Median experimental group	0.33	4.85	0.31	2.71
Median control group	-0.07	-1.09	-0.05	-6.92
Mann-Whitney U	12311.00	10863.00	16150.00	11259.00
Wilcoxon W	889461.00	888013.00	893300.00	888409.00
Z	-5.37	-5.98	-3.77	-5.81
Asymp. Sig. (2-tailed)	.00	.00	.00	.00
Exact. Sig. (2-tailed)	.00	.00	.00	.00

R	-0.15	-0.16	-0.10	0.16

4.3.2 Comparing the experimental and control group over time

In order to find out how the experimental and control group develop their TFCS over time, an LME analysis (Bates et al., 2015) and Bayesian hypothesis testing of intraclass correlations were conducted. The dependent variable (estimated factor scores on the subscales) was assumed to be normally distributed given the average subscale factor levels and the measurement moments. It was assumed that for each of the TFCS subscales (4 C's) and TFCS as a whole, the experimental group would variate from the control group in the random intercept score, and that this intercept will vary across the measurement moments. A random intercept model was conducted where each of the four (sub)scales had his own intercept value. For each of the scales, if the random intercept variance was greater than zero, then the mean varied across the three measurement moments. The Bayesian hypothesis test with interclass correlations as parameter showed how probable the results are for the null-hypotheses. When the ICC is identical to zero, this indicates that there is no variance in the intercept over time. When the ICC is greater than zero, this indicates that there is variance over the means and thus over the three measurement moments. This was tested for all the four (sub)scales at the same time.

4.3.2.1 Experimental group. The LME analysis is provided with a method to test and estimate random effect variances. The random effect variance estimates were all around zero in four decimals. No random intercept variance was identified in the LME analysis, indicating that it is probable that the students did not change over time. LME was only used to estimate random effect variances but not to test whether or not the intercept means differed. Therefore, a Bayesian hypothesis test was conducted. Based on the results of the LME analysis the hypothesis was formed that the construct's ICCs are all equal to zero, whereas the second hypothesis is the complement. The results showed that for each (sub)scale, the ICCs were all equal to zero with 96.4% probability as stated by hypothesis 1 (see Mulder & Fox, 2019, for more interpretation). It can thus with almost 100% certainty be concluded that the ICCs of the constructs of the experimental group are equal to zero. As a result, the time-specific intercepts did not differ and it can be concluded that the experimental group does not develop their TFCS over time.

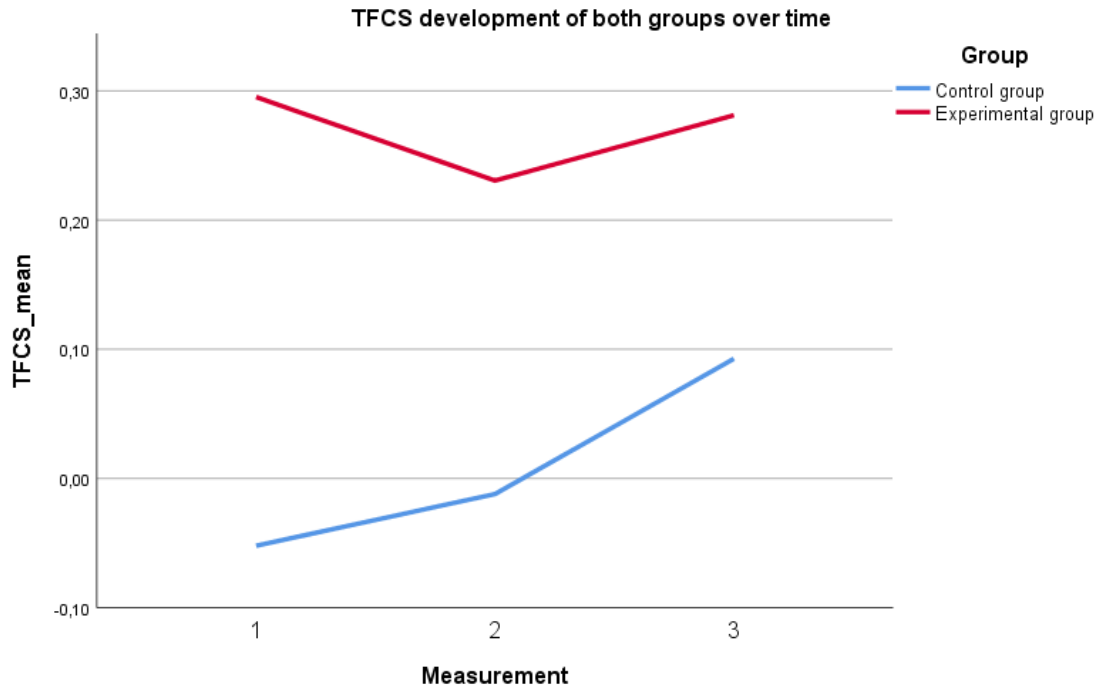
4.3.2.2 Control group. Also in the control group, the LME analysis is provided with a method to test and estimate random effect variances. The random effects showed variances in

the measurement moments over time on all the four (sub)scales. This indicates that it is probable that the students develop TFCS over time. Since LME is only used to estimate variances but not to test whether or not the random intercepts differ, a Bayesian hypothesis test was used again.

Based on the results of the LME analysis the hypothesis was formed that the ICCs of the four (sub)scales are all greater than zero, whereas the second hypothesis is the complement. The results of the Bayesian hypothesis test showed with 97.8% certainty that the ICCs are greater than zero, as stated by hypothesis 1 (see Mulder & Fox, 2019, for more interpretation). It can thus with almost 100% certainty be concluded that the average construct scores change over time for the control group. The experimental group did not develop their TFCS over time. A graph based on the estimated latent factor scores confirms this (see Figure 5 below), figures per subscale can be found in Appendix F.

Figure 5

TFCS development of both groups over time



4.3.2.3 Comparing the experimental and control group over time with school clustering. In order to test the three level growth model, an LME analysis was conducted for both groups including clustering as school. The LME analysis is provided with a method to test

and estimate random effect variances. For the experimental group, no variances were present amongst the random effects. This indicates that the honors group showed no development in TFCS, so schools also did not have influence on this.

For the control group, the random effects in the LME model showed variances in the measurement moments over time on all the (sub)scales, even though they are rather low (see Table 8). This indicates that it is probable that the schools have some influence in how students develop TFCS over time. It is not possible to exactly calculate the school influence. Normally, a Bayesian hypothesis test would be used, however the software of the BFpack (Mulder et al., 2019) cannot take into account the clustering in schools yet.

Table 8

Variances in schools of three level conditional growth model

Groups	Name	Variance	SD
School	(Intercept)	0.006	.08
Measurement	Creativity	0.007	.08
Measurement	Critical thinking	0.009	.09
Measurement	Communication	0.004	.06
Measurement	TFCS	0.004	.06
Residual		0.240	.49

That schools show some variance in the development over time can also be concluded from Figure 6 below (figures on scale level can be found in Appendix G). In this picture the few honors students per school were taken into account. In figure 7 only the non-honors students per school were incorporated. Some variance in school development was thus due to the honors students. All schools showed a pattern of decrease in number of participants, with the biggest gap in respondents per school during the last measurement. This was due to the fact that the schools were closed because of the Covid-19 outbreak and students had to fill in the

questionnaire from home. Only measurements per school with over ten participants were shown in the figures.

Figure 6

TFCS development of participating schools over time

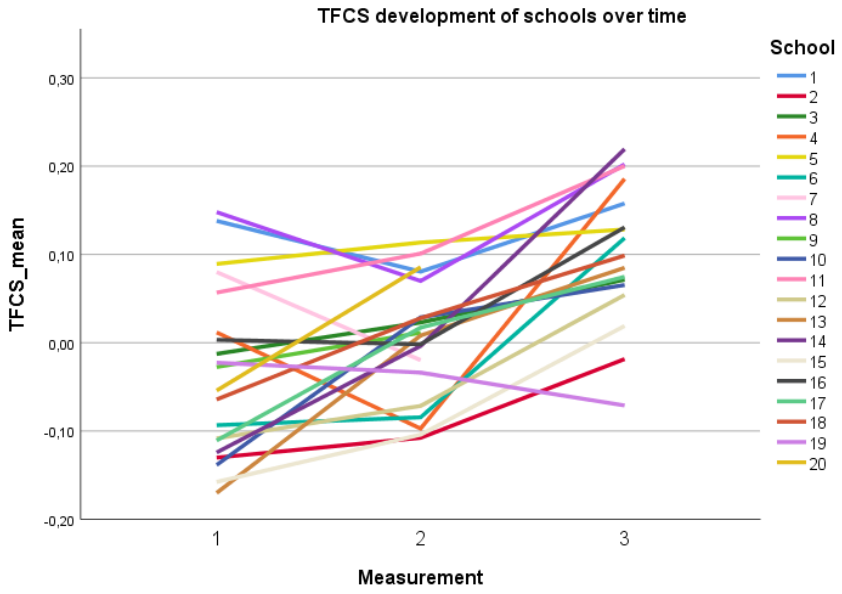
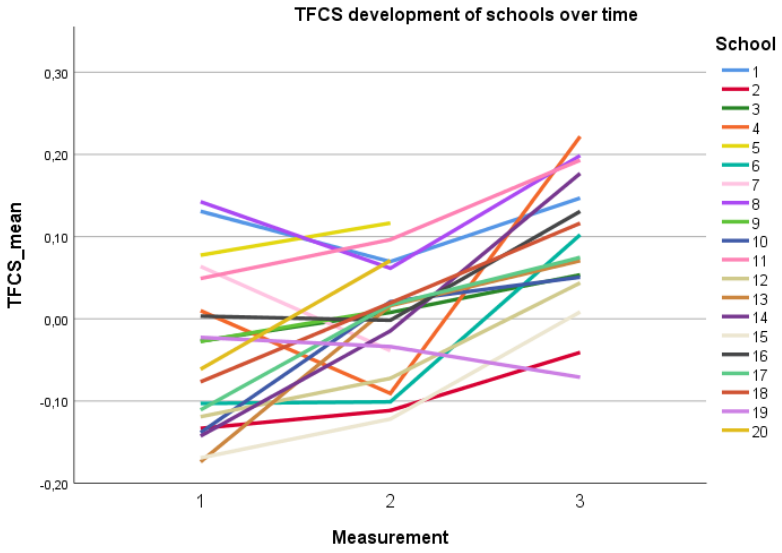


Figure 7

TFCS development of non-honors students in schools over time



4.3.3 Comparing the experimental and control group at the end of the intervention

The results above showed that there is no growth in TFCS in the experimental group. Since the experimental group did score significantly higher TFCS scores at the beginning of

the intervention, it was interesting to analyse whether these scores are still significantly higher at the end of the intervention. Therefore, another Shapiro-Wilk test was conducted. Just as with the data of the first measurement, the results showed significant departure from normality for all four constructs (see Table 9 below).

Table 9

Results of Shapiro-Wilk normality test

	Statistic	Df	Sig.
Creativity and innovation	.977	572	.000
Critical thinking and problem solving skills	.978	572	.000
Communication	.988	572	.000
Twenty First Century Skills	.976	572	.000

In order to continue further comparative analysis taken into account the non-normality of the data, Mann-Whitney tests were conducted for the four estimated constructs. The results, presented in Table 10 below, showed that the scores on TFCS in general, creativity and innovation (C1) and critical thinking and problem-solving skills (C2) were significantly greater for the experimental group than for the control group, indicating that these students scored higher on these constructs at the end of the intervention. In the case of communication (C4) a higher score is provided for the experimental group, but this higher score in communication was not significant.

Table 10

Results of Mann-Whitney test per construct

	Creativity and innovation	Critical thinking and problem solving skills	Communication	TFCS
Mann-Whitney U	7481.00	6691.00	8470.00	7024.00
Wilcoxon W	15086100	140071.00	151850.00	150404.00
Z	-2.49	-3.30	-1.47	-2.96
Asymp. Sig. (2-tailed)	.013	.001	.142	.003
Exact. Sig. (2-tailed)	.013	.001	.143	.003
R	-0.10	-0.14	-0.06	-0.12

5. Conclusion and discussion

5.1 Overview of the results

It appeared, based on the literature discussed in the introduction, that it is essential that students develop TFCS, but that learning these skills is currently insufficiently present in regular education. Since the focus on TFCS is visibly present in the Pre-U honors program, the goal of this study was to research the influence of following the extracurricular Pre-U honors program on the development of Twenty First Century Skills among 11th grade students.

Since the honors program focuses on enhancing TFCS but also selects its students, it was expected that at the start of following the honors program (the intervention) the honors students would have developed their TFCS to a higher extent than students following only regular education. Results of this study confirmed indeed that this was the case, taken into account that the "collaboration" scale was not included.

It was expected that students in the honors program would develop their TFCS to a higher extent than students who only followed regular education. The results showed that at the end of the intervention, honors students still had a higher TFCS score than non-honors students. However, the results showed that based on the used questionnaire and scales, students in the honors program did not develop their TFCS during the honors program, whilst students following regular education do develop their TFCS.

5.2 Discussion of results

Below the two main findings of this study are discussed, including their implications and suggestions for future research. It will start with the implications of the comparison of honors and non-honors students at the beginning of the intervention. Next, the focus will lie on the development of both groups over time and the comparison of the students at the end of the research. The role of self-assessment of the research, the role of collaboration as part of TFCS, the TFCS questionnaire and the development of both groups are therefore discussed below.

5.2.1 Comparing the honors and non-honors students at the start of the intervention

Since honors students were specifically selected to participate in the program and thus to some extent were open to develop themselves, it was expected that at the start of following the honors program (the intervention) the honors students would have developed their TFCS to a higher extent than students following only regular education. The current study confirms this.

Although the student should mainly want to develop themselves (and therefore not necessarily have higher TFCS scores), students who are already further in their TFCS development are selected according to the questionnaire. A reason for this can be the influence that many of the schools where the honors students come from have on the selection procedure. Schools conduct their own pre-selection or make demands on their students based on the grades obtained in order to be allowed to register for the Pre-U selection procedure at all (Pre-U, 2019c). The pre-selection of schools therefore also varies per school. Hypothetically, it could therefore be that schools pre-select students who score high in, for example, creativity and critical thinking.

Another reason can be a selection bias in the Pre-U selection committee. Even though students should mainly want to develop themselves, it might be that the selection committee selected students that were further in their TFCS development. For instance a student who did really well on the selection assignment "present yourself in an alternative way". A 'good' assignment in the eyes of the teacher could be a student who presented him or herself really creatively and originally, but in that case the student thus would already have developed their creativity and innovation skills to a further level. Future research should focus on improving the selection procedure on schools and in the Pre-U selection committee in order to give all motivated students an equal chance to participate in the honors program.

5.2.2 Comparing the honors and non-honors students over time

The results of this research showed that honors students do not develop their TFCS over time, whereas non-honors students do develop their TFCS to some extent and the school of the student plays, to some extent, a role in this development. Explanations of these findings are discussed below.

5.2.2.1 TFCS questionnaire. The results showed that in general, more than 75% of the students agreed with the provided statements. The first reason for this can be response bias due to lack of motivation when students filled in the questionnaire. The reversed items provide an indication for this since these statements showed the most spread and variation amongst the answers. When the reversed items were reversed back to normal, the data showed less spread amongst the answers. The second reason can be that the questions were framed 'too easy' and students therefore almost always agreed with the statements. The scales thus probably were not discriminating enough. Not all statements of the subscales functioned and contributed equally well. Even though the reliability of the scales could be increased by deleting 'insufficiently

performing statements' this choice was not made since this would decrease the validity of the scales. All items were based on specific criteria of that skill (P21, 2019), so if questions would have been deleted, the whole skill would not have been measured as it was defined by P21 (2019).

Even though the questionnaire was not yet perfect, up until now TFCS could not be assessed so the need to develop an instrument was therefore high (World Economic Forum, 2015). In this study a self-assessment questionnaire to assess TFCS was developed based on existing observation tools for teachers which uses the official definitions of the P21-model of the skills (CTE – Washington, 2019). The statements in our questionnaire were thus based on the official definitions of the four TFCS. After the questionnaire was developed, we tested the instrument in the same research amongst a large population at three moments, which lead to the questionnaire being filled out over three thousand times. Three of the four TFCS constructs contributed equally to TFCS. These points indicate that we did make a successful start in developing the questionnaire and contributed to current TFCS research, which are strong points of this research. Future research should logically be focused on further development of the four constructs in the questionnaire since it is crucial that measures for these essential skills can be developed in the future (World Economic Forum, 2015).

5.2.2.2 Collaboration as part of TFCS. The collaboration scale showed almost no dispersion at all and the TFCS model could not converge with collaboration as part of TFCS so it would provide no reliable results. Therefore the items of the collaboration scale were deleted from further analysis. This is not in line with the research of Germaine et al. (2016) who claimed that the four skills are connected to and influence each other. A reason for this lack of dispersion can be that nowadays, project-based learning in which students need to collaborate closely together is implemented in all forms of education (Lee, Huh & Reigeluth, 2015). Therefore, students in 11th grade might already be able to collaborate and therefore showed no dispersion.

Even though the TFCS model with collaboration could not be fitted on the data, it cannot be stated that collaboration is not a part of TFCS. The tested TFCS model was fitted by confirmatory factor analysis. In this analysis, the theory behind the model is crucial and helps deciding if the model is any good (Understanding Data, 2017). The theory discussed in the introduction and theoretical framework clearly present that collaboration is an essential part of TFCS (P21, 2019; Thijs et al., 2014). Therefore it can only be concluded that for this specific research, collaboration as construct and as part of TFCS could not be measured.

Since the dispersion on all the TFCS subscales were low, future research should focus on further development of the questionnaire in which the items can be adjusted in a way it still matches the theory but it can better discriminate students on levels of creativity and innovation C1); critical thinking and problem-solving skills (C2); collaboration (C3); and communication (C4).

5.2.2.3 TFCS development of non-honors students. The results of this research showed that non-honors students do develop their TFCS to some extent and the school of the student plays, to some extent, a role in this development. In the introduction it was discussed that students do not learn these essential skills enough at school to properly function in the global workplace of the future (Germaine et al., 2016). Even though the school only plays to some extent a role in TFCS development, this school variance must be acknowledged. At some schools, students develop their TFCS to a larger extent than at other schools. In order to find out how these school effects develop, it is interesting for future research to follow students for longer period of time, to compare education offers and teaching methods at schools, and to take background characteristics of students into account.

5.2.2.4 TFCS development of honors students. The results of this study showed that honors students do not develop their TFCS over time. This could indicate that the intervention was not successful, but based on the theory described in the introduction that is not expected. Furthermore, the high TFCS scores also do not provide indications for that. In the paragraphs above, several limitations regarding the TFCS questionnaire were discussed. It is therefore more probable that no results were found since the questionnaire was not levelled and scaled with high enough ranking options for honors students. A clear ceiling effect is also visible in the results of the honors group. It is therefore important for future research to further develop the TFCS questionnaire and take a close look at the answering options in order for the level of the answers to better fit the honors students.

Even though the results of this research showed that honors students do not develop their TFCS over time, they do score significantly higher on TFCS at the beginning and end of the intervention. The raw data output of the questionnaire also showed that the honors students scored quite a lot higher on the questionnaire and their minimum score was also severely higher than that of the control group. This all indicates that honors students do have higher levels of TFCS, but this study could not show that honors students also developed their TFCS. It was already discussed above that future research should focus on further adapting the TFCS

questionnaire. Once a reliable and valid assessment scale for TFCS exist, the research should be repeated, and even the entire one-and-a-half year honors intervention can be taken into account.

The literature described in the theoretical framework suggested that the honors program influences TFCS and high TFCS scores were found at the beginning and end of the intervention. Since it is very likely that the lack of development was based on a questionnaire that was not levelled and scaled with high enough ranking options for honors students, it cannot be stated yet that the Pre-U honors program thus does not contribute to TFCS. Future research should, as mentioned above, thus repeat the questionnaire once more in both groups once it is adapted.

Furthermore, a deeper analysis of the Pre-U honors program and its link to TFCS could help to develop the improved TFCS questionnaire and can assist in determining a level of TFCS in honors students. In order to do so, research can be done at policy- teacher- and student level. At a policy level, documents can be analyzed about the meaning and usefulness of this honors program with regard to teaching TFCS, and how regular schools approach this. In addition, it is relevant to interview the management of Pre-U and the regular schools regarding the objectives, what the program aims to achieve, et cetera. Subsequently, with the support of the teachers, it is relevant to look at all the information that is available to students: preparation assignments, PowerPoint presentations and it is useful to conduct further interviews with teachers about how their lessons and overall program now connect concretely with teaching and learning TFCS. This all in order to further determine the TFCS level and TFCS teaching of honors students. Lastly, it is relevant to conduct research at student level. Documents, motivation and application letters and portfolio assignments can be analyzed to look for TFCS development. All in all, it cannot be concluded yet that the Pre-U honors program does not contribute significantly towards learning TFCS. The above mentioned suggestions can make sure a definitive answer is provided.

5.3 Conclusion

In order for students to be prepared for the global workplace of the future, it is essential that students develop key TFCS (Germaine et al., 2016). Earlier studies showed that students learn these skills insufficiently and thus have not mastered these skills when they go into the job market (Saavedra & Opfer, 2012; World Economic Forum, 2015; WRR, 2014). The Pre-U honors program, which specifically focuses on developing TFCS, was seen as a possible solution to this problem. This current study made a good start in investigating other methods

and contents of teaching than current education in order to see if students can benefit from this by learning more TFCS. Next to that, the study contributed to science and practice by making a start at the development of a TFCS self-assessment questionnaire.

This research proved that students in the Pre-U honors program had significantly higher levels of TFCS than non-honors students at the beginning and end of the intervention. However, it could not be proven that the students also develop their TFCS further because of following the Pre-U honors program. This is most likely due to some limitations in the developed questionnaire. The results however do not prove the contrary of our expectations, that the honors program does not contribute to learning TFCS. Therefore, several suggestions were presented for future research regarding improvement of the questionnaire and further research into comparing honors and non-honors students.

Reference list

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... & Bolker, M. B. (2015). Package 'lme4'. *Convergence*, *12*(1), 2.

Bentler, P. M., & Bonett, D. G. (1980). Significance test and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*(3), 588-606.

Breetveld, I., Meijer, J. & and Koopman, P. (2012). *Effectonderzoek VWO-plus 'Onderzoek naar effecten van het Athaneum-plus programma van College De Heemlanden'*. Amsterdam, Nederland: Kohnstamm Instituut.

Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. New Jersey, USA: Lawrence Erlbaum Associates, Publishers.

Covey, S. (1989). *The seven habits of highly effective people*. New York, USA: Simon & Schuster.

CTE - Washington (2019). *21st Century Skills Standard Rubric.* Retrieved from https://www.k12.wa.us/student-success/career-technical-education/career-technical-education-pathways/21st-century-skills.

Gardner, H. (2007). *Five minds for the future*. Boston, USA: Harvard Business School.

Germaine, R., Richards, J., Koeller, M., & Schubert-Irastorza, C. (2016). Purposeful Use of 21st Century Skills in Higher Education. *Journal of Research in Innovative Teaching*, *9*(1).

Greiff, S., & Kyllonen, P. (2016). Contemporary Assessment Challenges: The Measurement of 21st Century Skills. *Applied Measurement in Education*, *29*(4), 243-244. https://doi.org/10.1080/08957347.2016.1209209

Gutman, L. M., & Schoon, I. (2013). The impact of non-cognitive skills on outcomes for young people. *Education Endowment Foundation*, *59*(22.2), 2019.

Kaufman, K. J. (2013). 21 ways to 21st century skills: why students need them and ideas for practical implementation. *Kappa Delta Pi Record*, *49*(2), 78-83.

Lamb, S., Jackson, J., & Rumberger, R. (2015). *ISCY technical paper: Measuring 21st century skills in ISCY*. Retrieved from http://vuir.vu.edu.au/31682/1/ISCY%2021st%20Century%20Skills%20Framework.pdf.

Larson, L., & Miller, T. (2011) 21st Century Skills: Prepare Students for the Future, *Kappa Delta Pi Record*, *47*(3), 121-123. https://doi.org/10.1080/00228958.2011.10516575

Ledoux, G., Meijer, J., Van der Veen, I., Breetvelt, I., Ten Dam, G., & Volman, M. (2013). *Meetinstrumenten voor sociale competenties, metacognitie en advanced skills. Een inventarisatie*. Amsterdam, The Netherlands: Kohnstamm Instituut.

Lee, D., Huh, Y., & Reigeluth, C. M. (2015). Collaboration, intragroup conflict, and social skills in project-based learning. *Instructional science*, *43*(5), 561-590.

Manpower Group (2014). *The Talent Shortage Continues: How the Ever Changing Role of HR Can Bridge the Gap.* Retrieved from: http://web.archive.org/web/20151029214731/http://www.manpowergroup.us/campaigns/talent-shortage-2014/assets/pdf/2014_Talent_Shortage_WP_US.pdf

Michels, B. I., & Eijkelhof, H. (2018). High-school students engaging with researchers within a pre-university programme: Motivations and experiences. *Research for All*, *2*(1), 43-61.

Mulder, J., & Fox, J. P. (2019). Bayes factor testing of multiple intraclass correlations. *Bayesian Analysis*, *14*(2), 521-552.

Mulder, J., Gu, X., Olsson-Collentine, A., Tomarken, A., Böing-Messing, F., Hoijtink, H., ... & van Lissa, C. (2019). BFpack: Flexible Bayes Factor Testing of Scientific Theories in R. *arXiv preprint arXiv:1911.07728*.

P21: Partnership for 21st Century Skills. (2019). *Framework for 21st century learning definitions*. Retrieved from http://static.battelleforkids.org/documents/p21/P21_Framework_DefinitionsBFK.pdf

Padget, S. (2012). *Creativity and critical thinking*. New York, USA: Routledge.

Pre-U (2019a). *Pre-U Honoursprogramma.* Retrieved from https://www.utwente.nl/onderwijs/pre-university/pre-u/honoursprogramma/

Pre-U (2019b). *Pre-U Honoursprogramma.* Retrieved from https://www.utwente.nl/onderwijs/pre-university/pre-/honoursprogramma/aanmelding-en-selectie/voor-wie/

Pre-U (2019c). *Informatiepakket aanmelden Pre-U Honoursprogramma lichting 19-21*. Retrieved internally from Pre-U.

Pre-U (2019d). *Handboek Pre-U Honoursprogramma lichting 19-21*. Retrieved internally from Pre-U.

Pre-U (2020a). *Aanbod Pre-U.* Retrieved from https://www.utwente.nl/onderwijs/pre-university/pre-u/

Pre-U (2020b). *Pre-U Partnerscholen.* Retrieved from https://www.utwente.nl/onderwijs/pre-university/over-pre-u/partnerscholen/

R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

Renzulli, J. S. (2012). Reexamining the role of gifted education and talent development for the 21st century: A four-part theoretical approach. *Gifted Child Quarterly*, *56*(3), 150-159.

Rijksoverheid [Netherlands]. (2014a). *Plan van aanpak toptalenten 2014 – 2018*. Retrieved from: https://www.rijksoverheid.nl/documenten/kamerstukken/2014/03/10/plan-van-aanpak-toptalenten-2014-2018

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *Journal of statistical software*, *48*(2), 1-36.

Rotherham, A. J., & Willingham, D. (2009). 21st century. *Educational leadership*, *67*(1), 16-21.

Saavedra, A. R., & Opfer, V. D. (2012). *Teaching and learning 21st century skills: Lessons from the learning sciences.* New York, USA: Asia Society.

Salmons, J. (2019). *Learning to Collaborate, Collaborating to Learn: Engaging Students in the Classroom and Online*. Sterling, USA: Stylus Publishing, LLC.

Samuel, D. F. (2019). Critical Thinking in Science and Technology: Importance, Rationale, and Strategies. In *Handbook of Research on Critical Thinking and Teacher Education Pedagogy* (pp. 177-193). https://doi.org/10.4018/978-1-5225-7829-1.ch010

Segers, E., & Hoogeveen, L. (2012). *Programmeringsstudie Excellentieonderzoek in primair, voortgezet en hoger onderwijs*. Nijmegen, the Netherlands: Behavioural Science Institute & Centrum voor Begaafdheids Onderzoek/Radboud Universiteit Nijmegen.

Silva, E. (2009). Measuring skills for 21st-century learning. *Phi Delta Kappan, 90*(9), 630-634.

Sirius expertcommissie. (2014). *Sirius Programma overall auditrapport 2013*. Retrieved from: https://www.siriusprogramma.nl/media/files/publicaties/overall-audit-rapport-v42013.pdf

Smith, T. D., & McMillan, B. F. (2001). *A Primer of Model Fit Indices in Structural Equation Modeling*. Retrieved from: https://files.eric.ed.gov/fulltext/ED449231.pdf.

Snyder, L. G., & Snyder, M. J. (2008). Teaching critical thinking and problem solving skills. *The Journal of Research in Business Education, 50*(2), 90.

Stichting het Zelfstandig Gymnasium (2020). *Honours Programma Gymnasia (HPG).* Retrieved from: https://www.gymnasia.nl/gymnasium-cum-laude

ten Cate, A.A., Niemeijer M. & Bijman, M.N.A. (2020, April 25). *Personal communication*.

Thijs, A., Fisser, P., & van der Hoeven, M. (2014). *21e eeuwse vaardigheden in het curriculum van het funderend onderwijs*. Enschede, the Netherlands: SLO. Retrieved from: http://downloads.slo.nl/Repository/21e-eeuwse-vaardigheden-conceptueel-kader.pdf

Tigchelaar, M. (2015). *Lezen, weten en niet vergeten*. Amsterdam, the Netherlands: Spectrum.

Trilling, B., & Fadel, C. (2009). *21st century skills for life in our times*. San Francisco, USA: Jossey-Bass

Understanding Data. (2017). *CFA in Lavaan – Writeup.* Retrieved from http://www.understandingdata.net/2017/03/22/cfa-in-lavaan/#writeup

Wagner, T. (2008). *The global achievement gap*. New York, USA: Basic Books.

Watson, G. (1980). *Watson-Glaser critical thinking appraisal*. San Antonio, USA: Psychological Corporation.

World Economic Forum (2015). *New vision for education: Unlocking the potential of technology*. Vancouver, Canada: British Columbia Teachers' Federation.

Willingham, D. T. (2008). Critical thinking: Why is it so hard to teach?. *Arts Education Policy Review*, *109*(4), 21-32.

Wolfensberger, M. V. (2015). *The Netherlands: Focus on excellence, honors programs all around. In Talent development in European higher education* (pp. 43-75). Cham, Switzerland: Springer International Publishing

Wolfensberger, M. V. C., Van Eijl, P., & Pilot, A. (2012). Laboratories for educational innovation: Honors programs in the Netherlands. *Journal of the National Collegiate Honors Council*, *13*(2).,149–170.

Wolfensberger, M., & Pilot, A. (2014). Uitdagingen voor onderzoek naar honoursonderwijs. *Tijdschrift voor hoger onderwijs*, *4*(1), 124-137.

WRR (2014). *Naar een lerende economie: Investeren in het inverdienvermogen van Nederland*. Den Haag, Nederland: WRR.

Appendix

Appendix A: Self-assessment questionnaire for Twenty First Century Skills (in Dutch)

Hieronder staan een aantal beweringen. Het is de bedoeling dat je je inleeft in de bewering en je aangeeft in hoeverre de bewering van toepassing is op jou. De beweringen gaan over hoe jij de dingen ziet en ervaart in je leven, dus niet specifiek alleen over je schoolwerk/schoolleven. Het is de bedoeling dat je het eerste antwoord wat in je opkomt aanklikt. Je kunt de beweringen een score geven van helemaal mee oneens tot en met helemaal mee eens. Er is geen goed of fout antwoord, het gaat erom welke bewering het beste bij jou past.

Creativity and Innovation

1. Ik gebruik verschillende creatieve technieken (zoals brainstormen) om tot nieuwe ideeën te komen
2. Ik ben goed in het verbeteren van bestaande concepten
3. Ik kom regelmatig met vernieuwende ideeën die niet voor de hand liggen (out of the box zijn)
4. Ik bedenk graag nieuwe ideeën en zoek daarvoor ook naar nieuwe kennis en bronnen
5. Ik analyseer en evalueer regelmatig mijn eigen ideeën
6. Ik slaag erin om mijn ideeën te verbeteren door ze te verfijnen en evalueren
7. REVERSED Ik vind het lastig om een nieuwe idee concreet te maken en gedetailleerd te beschrijven
8. REVERSED Het lukt me vaak niet om een nieuw idee praktisch tot uitvoering te brengen
9. Ik kan mijn ideeën goed aan anderen uitleggen
10. Ik vraag graag naar de ideeën van anderen om een probleem op te lossen
11. Ik verwerk de input van anderen graag in mijn ideeën en de uitwerking daarvan
12. Het lukt me om een idee (wat misschien abstract is) om te zetten naar een haalbare uitwerking
13. Mijn ideeën en de uitwerkingen daarvan bevatten vaak originele elementen
14. Ik probeer zo min mogelijk fouten te maken
15. Ik accepteer alle fouten die ik maak want die gebruik ik om mijn werk te verbeteren
16. Ik vind dat het een proces is van vallen en opstaan als ik nieuwe ideeën ontwikkel, toepas en uitwerk
17. Als ik een creatief idee heb zorg ik dat dat het in de praktijk uitgevoerd wordt

Critical Thinking and Problem Solving

18. Ik ben in staat om uit verschillende waarnemingen en gegevens een logische conclusie te trekken
19. Ik kan algemene kennis toepassen om in specifieke situaties een oplossing te verzinnen
20. Ik probeer een probleem of vraagstuk altijd te zien als een onderdeel van een groter geheel
21. Complexe vraagstukken breek ik op in eenvoudigere deelproblemen
22. Ik kan goed beslissingen nemen
23. REVERSED Ik vind het moeilijk om ergens een genuanceerd oordeel over te vellen
24. Ik ben in staat om bewijzen, argumenten, beweringen en overtuigingen kritisch te analyseren
25. Ik analyseer en evalueer graag perspectieven die heel anders zijn dan dat van mijzelf
26. Ik vind het nuttig om perspectieven die anders zijn dan mijn eigen perspectief kritisch te bekijken
27. Ik kan informatie en argumenten met elkaar in verband brengen om een standpunt te onderbouwen
28. REVERSED Ik vind het moeilijk om complexe informatie te analyseren en interpreteren en daar conclusies uit trekken
29. Ik reflecteer op leerervaringen en leerprocessen
30. Ik kan een probleem dat ik nog niet eerder ben tegengekomen oplossen door gebruik te maken van kennis en/of middelen die ik al heb
31. Ik kan een probleem dat ik nog niet eerder ben tegengekomen oplossen op een vernieuwende manier
32. Ik kan kritische vragen stellen om tot een betere oplossing van een probleem te komen
33. Ik kan kritische vragen stellen om verschillende perspectieven beter te begrijpen

Communication

34. REVERSED Ik vind het moeilijk in een gesprek mijn ideeën over te brengen
35. Ik ben altijd goed in staat om mijn ideeën schriftelijk over te brengen
36. Ik kan goed actief luisteren en haal er vaak meer informatie dan alleen wat wordt gezegd
37. Ik kan effectief communiceren om anderen te informeren en instrueren
38. Ik kan effectief communiceren om anderen te motiveren en overtuigen
39. Ik kan voor elke situatie goed bedenken op welke manier (bijvoorbeeld schriftelijk, mondeling of digitaal) ik mijn boodschap zo goed mogelijk over kan brengen

40. Ik kan bepalen wat de meest effectieve manier is om een boodschap digitaal over te brengen
41. Ik kan altijd goed verbaal en non-verbaal communiceren met anderen, in verschillende omgevingen (denk aan school, sport, bijbaantje)

Collaboration

42. REVERSED Ik werk liever alleen dan in een groep
43. Ik werk beter in een groep dan alleen
44. Ik kan goed en respectvol samenwerken, ook als de verschillen in onze groep groot zijn
45. In een groep ben ik bereid om een compromis te sluiten om een gemeenschappelijk doel te bereiken, ook als die oplossing niet mijn eerste voorkeur heeft
46. In groepswerk voel ik mij medeverantwoordelijk voor het eindproduct
47. Ik vind dat bij groepswerk de hele groep verantwoordelijk is voor het volledige eindproduct
48. Als ik in een groep werk waardeer ik de (individuele) bijdrage van anderen

Appendix B: Informed consent parents for students below the age of 16 (in Dutch)

Beste ouder(s)/verzorger(s),

Dit jaar zal de school van uw zoon/dochter meedoen aan een onderzoek van de Universiteit Twente naar de ontwikkeling van 21e eeuwse vaardigheden bij leerlingen in 5 VWO.

Alle leerlingen uit 5 VWO wordt gevraagd om in totaal 3x een vragenlijst in te vullen: in september, december en april. Het invullen van de vragenlijst gebeurt tijdens de les. Deelnemen aan dit onderzoek brengt geen risico's met zich mee. Alle gegevens zullen anoniem verwerkt worden.

In Nederland hebben kinderen onder de 16 jaar toestemming nodig van hun ouders om deel te nemen aan onderzoek. Daarom sturen we u dit toestemmingsformulier. Wanneer u hiermee toestemming geeft mag uw kind alle drie de keren deelnemen aan de vragenlijst. Uw kind mag te allen tijden stoppen met zijn/haar deelname aan dit onderzoek.

Wanneer uw kind mag deelnemen heeft de contactpersoon voor dit onderzoek, [naam contactpersoon], uiterlijk maandag 9 september dit ingevulde toestemmingsformulier nodig.

Mocht u vragen hebben dan kunt u contact opnemen met [naam onderzoeker] van de Universiteit Twente via [mail adres onderzoeker].

Met vriendelijke groet,
[Naam onderzoeker]

✂✂✂✂

Via deze weg geef ik toestemming dat mijn kind, _____, deel mag nemen aan het onderzoek naar de ontwikkeling van 21e eeuwse vaardigheden.

---------- ---------- -----/-----/2019

Naam ouder/verzorger Handtekening Datum van ondertekening

Dit formulier mag worden ingeleverd bij [naam contactpersoon], uiterlijk 9 september 2019.

Appendix C: R codes

R Code to build structural equation model

```
# load the lavaan package (only needed once per session)

library (lavaan)

#specify tfcs.model

tfcs.model <- '

creativity =~ CREA_1 + CREA_2 + CREA_3 + CREA_4 + CREA_5 + CREA_6 + CREA_7 + CREA_8 + CREA_9 + CREA_10 + CREA_11 + CREA_12 + CREA_13 + CREA_14 + CREA_15 + CREA_16 + CREA_17

critical_thinking =~ CRIT_1 + CRIT_2 + CRIT_3 + CRIT_4 + CRIT _5 + CRIT_6 + CRIT_7 + CRIT_8 + CRIT_9 + CRIT_10 + CRIT_11 + CRIT_12 + CRIT_13 + CRIT_14 + CRIT_15 + CRIT_16

communication =~ COMM_2 + COMM_1 + COMM_3 + COMM_4 + COMM_5 + COMM_6 + COMM_7 + COMM_8

tfcs =~ creativity + critical_thinking + communication'

#fit tfcs.model with categorical variables

fit_tfcs.model <- cfa(tfcs.model, data = df_est_data, ordered = c ("CREA_1", "CREA_2", "CREA_3", "CREA_4", "CREA_5", "CREA_6", "CREA_7",
                                                                   "CREA_8", "CREA_9", "CREA_10", "CREA_11", "CREA_12", "CREA_13", "CREA_14",
                                                                   "CREA_15", "CREA_16", "CREA_17",
                                                                   "CRIT_1", "CRIT_2", "CRIT_3", "CRIT_4", "CRIT_5", "CRIT_6", "CRIT_7",
                                                                   "CRIT_8", "CRIT_9", "CRIT_10", "CRIT_11", "CRIT_12", "CRIT_13", "CRIT_14",
                                                                   "CRIT_15", "CRIT_16",
```

"COMM_1", "COMM_2", "COMM_3", "COMM_4", "COMM_5", "COMM_6", "COMM_7", "COMM_8"))

#display summary output tfcs

summary(fit_tfcs.model, fit.measures = TRUE, standardized=TRUE)

R Code to calculate latent factor scores and make a new dataset

#Predict factor scores for latent variables

predict_tfcs.model <- predict(fit_tfcs.model)

#Check predictions

head(predict_tfcs.model)

#Create new dataframe with all data

newdata <- data.frame(predict_tfcs.model,df_est_data)

Appendix D: Descriptive statistics for each measurement on item level

Descriptive statistics measurement 1 on item level

Table 1

Descriptive Statistics for Observed Variables 'Twenty First Century Skills' during measurement 1, N=1362

Latent factor (scale)	Variable	Min	Max	Mean	SD	25%	50%	75%
Creativity and innovation	CREA_1	1.00	6.00	3.64	1.27	3.00	4.00	5.00
	CREA_2	1.00	6.00	3.86	1.05	3.00	4.00	5.00
	CREA_3	1.00	6.00	3.75	1.17	3.00	4.00	5.00
	CREA_4	1.00	6.00	3.79	1.21	3.00	4.00	5.00
	CREA_5	1.00	6.00	3.76	1.21	3.00	4.00	5.00
	CREA_6	1.00	6.00	3.86	1.10	3.00	4.00	5.00
	CREA_7	1.00	6.00	3.63	1.20	4.00	5.00	5.00
	CREA_8	1.00	6.00	3.76	1.18	4.00	5.00	5.00
	CREA_9	1.00	6.00	4.45	1.06	4.00	4.00	5.00
	CREA_10	1.00	6.00	4.37	1.09	3.00	4.00	5.00
	CREA_11	1.00	6.00	4.28	1.02	3.00	4.00	5.00
	CREA_12	1.00	6.00	3.89	0.98	5.00	5.00	6.00
	CREA_13	1.00	6.00	3.94	1.01	4.00	4.00	5.00
	CREA_14	1.00	6.00	5.14	0.96	4.00	5.00	5.00
	CREA_15	1.00	6.00	4.25	1.17	3.00	4.00	4.00
	CREA_16	1.00	6.00	4.43	1.04	4.00	5.00	5.00

	CREA_17	1.00	6.00	3.60	1.10	4.00	5.00	5.00
Critical thinking and problem solving skills	CRIT_1	1.00	6.00	4.59	0.88	3.00	4.00	5.00
	CRIT_2	1.00	6.00	4.58	0.88	3.00	4.00	5.00
	CRIT_3	1.00	6.00	3.80	1.10	3.00	4.00	5.00
	CRIT_4	1.00	6.00	3.81	1.12	3.00	4.00	5.00
	CRIT_5	1.00	6.00	4.08	1.32	4.00	5.00	5.00
	CRIT_6	1.00	6.00	3.69	1.18	3.00	4.00	5.00
	CRIT_7	1.00	6.00	4.40	0.96	4.00	4.00	5.00
	CRIT_8	1.00	6.00	3.90	1.22	4.00	5.00	5.00
	CRIT_9	1.00	6.00	4.27	1.05	3.00	4.00	5.00
	CRIT_10	1.00	6.00	4.54	0.87	3.00	4.00	5.00
	CRIT_11	1.00	6.00	3.65	1.13	4.00	4.00	5.00
	CRIT_12	1.00	6.00	3.97	1.08	3.00	4.00	4.00
	CRIT_13	1.00	6.00	4.34	0.86	4.00	4.00	5.00
	CRIT_14	1.00	6.00	3.88	0.96	4.00	4.00	5.00
	CRIT_15	1.00	6.00	4.15	1.01	3.00	5.00	5.00
	CRIT_16	1.00	6.00	4.21	0.99	3.00	4.00	5.00
Communication	COMM_1	1.00	6.00	4.22	1.27	4.00	4.00	5.00
	COMM_2	1.00	6.00	4.11	1.12	4.00	4.00	5.00
	COMM_3	1.00	6.00	4.12	1.08	4.00	5.00	5.00
	COMM_4	1.00	6.00	4.36	1.00	4.00	4.00	5.00
	COMM_5	1.00	6.00	4.40	1.01	4.00	4.00	5.00

	COMM_6	1.00	6.00	4.17	1.02	4.00	5.00	5.00
	COMM_7	1.00	6.00	4.14	0.99	2.00	3.00	5.00
	COMM_8	1.00	6.00	4.53	1.10	3.00	4.00	5.00
Collaboration	COLL_1	1.00	6.00	3.49	1.42	4.00	5.00	5.00
	COLL_2	1.00	6.00	3.64	1.23	4.00	5.00	5.00
	COLL_3	1.00	6.00	4.77	0.94	5.00	5.00	6.00
	COLL_4	1.00	6.00	4.67	0.95	5.00	6.00	6.00
	COLL_5	1.00	6.00	5.28	0.85	5.00	5.00	6.00
	COLL_6	1.00	6.00	5.31	0.91	3.00	4.00	5.00
	COLL_7	1.00	6.00	5.14	0.82	3.00	4.00	5.00

Descriptive statistics measurement 2 on item level

Table 2

Descriptive Statistics for Observed Variables 'Twenty First Century Skills' during measurement 2, N=1083

Latent factor (scale)	Variable	Min	Max	Mean	SD	25%	50%	75%
Creativity and innovation	CREA_1	1.00	6.00	3.77	1.22	3.00	4.00	5.00
	CREA_2	1.00	6.00	4.04	0.94	4.00	4.00	5.00
	CREA_3	1.00	6.00	3.87	1.10	3.00	4.00	5.00
	CREA_4	1.00	6.00	3.95	1.10	3.00	4.00	5.00
	CREA_5	1.00	6.00	3.95	1.11	3.00	4.00	5.00
	CREA_6	1.00	6.00	4.07	1.01	4.00	4.00	5.00

	CREA_7	1.00	6.00	3.54	1.18	3.00	3.00	5.00
	CREA_8	1.00	6.00	3.67	1.19	3.00	4.00	5.00
	CREA_9	1.00	6.00	4.40	1.10	4.00	5.00	5.00
	CREA_10	1.00	6.00	4.38	1.02	4.00	5.00	5.00
	CREA_11	1.00	6.00	4.35	0.96	4.00	4.00	5.00
	CREA_12	1.00	6.00	3.99	0.93	4.00	4.00	5.00
	CREA_13	1.00	6.00	4.11	0.95	4.00	4.00	5.00
	CREA_14	1.00	6.00	5.11	1.00	5.00	5.00	6.00
	CREA_15	1.00	6.00	4.27	1.11	4.00	4.00	5.00
	CREA_16	1.00	6.00	4.45	0.95	4.00	4.00	5.00
	CREA_17	1.00	6.00	3.73	1.09	3.00	4.00	4.00
Critical thinking and problem solving skills	CRIT_1	1.00	6.00	4.53	0.91	4.00	5.00	5.00
	CRIT_2	1.00	6.00	4.57	0.86	4.00	5.00	5.00
	CRIT_3	1.00	6.00	3.98	1.03	3.00	4.00	5.00
	CRIT_4	1.00	6.00	3.99	1.04	3.00	4.00	5.00
	CRIT_5	1.00	6.00	4.09	1.27	3.00	4.00	5.00
	CRIT_6	1.00	6.00	3.59	1.15	3.00	3.00	4.00
	CRIT_7	1.00	6.00	4.39	0.96	4.00	4.00	5.00
	CRIT_8	1.00	6.00	4.09	1.14	3.00	4.00	5.00
	CRIT_9	1.00	6.00	4.34	0.99	4.00	4.00	5.00
	CRIT_10	1.00	6.00	4.48	0.89	4.00	5.00	5.00
	CRIT_11	1.00	6.00	3.53	1.09	3.00	3.00	4.00

	CRIT_12	1.00	6.00	4.09	0.98	4.00	4.00	5.00
	CRIT_13	1.00	6.00	4.32	0.89	4.00	4.00	5.00
	CRIT_14	1.00	6.00	3.99	0.93	3.00	4.00	5.00
	CRIT_15	1.00	6.00	4.23	0.93	4.00	4.00	5.00
	CRIT_16	1.00	6.00	4.28	0.96	4.00	4.00	5.00
Communication	COMM_1	1.00	6.00	3.94	1.28	3.00	4.00	5.00
	COMM_2	1.00	6.00	4.11	1.07	4.00	4.00	5.00
	COMM_3	1.00	6.00	4.25	1.04	4.00	4.00	5.00
	COMM_4	1.00	6.00	4.37	0.99	4.00	4.00	5.00
	COMM_5	1.00	6.00	4.43	1.10	4.00	5.00	5.00
	COMM_6	1.00	6.00	4.23	0.97	4.00	4.00	5.00
	COMM_7	1.00	6.00	4.18	1.00	4.00	4.00	5.00
	COMM_8	1.00	6.00	4.45	1.08	4.00	5.00	5.00
Collaboration	COLL_1	1.00	6.00	3.31	1.32	2.00	3.00	4.00
	COLL_2	1.00	6.00	3.59	1.18	3.00	4.00	4.00
	COLL_3	1.00	6.00	4.70	0.96	4.00	5.00	5.00
	COLL_4	1.00	6.00	4.60	0.92	4.00	5.00	5.00
	COLL_5	1.00	6.00	5.15	0.95	5.00	5.00	6.00
	COLL_6	1.00	6.00	5.15	0.98	5.00	5.00	6.00
	COLL_7	1.00	6.00	4.98	0.92	5.00	5.00	6.00

Descriptive statistics measurement 3 on item level

Table 3

Descriptive Statistics for Observed Variables 'Twenty First Century Skills' during measurement 3, N=572

Latent factor (scale)	Variable	Min	Max	Mean	SD	25%	50%	75%
Creativity and innovation	CREA_1	1.00	6.00	3.98	1.14	3.00	4.00	5.00
	CREA_2	1.00	6.00	4.23	0.93	4.00	4.00	5.00
	CREA_3	1.00	6.00	3.96	1.04	3.00	4.00	5.00
	CREA_4	1.00	6.00	4.11	1.10	4.00	4.00	5.00
	CREA_5	1.00	6.00	4.19	1.05	4.00	4.00	5.00
	CREA_6	1.00	6.00	4.25	0.95	4.00	4.00	5.00
	CREA_7	1.00	6.00	3.91	1.16	3.00	4.00	5.00
	CREA_8	1.00	6.00	3.83	1.10	3.00	4.00	5.00
	CREA_9	1.00	6.00	4.37	1.12	4.00	5.00	5.00
	CREA_10	1.00	6.00	4.55	1.00	4.00	5.00	5.00
	CREA_11	1.00	6.00	4.51	0.95	4.00	5.00	5.00
	CREA_12	1.00	6.00	4.04	0.97	4.00	4.00	5.00
	CREA_13	1.00	6.00	4.17	0.98	4.00	4.00	5.00
	CREA_14	1.00	6.00	5.31	0.86	5.00	5.50	6.00
	CREA_15	1.00	6.00	4.30	1.13	4.00	4.00	5.00
	CREA_16	1.00	6.00	4.56	0.98	4.00	5.00	5.00
	CREA_17	1.00	6.00	3.70	1.07	3.00	4.00	4.00
	CRIT_1	1.00	6.00	4.73	0.84	4.00	5.00	5.00
	CRIT_2	1.00	6.00	4.73	0.83	4.00	5.00	5.00

Critical thinking and problem solving skills	CRIT_3	1.00	6.00	4.05	1.01	3.00	4.00	5.00
	CRIT_4	1.00	6.00	4.10	1.06	4.00	4.00	5.00
	CRIT_5	1.00	6.00	4.03	1.35	3.00	4.00	5.00
	CRIT_6	1.00	6.00	4.28	1.04	4.00	4.00	5.00
	CRIT_7	1.00	6.00	4.57	0.94	4.00	5.00	5.00
	CRIT_8	1.00	6.00	4.23	1.10	4.00	4.00	5.00
	CRIT_9	1.00	6.00	4.54	0.95	4.00	5.00	5.00
	CRIT_10	1.00	6.00	4.60	0.86	4.00	5.00	5.00
	CRIT_11	1.00	6.00	4.06	0.98	4.00	4.00	5.00
	CRIT_12	1.00	6.00	4.22	1.08	4.00	4.00	5.00
	CRIT_13	1.00	6.00	4.46	0.88	4.00	5.00	5.00
	CRIT_14	1.00	6.00	4.00	0.93	3.00	4.00	5.00
	CRIT_15	1.00	6.00	4.35	0.95	4.00	4.00	5.00
	CRIT_16	1.00	6.00	4.44	0.97	4.00	5.00	5.00
Communication	COMM_1	1.00	6.00	4.36	1.17	4.00	5.00	5.00
	COMM_2	1.00	6.00	4.30	1.10	4.00	4.00	5.00
	COMM_3	1.00	6.00	4.46	1.01	4.00	5.00	5.00
	COMM_4	1.00	6.00	4.51	1.00	4.00	5.00	5.00
	COMM_5	1.00	6.00	4.48	1.03	4.00	5.00	5.00
	COMM_6	1.00	6.00	4.33	0.98	4.00	4.00	5.00
	COMM_7	1.00	6.00	4.32	0.99	4.00	4.00	5.00
	COMM_8	1.00	6.00	4.54	1.11	4.00	5.00	5.00

Collaboration	COLL_1	1.00	6.00	3.87	1.27	3.00	4.00	5.00
	COLL_2	1.00	6.00	3.62	1.15	3.00	4.00	4.00
	COLL_3	1.00	6.00	4.92	0.87	5.00	5.00	5.00
	COLL_4	1.00	6.00	4.80	0.93	4.00	5.00	5.00
	COLL_5	1.00	6.00	5.44	0.83	5.00	6.00	6.00
	COLL_6	1.00	6.00	5.43	0.85	5.00	6.00	6.00
	COLL_7	1.00	6.00	5.23	0.82	5.00	5.00	6.00

Appendix E: Factor loadings Twenty First Century Skills on item level

Table 1

Factor loadings TFCS subscales on item level (with main latent variable)

Latent factor (scale)	Indicator	B (factor loading)	SE (standard error)	Z	Beta coefficient (standardized)	Sig P(>|z|)
Creativity and innovation	CREA_1	1.00			.53	.00
	CREA_2	1.27	0.04	35.78	.67	.00
	CREA_3	1.18	0.03	34.68	.62	.00
	CREA_4	1.26	0.04	35.66	.66	.00
	CREA_5	1.19	0.04	33.15	.63	.00
	CREA_6	1.25	0.04	33.87	.66	.00
	CREA_7	0.58	0.04	15.30	.31	.00
	CREA_8	0.58	0.04	15.04	.30	.00
	CREA_9	1.00	0.04	25.95	.53	.00
	CREA_10	0.89	0.03	26.08	.47	.00
	CREA_11	1.01	0.04	28.42	.53	.00
	CREA_12	1.21	0.04	32.78	.64	.00
	CREA_13	1.18	0.04	32.83	.62	.00
	CREA_14	0.73	0.04	18.69	.39	.00
	CREA_15	0.79	0.04	22.16	.41	.00
	CREA_16	1.03	0.04	28.69	.54	.00
	CREA_17	0.97	0.03	29.83	.51	.00

Critical thinking and problem solving skills	CRIT_1	1.00			.62	.00
	CRIT_2	1.03	0.02	49.36	.64	.00
	CRIT_3	0.85	0.02	37.21	.53	.00
	CRIT_4	0.80	0.02	33.42	.50	.00
	CRIT_5	0.62	0.03	23.68	.39	.00
	CRIT_6	0.35	0.03	12.15	.22	.00
	CRIT_7	1.07	0.02	44.98	.66	.00
	CRIT_8	0.91	0.02	37.32	.57	.00
	CRIT_9	0.91	0.02	37.80	.56	.00
	CRIT_10	1.07	0.03	43.50	.66	.00
	CRIT_11	0.47	0.03	16.14	.29	.00
	CRIT_12	0.82	0.03	32.58	.51	.00
	CRIT_13	0.97	0.03	38.94	.60	.00
	CRIT_14	0.89	0.03	34.63	.55	.00
	CRIT_15	1.14	0.03	45.49	.71	.00
	CRIT_16	1.13	0.03	44.37	.70	.00
Communication	COMM_2	1.00			.56	.00
	COMM_1	0.73	0.04	19.69	.40	.00
	COMM_3	1.05	0.04	29.88	.58	.00
	COMM_4	1.35	0.04	36.07	.75	.00
	COMM_5	1.32	0.04	34.20	.74	.00
	COMM_6	1.26	0.04	35.93	.70	.00

	COMM_7	1.14	0.04	32.53	.63	.00
	COMM_8	1.07	0.04	30.03	.60	.00

Appendix F: Figures of scale development between groups over time

Figure 1

Creativity development of both groups over time

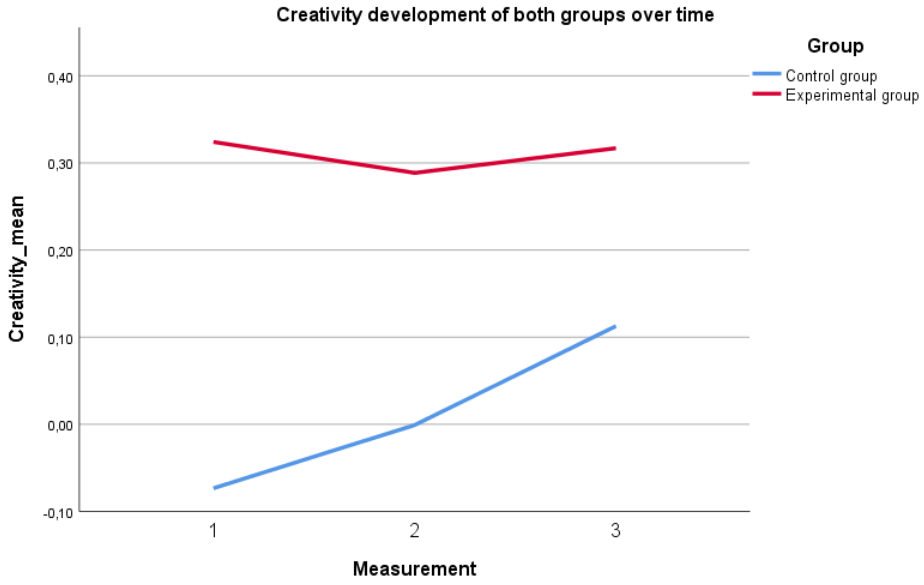


Figure 2

Critical thinking development of both groups over time

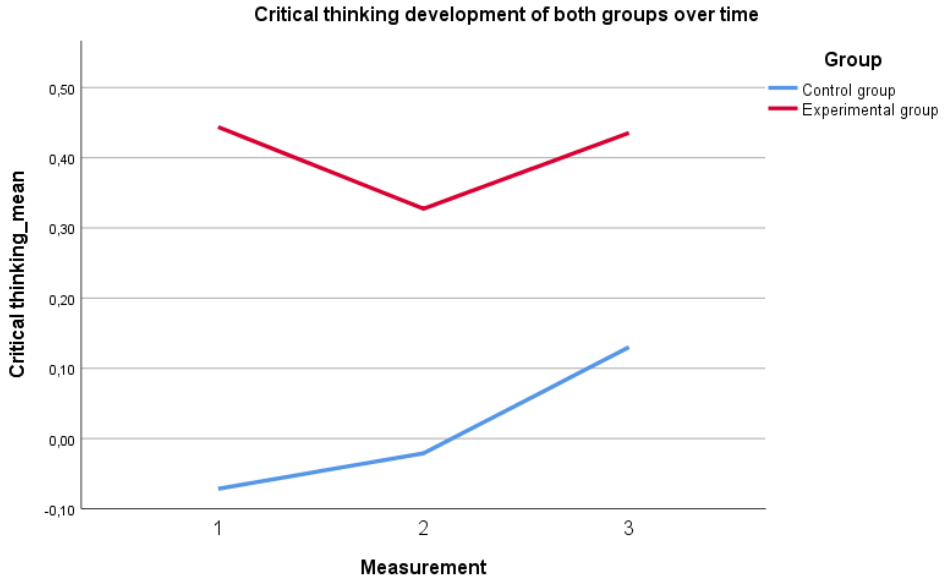
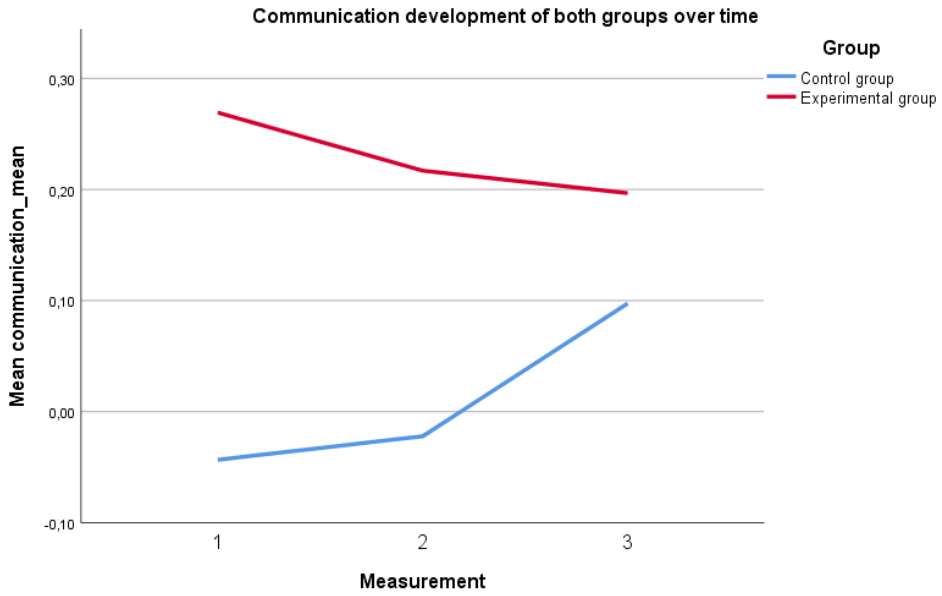


Figure 3

Communication development of both groups over time


Communication development of both groups over time
Group
Control group
Experimental group
Mean communication_mean
0,30
0,20
0,10
0,00
-0,10
1
2
3
Measurement

Appendix G: Figures of scale development between schools over time

Figure 1

Creativity development of schools over time

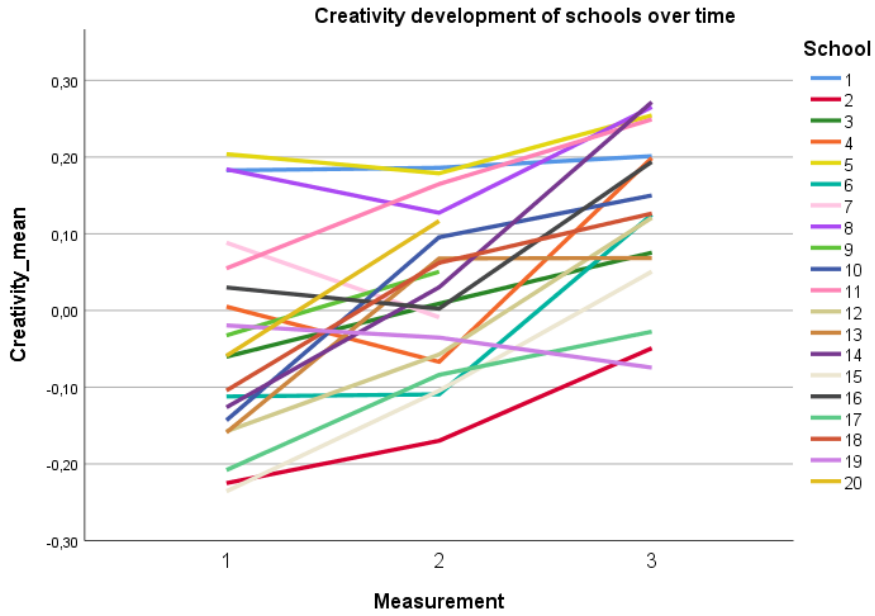


Figure 2

Critical thinking development of schools over time

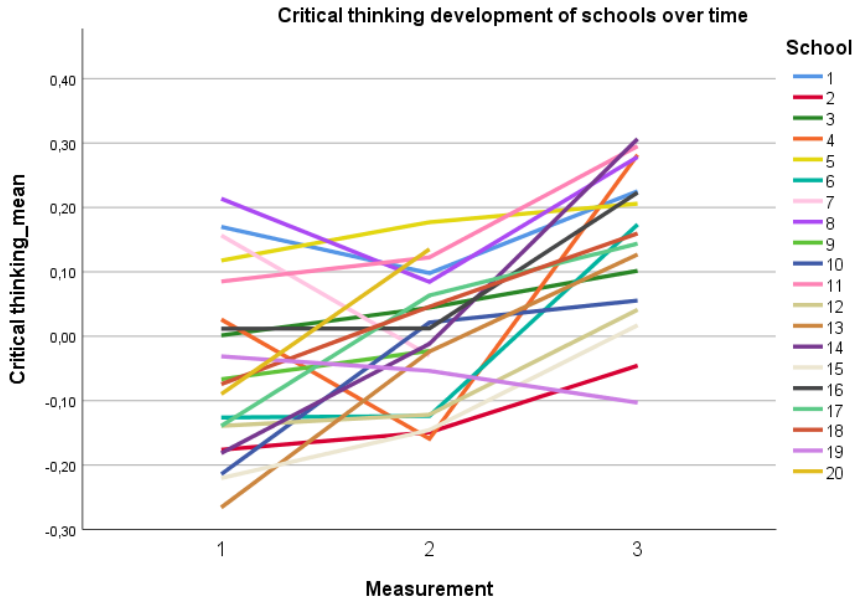
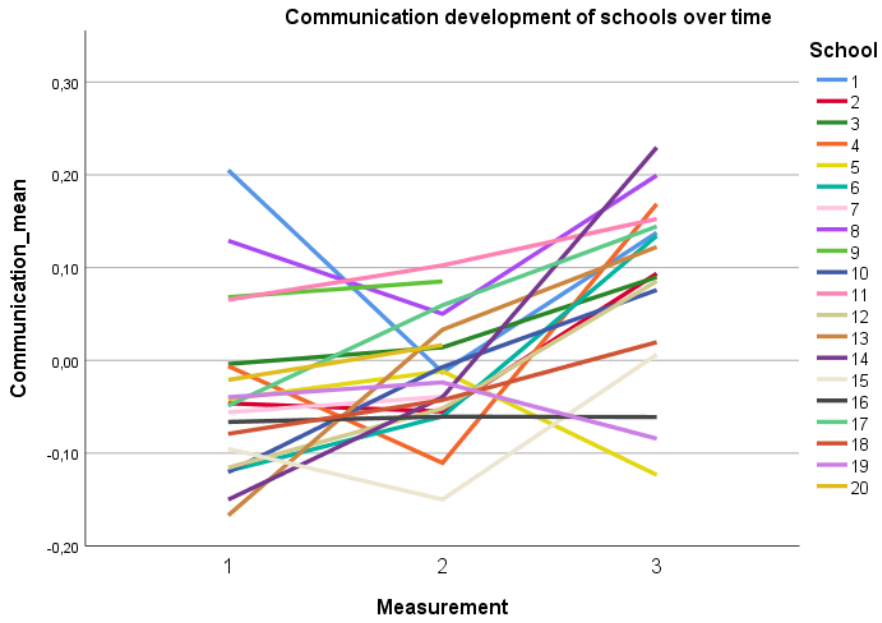


Figure 3

Communication development of schools over time


Communication development of schools over time
Communication_mean
0,30
0,20
0,10
0,00
-0,10
-0,20
1
2
3
Measurement
School
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20